

АКАДЕМИЯ НАУК СССР  
ИНСТИТУТ ЯЗЫКОЗНАНИЯ

Ю.Н.КАРАУЛОВ

ЛИНГВИСТИЧЕСКОЕ  
КОНСТРУИРОВАНИЕ  
И ТЕЗАУРУС  
ЛИТЕРАТУРНОГО  
ЯЗЫКА



ПИДАТЕЛЬСТВО «НАУКА»  
МОСКВА  
1981

В книге разрабатываются приемы построения новых лингвистических объектов, названные **принципами** лингвистического конструирования, характеризуется тенденция к лексикографической параметризации языка в современном языкознании, проходится параметрический анализ словарей. Предлагаемое понимание параметра находит применение в процессе построения русского тезауруса на ЭВМ. Результаты автоматической обработки информации, содержащейся в словарных дефинициях, анализируются затем с целью выявления соотношения лексического и синтаксического компонентов в их семантике.

Ответственный редактор

член-корреспондент АН СССР  
С. Г. БАРХУДАРОВ

## ОТ РЕДАКТОРА

Представляя читателю эту книгу, я хотел бы несколько слов сказать о ее авторе — Юрии Николаевиче Караплове. Это третья его книга. Две первые — «Общая и русская идеография» (1976) и «Частотный словарь семантических множителей русского языка» (1980) — пользуются заслуженной известностью и среди русистов, и среди специалистов по общему языкознанию, не только, как я думаю, в силу актуальности тех проблем, которые они рассматривают (т. е. относящихся к идеографии), или эффективности приемов их анализа (с использованием ЭВМ), но прежде всего благодаря оригинальности развиваемых в них идей. Разработанная Ю. Н. Карапловым схема лексико-семантического поля, сыгравшая первоначально роль средства системного описания объекта, теперь получила распространение в ряде работ по лексикологии и психолингвистической семантике, но уже в качестве характеристики, отражающей сущностное свойство самого объекта — языка. Наблюдения над двумя принципиально расходящимися классами лексики, обозначенными в «Общей и русской идеографии» как денотативный и десигнативный (сигнификативный), также получили развитие в исследованиях последнего времени. Вообще, можно сказать, что разработка многих проблем идеографии, ее теория и практика в значительной мере определяются научным вкладом автора данной книги. Это подтверждается и рецензиями на его работы, опубликованными в журналах «Филологические науки» (1977, № 4); «Вопросы языкознания» (1978, № 5); «Karogs» (1977, № 3); «Bulletin de la société de linguistique de Paris» (1979, f. 2); ежегоднике «Wissenschaft und Menschheit» (Leipzig u. a., 1978) и др.

Если попытаться в общей форме ответить на вопрос, специалистом в какой области языкознания является Ю. Н. Караплов, то оказывается в определенном затруднении. У нас есть ученые, которые попадают в разряд фонетистов и грамматистов, морфологов и синтаксистов,

словообразователей, историков и теоретиков. Однако пока не подобрано слово для называния человека, работающего в области лингвистической семантики: «семантику» — занято философами, «семантист» — выглядит непривычно и странно. Между тем именно этой сфере — сфере языковой семантики, исследователи которой не получили пока общепринятого обозначения, ученый отдает предпочтение, она является основной для Ю. Н. Карапулова.

Уже в первой его большой работе «Таксономия падежей и предлогов (семантика)» (1967) им была предпринята попытка системного анализа падежных отношений путем разложения грамматического значения падежа на элементарные смыслы, своеобразные «атомы» значения, с тем чтобы в их комбинации представлять любое значение, выражаемое предложно-падежной системой русского, падежной — эстонского и предложной — испанского языков. Разработанная им методика для вычленения этих «падежных примитивов» (метод трех оппозиций), на мой взгляд, нетривиальна и с успехом применяется специалистами.

Обращение Ю. Н. Карапулова к лексикографии не было случайным. В слове, как в капле воды океан, отражена вся языковая система, и познание ее содержательных свойств оказывается наиболее эффективным на конкретном объекте, отражающем эти системные отношения, а таковым, в частности в лексике, является тезаурус. Теоретические основы идеографических словарей, или тезаурусосстроения в сфере нетерминологической лексикографии, были намечены в «Общей и русской идеографии». Новая работа автора, продолжающая первоначальный замысел, отличается двумя чертами: во-первых, в центре анализа находится слово в совокупности его семантических и синтаксических характеристик, рассмотренных в связи с позицией слова в тезаурусе (системе); во-вторых, эти характеристики определяют предложенный автором, а затем реализованный им в эксперименте формализм машинного построения тезауруса.

Работа с ЭВМ — не просто как с бездушным, механически точным инструментом исследования, но почти как с соавтором — привела к установлению интересных статистических закономерностей в распределении семантических множителей в дефинициях толковых словарей. Анализ большого конкретного материала и осмысливание этих закономерностей в общетеоретическом плане позво-

или Ю. Н. Каулову сделать принципиально новые выводы о соотношении семантической и синтаксической информации в словарных определениях.

Важным результатом, на мой взгляд, является и характеристика впервые выявленной Ю. Н. Кауловым тенденции к лексикографической параметризации языка в современной лингвистике, т. е. стремления представлять результаты изучения языковых явлений разных планов и уровней в виде словарей. Выделение и рассмотрение на фоне этой тенденции совокупности словарных параметров, осуществленное автором, может послужить основой новой словарной типологии, элементы которой уже намечены в данной книге.

Итак, о чём же эта книга? На этот вопрос нельзя ответить однозначно. О семантике — да, о словарях и отражении семантики в их структуре — тоже да. Но что же такое «лингвистическое конструирование» — понятие, вынесенное в ее заглавие? С одной стороны, это совокупность приемом, обосновывающих эксперимент в языкоznании, с другой, — это одновременно и принципы трактовки экспериментальных результатов, их включения в обще-теоретический контекст науки, перехода от методик — к сущностным свойствам языка-объекта. Я бы назвал лингвистическое конструирование особым способом создания исследовательского пространства, которое удачным образом заполняет собой то звено между достижениями теории и методологии советского языкоznания, с одной стороны, и современной «лингвистической технологией», — с другой, которое пока является самым слабым в цепи автоматизации эксперимента в лингвистике и автоматического построения словарей. И на пути решения возникающих здесь проблем автор значительно продвинулся вперед.

*С. Г. Бархударов*

## ВВЕДЕНИЕ. ЦЕЛИ И ПРОБЛЕМЫ

Основное значение гипотез и теорий — кажущееся. Несмотря на то огромное влияние, которое они оказывают на научную мысль и научную работу данного момента, они всегда более преходящи, чем непрекаемая часть науки, которая есть научная истина...

*В. И. Вернадский. Размышления натуралиста*

Советское языкоznание характеризуется значительными достижениями в разных областях: в разработке принципов научного описания разносистемных языков, в создании грамматик и словарей многих языков народов СССР и мира, в изучении процессов развития и функционирования национальных языков в нашей стране, особенно в их взаимодействии с русским как языком межнационального общения. Заметными результатами отмечено и развитие теории, строящейся на базе марксистско-ленинской методологии, — прежде всего, в исследовании сущности и свойств языка как специфического структурного и общественного явления, в изучении исторических закономерностей его развития в связи с развитием общества. Вместе с тем в эпоху финализации наук, распространения программно-целевого подхода к научным исследованиям следует по-новому взглянуть на возможности лингвистики, как и всякой науки, в известном смысле заново оценить ее внутреннюю готовность к качественно иному этапу развития, к решению выдвигаемых жизнью практических задач.

### «ПОЧЕМУ» ИЛИ «ДЛЯ ЧЕГО»?

Смысл работы лингвиста, как и всякого ученого, — в поисках научной истины, которая в идеале должна представлять собой окончательные ответы на вопросы «что это такое?» и «почему это так?» Поиски эти в «нормальной науке», как отмечает Т. Кун, идут по линии анализа аномалий, рассмотрения парадоксов, решения задач-головоломок<sup>1</sup>. Лингвистическая теория, вырастая из эмпирии и предполагая экстраполяцию наших представлений об исследованных фактах, является тем не менее всего лишь обобщением уже известного и потому не всегда обладает достаточной прогностической силой, не всегда

дает ответы на новые вопросы, но зато помогает формулировать их; с ее помощью загадки если и не решаются, то обнаруживаются. Так возникают научные проблемы, и в их исследовании ученые отправляются от достигнутого уровня знаний. Естественно, этот путь далеко не всегда приводит к решению самих проблем, поскольку неожиданно открываются неизведанные области, выявляются новые трудности, возникают дополнительные неразрешимые загадки. Но вместе с тем, хотя какая-то проблема и остается нерешенной, растет сумма наших знаний, изменяется система представлений о предмете, а с нею может видоизмениться и сама исходная проблема. Так последовательно — кирпичик за кирпичиком, без пробелов и пропусков — строится целостное здание науки о языке.

На этом пути ученый как бы находится в свободном научном поиске: достигнув определенного результата, он при выборе следующей проблемы целиком опирается на этот полученный результат и делает следующий шаг часто не в направлении принципиально важного, а в направлении реально, с его точки зрения, достижимого. Безусловно, в этом есть своя логика; на таком пути мы рано или поздно обязательно получим систематический ответ и на вопрос «что это такое?», и на вопрос «почему?». Подобный путь предполагает поступательное движение всей науки и последовательное, без пробелов и пропусков, возведение всего здания. Но, как говорил Сенека, «великие оставили нам не только открытия, но и много ненайденного. Может быть, они и нашли бы необходимое, если бы не искали лишнего»<sup>2</sup>. Абсолютно все исследовать нельзя, и в наше время, — время программно-целевого подхода к организации научных исследований — зачастую более нужным и важным бывает ответ на вопрос «для чего это?» Тогда исходным пунктом становится не достигнутый уже уровень знаний, не полученный ранее результат, а общественная и научная необходимость конкретного решения данной проблемы, — т. е. будущий результат — и как отправной момент и как конечная цель исследования. Здесь самое время ввести различие «проблемы» и «цели» в науке. При этом мы, естественно, далеки от телеологической трактовки «цели» и понимаем ее в материалистическом и кибернетическом смысле: во-первых, как извест-

---

<sup>1</sup> Кун Т. Структура научных революций. М., 1975, с. 57—66.

<sup>2</sup> Сенека Л. А. Нравственные письма к Луцилию. М., 1977, с. 75.

ную пользу для человека, для общества, и во-вторых — как фиксированную в определенных материальных объектах информацию<sup>3</sup>.

Итак, проблема формулируется в ходе самостоятельного (можно даже сказать — изолированного) развития лингвистических идей и возникает как бы «изнутри» науки, знаменуя определенный этап ее роста, когда концепции перестают удовлетворять новым фактам, или теории, построенные на разных фрагментах языковой действительности, приходят в противоречие друг с другом. Цель же, наоборот, привносится в науку «извне», представляет собой как бы внешнюю по отношению к науке задачу, решение которой, однако, оказывается невозможным без использования средств и результатов разработки проблем именно данной научной области. Формулирование проблемы — дело специалистов в определенной области знаний, и любая проблема — это всего лишь имя собственное для очередного узла противоречий. Цели же формулируются, как правило, непрофессионалами, и поэтому можно сказать, что наука сама себе не ставит целей<sup>4</sup>. Рассмотрим некоторые примеры.

Проблемы в языкоznании, как и в других науках, бывают разных рангов. Есть проблемы вечные, к которым обращаются всегда и для которых каждый период развития науки предлагает новые решения, в идеале, очевидно, все более и более приближающиеся к истине. Таков, например, вопрос о происхождении языка, о самом механизме его возникновения. Попытка реконструировать этот механизм, основываясь на реликтовых явлениях в живых языках, на моделях других семиотических систем, осуществляется постоянно. Проблема эта соизмерима, вероятно, с проблемой происхождения жизни.

Есть проблемы специфические, приобретающие подчас форму парадоксов. Так, согласно наблюдениям грамматиков, подтвержденным на материале большого числа языков мира, во всех языках должна прослеживаться универсальная типологическая закономерность в эволюции грамматического числа — тенденция к сокращению противопоставленных друг другу форм, к уменьшению чисел, т. е. развитие языковой системы идет в направлении от оппозиции нескольких грамматических чисел (единственное vs. двойственное vs. тройственное vs. четвертое или множественное) к бинарной оппозиции единственного и множественного числа, но не наоборот. Однако в финно-угор-

ских языках как будто обнаруживается некоторая аномалия, отклонение от указанной универсалии: в саамском, хантыйском и мансийском языках имеется двойственное, и оппозиция в области грамматического числа тернарна — од. —дв. —мн., тогда как в других языках этой семьи<sup>5</sup> при наличии оппозиции ед.—мн. нет уверенных доказательств в пользу существования в них в прошлом двойственного числа. Если так, то либо неверна типологическая универсалия<sup>6</sup> и направление эволюции может быть реверсивным (от оппозиции ед.—мн. к оппозиции ед.—дв.—мн.), а значит, двойственное число может быть вторичным образованием; либо не подтверждается гипотеза о существовании единого прафинно-угорского языка-основы, из которого образовались современные языки этой общности. Другой пример проблемы-парадокса представляет собой соотношение слога и морфемы: слог — это единица фонетической, в конечном счете физиологической природы, и, казалось бы, никак не должен быть связан со значением, с семантической информацией, минимальной единицей которой как раз и является морфема. Однако имеют место факты, которые как будто опровергают этот вывод. При определенных видах афатических расстройств, характеризующихся нарушением некоторых семантических соответствий, у больного могут возникать трудности в идентификации именно слогов, а не морфем. Кроме того, в ряде языков Африки и Австралии зафиксированы случаи, когда табуирование какого-то имени вле-

<sup>3</sup> Макаров М. Г. Категория «цель» в марксистской философии. Л., 1977, с. 117; Щонев В. А. Необходимост и случайност в перехода от нежива към жива материя. Философски очерк за възникването на биологичната целесообразна активност. София, 1973, с. 27—37.

<sup>4</sup> Конечно, общая благородная цель науки — исследование закономерностей природы и общества для использования их на благо человека и человечества. Но и эта цель не порождается наукой в ходе саморазвития, ее ставит перед наукой общество.

<sup>5</sup> За исключением венгерского, где исторические и семиотические свидетельства наличия в прошлом двойственного числа достаточно надежны. Но венгерский не показателен в том отношении, что наиболее близок к хантыйскому и мансийскому, составляя с ними единую угорскую ветвь общей финно-угорской семьи.

<sup>6</sup> Этот парадокс можно сформулировать и мягче, если считать, что названная универсалия имеет не абсолютный («для всех без исключения языков мира верно, что . . .»), а статистический характер («для большинства языков верно. . .»). Но появление двойственного числа в языке как результата вторичного образования, т. е. после множественного, все равно представляется сомнительным.

чет за собой табу и на слог (или слоги) из этого имени. Табуирование же в чисто лингвистическом аспекте (отвлякаясь от его социального и этнологического смысла) — явление семантического порядка. Так как же связан слог с семантикой?

Бывают проблемы софистского толка. Такой я считаю, например, так называемую гипотезу Сепира—Уорфа, согласно которой каждый язык навязывает его носителям свою собственную логику в восприятии и оценке мира, обусловленную особенностями его грамматической структуры и лексической организации. В крайнем выражении, несколько утрируя проблему, ее формулируют так: будут ли совпадать картина некоторого внеземного мира, описанная на языке, допустим, бамбара, и картина того же мира, переданная, например, по-русски? Проблема породила громадное количество исследований, безусловно, плодотворных, позволивших прояснить многообразие связей логики и грамматики, выявить глубинные особенности структурной организации языков разных типов, возникло целое направление — этнолингвистика, но ... проблема осталась. Дело в том, что по самой своей логической структуре она весьма напоминает зеноновские парадоксы, поскольку молчаливо предполагается, что сравнение двух описаний (на языке бамбара и русском) проводится с некоторым третьим — эталонным описанием (метаописанием), созданным пусть на каком-то идеальном логическом языке (метаязыке), но все равно на языке, представляющем собой, следовательно, явление принципиально того же ряда, того же порядка, что и языки описаний, хотя и более высокого уровня. Но ведь в принципе можно проводить сравнение не самих описаний, а реальных объектов, на основе этих описаний построенных. Едва ли отличия будут касаться таких объектов.

Есть, наконец, проблемы-антиномии, которые сами по себе, вероятно, не имеют решения, но рассмотрение которых чревато для языкоznания плодотворными идеями и сдвигами. Такова антиномия синхронии и диахронии («язык существует, постоянно изменяясь во времени, но функционировать он может, только оставаясь неизменным»), осознание и анализ которой позволили решить ряд конкретных вопросов языковой номинации, выработать критерии системного рассмотрения исторических изменений в языке, сформулировать условия приятия или неприятия языком структурных инноваций и др.

Подобный перечень лингвистических проблем разного масштаба и разной степени важности можно было бы продолжать, но для нашей задачи — проведения разграничений между «проблемой» и «целью» — сказанное представляется достаточным.

Что касается целей, то они диктуются изменившимися условиями общественного развития (например, создание письменности, перевод письма на другой алфавит, усовершенствование существующей орфографии), общим прогрессом цивилизации, расширением и интенсификацией международных связей государства (повышение эффективности обучения родному языку, языку межнационального общения, иностранным языкам), успехами и достижениями других наук ( осуществление автоматического перевода, непосредственный ввод в машину звучащей речи, лингвистические проблемы искусственного интеллекта), иногда политическими соображениями (доказательства самостоятельности, достаточной развитости и высокой собственной культуры языка данной нации по отношению к инонациональному языку господствующего класса, например, эстонского в Эстонии XVIII—XIX вв. — по отношению к немецкому, или в Финляндии того же периода финского — по отношению к шведскому).

Как видим, подобные цели не могут возникнуть в ходе внутринаучного развития, и этот в общем-то случайный перечень «проблем» и «целей» с достаточной наглядностью иллюстрирует главное различие между ними: разные источники их возникновения, порождения. Этим различием в конечном счете определяются и другие свойства и признаки, характеризующие расхождения между проблемой и целью в науке.

Если цель по самой своей природе предполагает наличие определенного прагматического эффекта, определенного результата, то проблема лишена такого обязательного свойства, для нее прагматический аспект факультативен, хотя в принципе любая фундаментальная проблема может и должна иметь прикладной выход (который, впрочем, для самих языковедов, занимающихся разработкой данной проблемы, представляется, как правило, чем-то второстепенным, не заслуживающим серьезного внимания), поскольку наука никогда не развивается независимо от ее практических применений. Так, необходимость периодического совершенствования орфографии — это воплощение одного из следствий фундаментальной антиномии

синхронии и диахронии; получившие в последнее время широкое распространение контрастивно-сопоставительные грамматики, контрастивно-сопоставительные описания языков в учебных целях — пример того, как идеи фундаментальной типологии, став элементом обыденного лингвистического сознания преподавателей-практиков («внедрившись» в повседневное сознание), позволили повысить эффективность обучения иностранным языкам.

Далее, цель и проблема разнятся по глубине одновременного охвата уровней изучаемого объекта, по глубине проникновения в объект. Проблема, даже сложная, комплексная, никогда не является сквозной, в сфере ее действия попадает либо большая, либо меньшая, но всегда только часть объекта. При формулировании проблемы специалист ограничен рамками логики развития своей научной области, связан пониманием того, что можно и чего нельзя при данном уровне знаний об объекте. Скажем, ни один серьезный специалист не поставит сейчас задачу — выявить исходные звуковые и понятийные комплексы, универсальные для всех языков мира, построить, так сказать, основы лингвистической ойкумены, хотя историк конкретного языка даст последовательную и точную картину эволюции звуковой и содержательной стороны слова; этимолог путем сравнений зафиксированных письменных форм установит их иерархию во времени, а также фонетико-семантические соответствия в группе родственных языков; далее, компаративист, используя результаты первых двух исследователей, а также путем внутренней и внешней реконструкции выявит общие закономерности развития целой языковой семьи и восстановит архетипы, т. е. исходные формы пражязыка данной общности; наконец, специалист по проблемам номинации может установить некоторые общие для группы языков или универсальные вообще способы присвоения названий вещам и понятиям, способы соединения звука и значения. Таким образом, на каждом этапе, включающем разветвленную комплексную проблематику, имеется в виду определенный фрагмент объекта. Цели же, как правило, имеют сквозной по отношению к объекту характер и формулируются вне зависимости от разработанности — неразработанности той или иной области. В этом смысле цель как будто нарушает, внутреннюю логику развития науки, но вместе с тем и стимулирует самое это развитие. Такая, на первый взгляд, простая цель, как изменение существующей орфографии <sup>7</sup>,

проведение орфографической реформы требует рассмотрения и решения целой цепочки проблем, пронизывающих все уровни языка и выходящих за его пределы — в сферы психологии, педагогики, общественных отношений, не говоря уже о трудно предсказуемом отражении последствий этой реформы на громадном по объему комплексе терминосистем естественных и технических наук. Один только перечень этих проблем показывает, что орфография представляет собой специфический социальный инструмент, изменить который далеко не просто. Проблемы эти следующие: соотношение нового графемного изображения с современными произношением (т. е. проблемы фонетики и фонологии) и морфологическим членением слова, а новой пунктуации — с законами синтаксиса; отражение изменений на основном принципе письма на данном языке (т. е. сохранение либо звукового, либо слогового, либо морфемного по преимуществу, либо идеограммного типа письма); соответствие вводимых изменений исторически сложившемуся облику и тенденциям в развитии фонетики слов<sup>8</sup>; соблюдение в новой орфографии закономерностей звукового (а значит — и семантического) освоения заимствуемых слов в их соотношении со словами исконными. Далее, в этот перечень проблем включаются психологические моменты, связанные с необходимостью учитывать последствия неизбежной ломки установившегося у носителей языка навыка письма и навыка в восприятии внешнего вида слов, непосредственно ассоциируемого с их семантикой. Наконец, надо учитывать общественную реакцию, которая, как показывает опыт ряда стран (например,

---

<sup>7</sup> Кажущаяся простота изменения правописания внушается отчасти, видимо, тем обстоятельством, что общенародный язык, язык повседневного общения, который и должен в первую очередь стать объектом этих изменений, в условиях НТР занимает по своему объему (т. е. составу лексем и правил оперирования ими) ничтожную и постоянно уменьшающуюся долю в ряду терминологических систем специальных наук. Если 30 тыс. слов (лексем) литературного языка хватает для эффективной культурно-социальной коммуникации, то число терминов в области одной только органической химии превышает 300 тыс. Причем терминологические системы обладают четко выраженным специфическими морфологическими и синтаксическими особенностями, т. е. образуют относительно самостоятельные подъязыки.

<sup>8</sup> Дело в том, что анализ ошибок школьников в родном языке позволяет установить совершенно четкие, однозначные корреляции некоторых типов этих ошибок с формами, исторический засвидетельствованными на ранних этапах развития данного языка; т. е. встает проблема научного осмыслиения статуса самой ошибки,

Японии), может быть очень бурной и приобретать *крайние* формы выражения. Как видим, постановка и разработка целей в языкоznании требует использования значительного числа (если не всех) достигнутых результатов, мобилизации всех ресурсов науки, а кроме того, выдвигает и новые проблемы, стимулирующие ее развитие<sup>9</sup>.

Сказанное, однако, вовсе не означает какого-то пре-восходства, приоритета целей перед проблемами из-за глобальности и всеобъемлемости первых. Вне зависимости от их стимулирующего действия цели могут быть как истинными, так и ложными. Поиски «философского камня» хотя дали определенный толчок развитию химии и ее практиче-ских приложений, но все же стали и тормозом в раз-витии целой научной эпохи. Далее, если научные проб-лемы существуют как бы вне времени, то цели носят всегда временный, преходящий характер, и достижение целей (скажем, в установленный планами срок) еще не оз-начает решения всех входящих в них проблем. Некоторые звенья из цепочки проблем, ведущих к цели, либо при этом вовсе исключаются, либо получают временное, тех-ническое решение. Вот здесь-то мы и подходим вплотную к понятию лингвистического конструирования. Но прежде чем перейти к этому понятию, закончим рассмотрение различий между проблемой и целью.

Будучи выдвинутыми, как отмечалось выше, самой жизнью, т. е. экстралингвистической реальностью, цели формулируются часто не на профессиональном языке науки, в данном случае — языкоznания, и задача специа-листов — перевести их на язык собственно лингвистиче-ской проблематики, переформулировать таким образом, чтобы они представили в виде четкой последовательности подлежащих решению проблем. Например, важная обще-государственная программа, направленная на соверше-ствование функционирования в стране русского языка как средства межнационального общения народов СССР, включает ряд частных целей — повышение уровня зна-ния русского языка в национальных республиках путем развития всеобщего двуязычия, расширение сфер его применения во внутриреспубликанском и межреспубли-канском общении, улучшение методики обучения ему в высшей и средней национальной школе. Известно, что полноценное двуязычие возможно только на базе совер-шенного владения родным языком, поэтому одновременно встает задача совершенствования преподавания националь-

ных языков, повышения национальной речевой культуры народа. На этом пути новые решения проблем, новые идеи и рекомендации должны последовать не только от уже имеющей определенные достижения социальной лингвистики («язык в обществе»), но прежде всего от психолингвистики («язык в индивидууме»), которая при разработке основ владения двумя языками и двуязычного обучения может вскрыть неизвестные резервы как в человеческой психике, так и в особенностях организации самой языковой материи. В частности, заслуживает внимания высказываемая в литературе последних лет гипотеза о том, что овладение вторым языком проходит через этап его пиджинизации и креолизации в сознании индивида<sup>10</sup>. Если это предположение подтвердится, то из него можно извлечь конструктивные методические следствия для промежуточных этапов обучения второму языку.

Далее, для повышения эффективности двуязычного обучения языковедам предстоит решить также проблему достижения максимальной компрессии языковых фактов и поиска наиболее экономных способов фиксации и представления языковой структуры. Решения этой проблемы скорее всего можно ожидать, вероятно, на пути построения либо порождающих, либо контенсивно-типовологических (структурно-семантических) грамматик, на пути создания национально-русских и русско-национальных тезаурусов. Однако сдвиги в этом направлении пока минимальны, поскольку лингвистика, оставаясь наукой по преимуществу описательной<sup>11</sup>, до последнего времени не разви-

---

<sup>9</sup> Причем соотношение в данном случае между «стимулом» и «реакцией» может быть самым неожиданным. Так, источником учения о фонеме и последующего развития фонологии как новой отрасли языкоznания, обогатившей в первой половине нашего столетия своими идеями всю теоретическую лингвистику, была цель создания всего лишь практической транскрипции, которая возникла в свое время в связи с возрастанием интереса к изучению живых языков и диалектов.

<sup>10</sup> *Adjeman Ch.* On the nature of interlanguage systems. — Language learning, Ann Arbor, 1976, v. 26, № 2; *Schumann .. H.* Second language acquisition. The pidginization hypothesis. — Idem.

<sup>11</sup> Характеризуя языкоznание подобным образом, я имею в виду работы не только констатирующего, собственно дескриптивного плана (т. е. выполненные на уровне наблюдения непосредственно над фактами языка), но и широкий круг обобщающих исследований, в которых развиваются авторами идеи, концепции, теории служат задачам анализа и объяснения языковых закономерностей. Это ведь по сути дела тоже описание, но описания «второго порядка». Увлечение таким объективированным описательством

вала взгляда на свой предмет — язык как на возможный субъект научного исследования. Естественный язык, будучи только объектом изучения, долгое время оставался для языковедов феноменом неприкосновенным, некоторым образом «вещью в себе». Лишь в последнее время — в связи с возрастанием общей суммы знаний и определенным изменением наших представлений о языке вообще, в связи с возникновением новых задач и попытками экспериментирования над естественным языком — появились основания для пересмотра устоявшегося взгляда, поскольку намечается особое направление исследовательской деятельности — построение новых лингвистических объектов с заданными свойствами. Реализация этих задач и должна осуществляться на принципах лингвистического конструирования (ЛК).

### ЭВРИСТИКИ ЛК

Лингвистическое конструирование — это не инженерная лингвистика, которая решает вполне определенные, конкретные задачи машинной обработки языка и его машинного использования, но тем не менее некоторые решения и инженерной лингвистики получены путем ЛК. Лингвистическое конструирование — это совокупность обобщенных способов и приемов компиляции и комбинирования «образцов решений проблем», экстраполяции уже имеющихся, готовых теоретических и практических результатов, полученных в разных областях лингвистики, и их прямого или эвристического использования для преодоления трудностей и решения проблем, возникающих в тех же или других областях при построении новых лингвистических объектов. Создать, построить какую-то «вещь», значит не только уметь объяснить те свойства языка, которые в ней использованы и на которых она основана, не только объяснить определенные закономерности языковой структуры, но и выяснить новые свойства построенного объекта, так или иначе характеризующие исследуемый язык, а значит — расширить наши знания о человеческом языке вообще. Таким образом, новым лингвистическим объектом будем называть такое представление фактов, языковых данных, которое генерирует новую информацию о языке. Как правило, такие объекты получаются не в результате описания некоторого материала, а возникают как результат эксперимента, причем

эксперимента, понимаемого широко. В этом смысле и ностратическую теорию, и порождающую грамматику, имеющие целью каждая создание нового объекта, следует трактовать как лингвистический эксперимент.

Глининый принцип лингвистического конструирования — «как сделать» тот или иной объект, — только на первый взгляд может представляться чисто техническим по своему содержанию. На деле такой подход таит много новых возможностей, новых выходов на фундаментальные проблемы науки о языке. В последнее время многие новые движения, идеи, даже достижения в языкоznании получены именно на пути построения новых объектов. Когда возникла идея Лингвистического атласа Европы или энциклопедии «Языки мира», то прежде всего встал вопрос «как сделать». Для энциклопедии, например, решение этого вопроса предполагает анализ соотношения разносистемных языков, возможностей трансформируемости единиц, их структурирования и критериев описания, возможностей использования при этом уже известных универсалий и понятийных категорий высокой степени общности (типа бытия, наличия, обладания, лица—нелица и др.) и выделения новых универсалий. Все это проблемы не технические, а фундаментальные. Или, скажем, идеи генеративной грамматики, которые сами по себе могут вызывать разное отношение. Основания для скепсиса здесь существенны: до сих пор на основе порождения не удалось сконструировать достаточно серьезной грамматики ни для одного языка. Тем не менее исследования в этом направлении оказались весьма плодотворными, поскольку, во-первых, выявили ошибочность позиции Хомского, ориентированного только на синтаксис, и тем самым как бы заново открыли для лингвистов важную

---

хотя и довольно высокого уровня было, пожалуй, особенно широко представлено в языкоznании конца 60-х годов, когда в значительной части теоретических работ научность подменялась свойством, которое я называю «комплексом цепречастности автора» и которое проявляется в том, что автор как бы самоустранился: вот, дескать, материал, вот совершенно объективная методика, приводящая вот к такому-то результату, а я здесь не при чем. Это и есть позиция чистого описательства. Альтернативу подобному подходу составляют работы, характеризующиеся адресностью, итогом которых является новый лингвистический объект вполне определенной предназначенностю — словарь, атлас, грамматика нового типа, действующая модель некоторого процесса, синтезаторы языковых феноменов (звуков, форм, высказываний) и др.

область — изучение взаимоотношений синтаксиса и семантики, которая интенсивно развивается и наши дни наиболее богата идеями и результатами. Во-вторых, стала очевидной несостоятельность притязаний конструктивной по своему существу порождающей модели «смысл—текст» на роль теоретической лингвистической концепции. Ограничивааясь построением системы правил перехода «от значения к звуку» (для говорящего) и «от звука к значению» (для слушающего), она не покрывает в качестве теории задач изучения человеческого языка, и вне ее компетенции остаются прежде всего два важнейших его аспекта: 1) откуда берется изначальный смысл (это область взаимоотношений языка и сознания, в том числе проблема происхождения языка); 2) как один и тот же смысл может выливаться в разные тексты, а один текст интерпретироваться разными смыслами (это область стилистики, а шире — взаимосвязей языка и общества).

Следовательно, лингвистическое конструирование — это не только способ создания новых «вещей», но и метод изучения, выявляющий существенные стороны языкового материала, предполагающий как этап его анализа, так и этап синтеза. Значит, при лингвистическом конструировании, в ходе построения новых лингвистических объектов, сам язык выступает и как предмет эксперимента, и как субъект научного исследования.

Каковы же эти возможные «новые лингвистические объекты» и в чем заключаются другие (помимо принципа «как сделать») приемы лингвистического конструирования? В числе таких объектов — тезаурус общелитературного русского языка, построенный с помощью электронно-вычислительной машины, действующая модель порождения словоизменительных форм (т. е. автоматический морфологический словарь, предназначенный как для анализа, так и для синтеза) эстонского языка, автоматический словообразовательный словарь русского языка, подготовленный в Институте русского языка АН СССР.

О ряде других искусственных лингвистических объектов можно говорить пока предположительно, как о результате разработки специализированных долгосрочных программ. Но общее для всех них то, что они относятся не к сфере языкового кода, языковой системы, которая исследована к настоящему времени достаточно хорошо, а к сфере ее функционирования, т. е. к сфере перехода от кода — к высказыванию, тексту. Среди этих объек-

топ — исследование и построение механизмов освоения второго (неродного) языка и владения им (или шире — лингвистической модели двуязычия), создание моделей порождении речи вообще, речевого воздействия, наконец, глобальной модели связи и взаимодействия языка с мышлением, т. о. искусственного интеллекта. При построении этих объектов принимаемые в каждом отдельном случае конкретные решения могут в чем-то отклоняться от общепринятых теоретических положений, от известных закономерностей языковой структуры, в чем-то, может быть, даже нарушать их, но при этом будет достигаться главная цель, ради которой данный объект и создается — обучение языку, порождение речи и т. п. Эти нарушения могут выражаться во введении специфических ограничений (например, в алгоритме тезауруса полисемия дескриптора снимается тем, что кодирование дескриптора производится только по одному его — главному или же общему — значению); в использовании в таких объектах (например, в модели обучения) некоторых искусственных промежуточных единиц, не совпадающих с традиционно выделяемыми — фонемой, морфемой, словом, предложением; в установлении необычных связей между элементами структуры. Нисколько не стремясь парадоксальным образом акцентировать «право на ошибку», хотелось бы отметить в этой связи принципиальную полезность некоторых ошибок (плохих чертежей, некорректных схем и т. п.)<sup>12</sup> на пути к психическому мутагенезу, порождению новых идей: «Наконец, возможно и сознательное управление механизмом прихода к „мутации“... В частности, поиск инвариантов есть сознательное применение механизма некорректного склеивания объектов, т. е. получение „ошибочных“ промежуточных исчислений не только допустимо, но и желательно»<sup>13</sup>. Недаром бытует такое представление о творчески работающем ученом, что он должен каким-то образом ухитряться забыть о некоторых вещах и в то же время помнить о том, что забыл их.

В поисках новых технических решений лингвистическое конструирование может прибегать даже к *ad hoc* гипотезам и *ad hoc* конструкциям, что методологически

---

<sup>12</sup> Ср. наблюдения о пользе переводческих ошибок: *Valesio P. The virtues of traducement: sketch of a theory of translation.* — *Semiotica*, The Hague, 1976, v. 18, N 1.

<sup>13</sup> *Маслов С. Ю.* Теория поиска вывода и вопросы психологии творчества. — Семиотика и информатика, вып. 13. М., 1979, с. 45.

представляется в данном случае оправданным, поскольку запрет на их использование в значительной степени объяснялся, на наш взгляд, кажущейся авторитетностью неопозитивистских требований обязательного гипотетико-дедуктивного характера теории. Между тем *ad hoc* гипотезы, использованные при движении к цели в рамках лингвистического конструирования, в ретроспективном плане, при последующих обобщениях, зачастую приобретают содержательную интерпретацию и могут становиться полноправными элементами теории. Введенное нами конструктивное «правило шести шагов», объясняющее не «почему», а «как» связаны единицы в словаре<sup>14</sup>,казалось на первых порах каким-то «трюком», игровым семантическим приемом. Однако при дальнейшем осмыслении обнаружилась его связь со свойствами человеческой памяти, с одной стороны, и с некоторыми фундаментальными принципами системологии, с другой. В ряду эвристических процедур *ad hoc* гипотезы укладываются в так называемый «механизм прыжка», характеризующий нелевополушарное (недискретное) мышление: «„Содержательные“ производные правила чаще возникают при нацеленном „прыжке“ к желательному выводимому объекту с последующим восстановлением „вставки“... „Прыжок“ хорошо согласуется с принципом нацеливания и, следовательно, с упомянутой моделью взаимодействия сознательного и бессознательного. „Желательный объект“ может возникать на основе волевого выбора цели и/или некоторого перебора»<sup>15</sup>. Напомним также хорошо известный из истории науки факт, когда М. Планк, пытаясь найти выход из теоретического противоречия, согласно которому равновесие между веществом и энергией казалось невозможным, несколько видоизменил классическую формулу излучения, введя новую постоянную («постоянная Планка»), и показал, что противоречие устранимо. Понять настоящий глубокий и универсальный смысл этой постоянной, введение которой представлялось многим чисто формальным шагом, удалось лишь позже. Физика обязана этим Эйнштейну — он первый понял фундаментальное значение открытия Планка и дал ему более общее физическое толкование<sup>16</sup>.

Здесь мы в наших рассуждениях подошли к рассмотрению важнейшей закономерности в движении человеческого познания — принципа оборачивания метода, широко используемого, в частности, и лингвистическим конструированием. Сущность этого познавательного приема впер-

вые была раскрыта К. Марксом, когда, занимаясь дифференциальным исчислением, в своих «Математических рукописях» он показал диалектическую природу одного из важнейших математических операторов — дифференциала. Маркс выявил двойственную природу дифференциального коэффициента, диалектическое единство двух его сторон — предметной и оперативной, переход которых друг в друга он назвал оборачиванием метода (*Um-schlag in der Methode*)<sup>17</sup>. Этот прием характеризуется двумя эвристически универсальными моментами. Во-первых, он предполагает превращение одного метода (в данном случае алгебраического) в другой, противоположный ему (в данном случае дифференциальный), превращение, которое воплощается в том, что содержательные алгебраические символы становятся формальными символами нового, дифференциального исчисления. Во-вторых, результат, полученный старым методом, выступает как нечто вторичное по отношению к исходному материалу познания, и «оборачивание метода», по существу, состоит в том, что вторичное как нечто производное, несамостоятельное, как результат, зависящий от некоторого исходного материала, превращается в свою противоположность и тем самым становится исходным самостоятельным пунктом нового процесса, своего рода «независимой переменной», а то, что ранее выступало в качестве исходного пункта познавательного процесса, было первичным в старом методе, наоборот, становится чем-то производным, вторичным, своего рода «зависимой переменной»<sup>18</sup>. В. С. Черняк показывает далее, что «аналогичный принцип, который, пользуясь выражением самого Маркса, лучше назвать „оборачиванием ролей“», Маркс использовал и ранее — уже в работах «К критике политической экономии» и «Капитал» — при анализе функционирования денег в качестве средства обращения. Таким образом, оборачивание метода представляет собой важнейший прием исследования и одновременно форму развития не только математической мысли или эмпирических наук, но и мысли философской, и наук общественных.

---

<sup>14</sup> Караплов Ю. Н. Общая и русская идеография. М., 1976, с. 77—79.

<sup>15</sup> Маслов С. Ю. Указ. соч., с. 41.

<sup>16</sup> Капица П. Л. Эксперимент, теория, практика. М., 1974, с. 20.

<sup>17</sup> См.: Черняк В. С. Оборачивание метода и диалектика развития знания. — Вопросы философии, 1975, № 8, с. 101.

<sup>18</sup> Там же, с. 102.

Оборачивание метода продемонстрировано было лишь переходом от содержательного истолкования символов к их операционному определению и использованию. Но сам принцип допускает, очевидно, и возможность противоположного перехода. В этом убеждает нас, в частности, рассмотренный выше пример с «постоянной Планка», которая, возникнув в скромной роли операционного символа, обрела затем фундаментальный физический смысл. Это не означает, естественно, что за каждым методическим оператором можно искать определенное содержание, как неверным было бы утверждение, будто любой содержательный элемент благодаря оборачиванию метода способен выполнять роль формального оператора<sup>19</sup>. Однако, если рассматривать тот или иной метод научного исследования, с одной стороны, как производное от соответствующей теории, а с другой, — как зеркальное отображение какой-то стороны или сторон изучаемого объекта (а только так, на наш взгляд, метод и надо рассматривать<sup>20</sup>), то мы вправе ожидать наличия в любом методе наряду со вспомогательными, чисто формальными операторами и таких операционных символов, которые могут соответствовать не обнаруженным еще, но открытым наукой реальным свойствам изучаемого объекта, т. е. могут получить содержательную интерпретацию по принципу оборачивания метода.

Обращаясь к лингвистическому конструированию, проиллюстрируем принцип оборачивания метода примером с трактовкой семантического множителя при построении с помощью ЭВМ нового лингвистического объекта — тезауруса русского литературного языка. Идея содержательных множителей, или семантических компонентов, из которых слагается смысл слова и словосочетания, было придано операционное наполнение таким образом, что множителем предложено было считать каждое слово, входящее в состав словарной дефиниции<sup>21</sup>. Функционирование так понимаемых семантических множителей уже в роли операторов, во-первых, привело к решению поставленной задачи, а во-вторых, обнаружило такие их свойства, которые при возвращении к содержательной интерпретации компонентов четко выявили три их типа: денотативные, реляционные и синтаксические (подробнее см. ниже, § 9–11).

Перейдем к характеристике двух следующих взаимосвязанных принципов лингвистического конструирова-

ния. Во-первых, это неразличение обычных уровней в языковой структуре, что дает возможность нетрадиционно, по-новому категоризировать исследуемый объект в ходе решения определенной задачи. Так, новая категоризация лексикографической информации позволяет построить словарную типологию на содержательной основе и предложить принципиально новые типы словарей<sup>22</sup>; иная категоризация языкового строя оказывается необходимой при машинном переводе, когда лингвистическое описание делается в алгоритмических целях.

Во-вторых, это введение новых единиц описания. Так, при построении русского тезауруса для кодирования

---

<sup>19</sup> «В этой связи можно сослаться на парадокс, который указал Арно в письме Лейбницу. Этого автора смущила пропорция  $\frac{1}{-1} = \frac{-1}{1}$ ,

в которой, как он писал, большее, деленное на меньшее, равно меньшему, деленному на большее. Характерно, что и сам Лейбниц считал эту пропорцию не истинной, а лишь воображаемой. Причиной появления этого мнимого парадокса было истолкование положительных чисел как больших, а отрицательных — как меньших количеств. Никакого парадокса не будет, если рассматривать числа как оперативные символы, если не сводить оперативное значение математической символики к какой-то одной интерпретации» (Черняк В. С. Указ. соч., с. 105).

<sup>20</sup> Для подтверждения справедливости такой позиции сошлемся на предложенное Ю. С. Степановым трехчастное строение метода, включающего «методику» (вопрос о способах выявления нового материала и его введения в научную теорию), собственно «метод» (вопрос о способах систематизации и объяснения этого материала) и «методологию» (вопрос о соотнесении и способах соотнесения уже систематизированного и объясненного материала с данными смежных наук и прежде всего с философией) (Степанов Ю. С. Методы и принципы современной лингвистики. М., 1975, с. 5). Это расчлененное представление структуры метода позволяет показать, что две первые его части обращены, несомненно, к объекту и направлены прежде всего на поиск ответов на вопросы «что» и «как». Третья же часть, или уровень организации, метода относится главным образом к теории и предполагает получение ответа на вопрос «почему». Одновременно благодаря такому представлению становится ясной принципиальная возможность заимствования методов из других наук и использования одинаковых методов в рамках несовместимых одна с другой теорий.

<sup>21</sup> См.: Карапулов Ю. Н. Частотный словарь семантических множителей русского языка. М., 1980; см. также ниже, § 6, 7.

<sup>22</sup> Карапулов Ю. Н. Об одной тенденции в современной лексикографической практике. — В кн.: Русский язык. Проблемы художественной речи. Лексикология и лексикография. М., 1981.

семантической информации использовано понятие квази-  
основы (аппаратный, победитель, неожиданный и т. п.), которая определяется правилами а) минимизации  
числа букв, б) возможности семантического объединения  
словоизменительных и части словообразовательных форм,  
в) обратимости, т. е. однозначности декодирования. Она  
не совпадает ни с корнем слова, ни с его морфемным членением и служит для точного опознания соответствующей гиперлексемы и идентификации обобщенного смысла составляющих ее единиц.

Наконец, важной характеристикой принципов лингвистического конструирования является понятие «усреднения», средней величины. Казалось бы, всякое отступление от логической строгости, использование огрубленных, приблизительных величин должно вести к нарушению точности научного исследования. Однако при построении новых объектов, предполагающем использование очень большого числа данных, усреднение по некоторым признакам оказывается не только допустимым, но и неизбежным, а в итоге может привести к большей точности при целостном восприятии объекта. Известно, как много противоречий и непоследовательностей встречается в словарных дефинициях больших толковых словарей. Но эти противоречия становятся очевидными лишь при системном анализе словаря, т. е. при его целостном восприятии. Наличие этих противоречий есть закономерный итог основного принципа лексикографической работы, являющегося по сути дела атомистическим, согласно которому следует стремиться к максимальной точности в описании семантики отдельного слова. Сам по себе этот принцип, безусловно, хорош, однако последовательное его осуществление ведет, как ни странно, к искажениям в системе в целом. Иными словами, лексикограф всегда стоит перед альтернативой: либо «точность» по отношению к отдельному слову и заведомые нарушения в интерпретации лексической системы, либо логическая строгость в представлении системы и некоторая огрубленность в характеристике отдельных слов. Компромисс достигается на путях лингвистического конструирования. Так, для выявления системных отношений в сфере лексической семантики используется агрегированное (укрупненное и огрубленное)

понятие квазиосновы (ср. побе<sup>Да</sup>  
дитель), которое «усред-  
нить  
дно

ляет», в частности, категориальную семантику, не проводя различий между частями речи. Для других целей, например, при построении алгоритма автоматического морфологического анализа и синтеза эстонского языка, задача распознавания частей речи может стать одной из главных, а усреднение проводится по другим признакам<sup>23</sup>. К идеи усреднения (правда, без использования этого термина) все чаще обращаются сами словарники. Так, К. С. Горбачевич, рассматривая пути экономизации иллюстративного материала к дефиниции, обосновывает необходимость более широкого использования типовых речений и фраз (т. е. «усредненных» образцов). — Ю. К.) вместо беллетристических цитат<sup>24</sup>.

Эвристическая значимость принципа усреднения выходит, конечно, за рамки лингвистики, а тем более лексикографии. Являясь широко распространенным в технических и естественных дисциплинах, прием этот находит использование и в науках гуманитарных, например, при создании обобщенного фото- или рисованного портрета с целью установления антропологического сходства при анализе этнических и расовых групп. Сущность этого биометрического метода состоит в конструировании из снимков (портретов) отдельных людей некоторого типового лица группы, воссоздании внешности, так сказать, «среднего человека» — представителя одной семьи на протяжении ее разных поколений, или замкнутой локальной группы населения, или населения большого региона и т. п. (например, западного грузина, восточного славянина, каракалпака)<sup>25</sup>. Оказывается, обобщенный портрет обладает высокой научной информативностью: хотя он абстрактен, но вполне представителен, узнаваем и, не будучи схож

<sup>23</sup> Vicks Ü. Klassifikatoorse morfoloogia põhimõtted. Preprint KKI-9. Tallinn, 1977.

<sup>24</sup> Горбачевич К. С. Словарь и цитата (о рационализации иллюстрирования слов и значений во втором издании семнадцатитомного словаря). — ВЯ, 1978, № 5.

<sup>25</sup> Абдушишвили М. Г., Павловский О. М. Интегрирование схемографического и фотографического методов обобщения изображения лица и использование полученного портрета в качестве источника антропологической информации. — Советская этнография, 1979, № 1,

с внешностью ни одного из конкретных лиц, вошедших в выборку (поскольку на деле такое обобщенное «среднее лицо» и не существует в реальности), он в то же время типичен, соотносим с обликом любого представителя данной группы.

В филологии к приему усреднения данных постоянно обращаются диалектологи<sup>26</sup>, он необходим также в практике обучения иностранным языкам при определении общих контуров и важнейших параметров преподаваемого варианта языка, к нему прибегают, наконец, в эдиционной практике, обобщая, например, варианты текста разных исполнителей при популяризации произведений народного творчества (былин, исторических песен, сказок).

В лингвистическом конструировании с принципом усреднения мы встречаемся, помимо приведенных выше случаев, также при кодировании семантической информации о слове (см. ниже, § 6), которое осуществляется без учета числа его значений, выделяемых тем или иным словарем, а только на основе используемых в толковании понятий. Так, если два первых значения слова *холод* поданы в словаре следующим образом: 1. Низкая температура воздуха. 2. Воздух низкой температуры, а также погода с низкой температурой воздуха, — то в семантической записи мы имеем всего четыре множителя: воздух низ темпер погод. Одновременно такое усреднение позволяет поставить и вопрос о правомерности включения в запись, или о дифференциальной полноценности, множителя погод, т. е. о целесообразности разграничения двух значений. Такое разграничение мотивируется, как кажется, лишь наличием сочетания *зимние холода* для второго, где определяемое слово допускает форму множественного числа, тогда как в остальных случаях всегда выступает в единственном. Но это уникальное сочетание вполне могло бы быть квалифицировано как идиоматическое, и, таким образом, число значений уменьшилось, а описание упростилось бы.

### ЛИНГВИСТИЧЕСКОЕ КОНСТРУИРОВАНИЕ, ЯЗЫКОВАЯ ТЕОРИЯ, ИССЛЕДОВАТЕЛЬСКАЯ ПРАКТИКА

Естественно, возникает вопрос о соотношении лингвистического конструирования с эмпирическими обобщениями: в самом деле, словари и грамматики создавались

всегда, но при чем здесь лингвистическое конструирование? Вопрос вполне закономерен, но нужно помнить, что («логарифмы и грамматики традиционно строятся на эмпирической и общем-то основе. Как справедливо отмечает Г. Вариг, кто только теоретикам представляется, что применение и конкретизация абстрактных положений теории, зачастую еще затуманенных специфической терминологией, требует значительно большего труда, чем решение тех же вопросов с помощью известных уже практических навыков, обычай, традиций<sup>27</sup>. Приемы лингвистического конструирования и позволяют ликвидировать этот разрыв, поскольку направлены они, в частности, и на развитие, улучшение, совершенствование отстоявшихся практических навыков, но на базе современной лингвистической теории. Простой пример: известно, что толковые словари созданы не для всех языков народов нашей страны, и это обстоятельство объясняется определенной консервативностью исторически сложившейся методики их составления. Хотя «литературный язык» и «язык литературы» не считаются тождественными, но картотека словаря литературного языка традиционно составляется на основе расписывания произведений (т. е. языка) литературы, и если у малого народа, скажем, у хантов, нет достаточно богатой художественной литературы, то как будто уже и нет возможности составить толковый словарь языка. Однако, как было показано в докладе автора этих строк на IV Международном конгрессе финно-угроведов, с помощью лингвистического конструирования можно построить методику, которая позволяет непосредственно от полевых записей переходить к словарным дефинициям толкового словаря, опираясь на некоторые положения и приемы, используемые в теории семантического поля.

Современное состояние науки о языке характеризуется парадоксальным фактом: сегодня написать теоретическую работу, выдвинуть новые идеи, даже создать целую новую теорию легче (заметим еще — и престижнее), чем внед-

---

<sup>26</sup> Скрепина Л. М. Диалектные данные в истории языка и метод «усреднения». — В кн.: Романско-германское языкознание, вып. 6. Минск, 1976, с. 207—210.

<sup>27</sup> Wahrig G. Anleitung zur grammatisch-semantischen Beschreibung lexikalischer Einheiten. Versuch eines Modells. Tübingen, 1973, S. XV—XVI.

рить уже имеющиеся идеи, воплотить готовые теории. Можно даже сказать, что в языкоznании сегодня наблюдается перепроизводство идей: есть порождающая теория, но нет ни порождающей, ни трансформационной грамматики какого-то конкретного языка, есть теория тезаурусов, но нет тезауруса хотя бы одного из литературных языков народов нашей страны; более того, не для всех языков, как отмечено выше, созданы и толковые словари. Таким образом, при наличии известной нужды в таких лингвистических объектах, воплощающих теорию, как грамматика нового типа или словарь нового типа, в сегодняшнем языкоznании тем не менее стихийно складывается тенденция к преобладанию теоретических работ, т. е. к созданию избыточного числа фактических описаний, но более высокого уровня, так сказать, второго порядка, в которых обычно только намечаются новые подходы, иллюстрируются некоторые возможности выдвигаемой идеи, указывается направление, в котором надо идти, чтобы получить определенный результат. Авторы учат, как надо делать, но не делают сами, поэтому результата как такового не бывает, новая вещь в итоге такого исследования не получается, создается лишь, как замечает остроумно один известный лингвист, «информационный шум».

На пути расширения лингвистических работ практической ориентации возникает ряд трудностей. Прежде всего, внедрение идеи всегда оказывается делом значительно более трудоемким, а часто требующим и большей научной смелости, чем выдвижение идеи и демонстрация ее возможностей. Поэтому авторов, отдающих предпочтение исследованиям теоретического характера, можно понять: ведь в конце экспериментального пути, нацеленного на внедрение идеи, на построение некоторого объекта, всегда ждет вопрос — «выйдет или не выйдет?» Ну, а описание большей или меньшей степени теоретизированности получится в любом случае. Издание наших книг способствует утверждению иллюзии, что они-то и есть если не единственный, то основной практический результат.

Однако подобный результат находит применение лишь в работах небольшого числа родственных специалистов, причем иногда его использование не идет дальше формальных ссылок в библиографии. Круг, так сказать, замыкается. Это работа на себя, работа, направленная на самосовершенствование, саморазвитие, своего рода лингвистический культуризм. Конечно же, без этого не может су-

ществовать и двигаться вперёд никакая наука. И под внедрением наряду с обычным его смыслом — непосредственным использованием в лингвистической, преподавательской, общественно-культурной практике следует понимать также «опосредованное использование», т. е. внедрение новых теоретических положений в самое лингвистику, введение новых идей в научный обиход, в систему лингвистических понятий, в систему наших знаний о языке вообще. Речь идет только о том, что подобная направленность исследований не должна становиться, как это бывает, к сожалению, в сознании самих ученых, единственной целью и самоцелью лингвистики. Нам надо бы помнить, что решение далеко не всякой проблемы увеличивает наши знания, расширяет круг представлений о языке вообще, а потом ведь и не все знания оказываются полезными и нужными. В «Путешествиях Гулливера», рассказывая об элите Лапуты — математиках, музыкантах, мудрецах, Свифт сообщает читателям, что у них были слуги, которые хлопали их время от времени бычьими пузырями, чтобы отвлечь от пустых размышлений и напомнить о сложности повседневной жизни. «Возможно, — говорит в этой связи Дж. Пассмор, — нам тоже нужны такие слуги в наших самых высоких научных советах»<sup>28</sup>.

Что касается трудоемкости непосредственного внедрения, т. е. трудоемкости работы по созданию новых лингвистических объектов, то в наш век коллективного творчества, всесоюзной и международной кооперации и координации исследований, это препятствие на пути расширения практических выходов науки о языке, казалось бы, легко преодолимо. Однако, как показывает опыт такой кооперации и координации, объединение усилий ученых — от коллектива, работающего над какой-то темой в пределах одного института, до международного сотрудничества целых учреждений — приводит пока только к экстенсивному росту, увеличению масштабов, так сказать, обрабатываемых площадей. Это, безусловно, тоже достижение, которое позволило в последнее время осуществить такие грандиозные предприятия, как многие республиканские диалектологические атласы, коллективный труд всесоюзного масштаба «Языки народов СССР» или исследование международного характера «Общеславянский лингвисти-

---

<sup>28</sup> Passmore J. Science and its critics. New Brunswick, Rutgers University press, 1978, p. 76.

ческий атлас», и наметить еще более крупные темы ~ «Лингвистический атлас Европы» и энциклопедию «Языки мира». Однако особенность такого рода кооперации состоит в том, что соотношение соисполнителей при разработке общей темы (или проблемы) строится на принципе «и . . и . . .», на принципе простого сложения результатов. Но сегодня этого уже недостаточно, надо идти дальше: реализация целевых программ требует усиления «аддитивного» эффекта в нашей науке, т. е. установления иного отношения между соисполнителями — по принципу «если . . , то», когда результаты, полученные одними исполнителями, служат исходным пунктом для других, используются другими исполнителями. Иными словами, встает необходимость разделения труда не «по площадям», а по проблемам внутри одной специализированной программы. Следовательно, в пересмотре нуждаются не только основы планирования, но и сами способы кооперации и координирования научных работ.

Прогнозирование развития науки о языке, в том числе и долгосрочное планирование, всегда осуществлялось либо по проблемам, а значит не предполагало какого-то практического выхода за исключением учебников, либо по темам, а значит — «по площадям», когда практическим результатом должна стать сама проработка темы, некоторое описание, книга, серия, подразумевающая, естественно, рассмотрение и решение входящих в тему проблем. Планы фактически никогда не включали целей. Отсутствие целевой ориентации приводило, в частности, к тому, что материал, всегда имеющий тенденцию расти до бесконечности, в ряде случаев начинал господствовать над исследователем, уводил от основной идеи. Определяющим становился ложный пафос полноты описания, которая на деле никогда не может быть исчерпывающей, и в итоге важные и нужные работы растягивались иногда на десятилетия. Цели же, как было сказано выше, не только предполагают получение конкретных результатов, создание некоторых новых лингвистических объектов, но и стимулируют развитие теории. Недаром П. Л. Капица подчеркивает: «История науки показывает, что даже большое творчество всегда было направленным, и его направление можно заранее определить как по запросам мировой науки, так и по запросам культуры страны» <sup>29</sup>.

Таким образом, к лингвистическому конструированию как одному из возможных путей преодоления замкнутости

теоретических работ и расширения практических выходов науки о языке нас приводят два фактора. Во-первых, это — состояние самой лингвистики, и здесь лингвистическое конструирование выступает как внутренний ресурс науки, направленный на преодоление ее собственных трудностей и противоречий; во-вторых, — программно-целевой подход к научно-исследовательской деятельности, который должен опираться на лингвистическое конструирование как на средство построения лингвистических объектов с заданными свойствами. Последняя задача — по отношению к главной линии развития науки, направленной на открытие научной истины, — выступает как некоторый паллиатив, как неточно сформулированная цель научных изысканий. Но подобный паллиатив зачастую бывает единственным результативным путем там, где прямые подступы к истине невозможны<sup>29</sup>.

Цель создания лингвистических объектов с заданными свойствами, будучи всегда сквозной для языка, вовлекает в действие все его уровни, требует практического решения большого числа проблем, каждую из которых надо либо рассматривать и решать заново, либо эвристически использовать уже имеющиеся, готовые теоретические и практические образцы решений из других отраслей языкоznания. Вот здесь-то и приходит на помощь лингвистическое конструирование. Опираясь на весь арсенал лингвистических знаний, приемов, методов, «образцов», необычным образом категоризуя исследуемую область, используя принцип оборачивания метода, вводя новые единицы и их комбинации, прибегая к статистическому усреднению некоторых важных параметров, лингвистическое конструирование позволяет преодолеть объективные трудности, связанные с современным состоянием науки о языке, с неразработанностью отдельных проблем, т. е. принимать конструктивные, технические решения там, где теория оказывается бессильной.

Естественно встает вопрос и об отношениях лингвистического конструирования с теорией. Из предыдущего ясно, что это прежде всего отношение взаимодополнительности: принципы лингвистического конструирования работают там, где общая теория еще не создана или принципиально

---

<sup>29</sup> Капица П. Л. Указ. соч., с. 85.

<sup>30</sup> Ботвинник М. М. О кибернетической цели игры. М., 1975, с. 33—39.

не может быть создана (как, например, в сфере научно-технической терминологии). Теория, отправляясь от суммы ранее полученных результатов, ищет закономерности. Лингвистическое конструирование, отправляясь от конечного продукта (который необходимо получить), находит решения. В этом плане, казалось бы, можно сказать, что лингвистическое конструирование идет в русле тенденции «модель вместо закона»<sup>31</sup>, но результатом лингвистического конструирования должен быть реальный объект, «новый лингвистический объект», а не его модель. То есть лингвистическое конструирование дает не только качественное (модель), но и количественное решение задачи.

Не хотелось бы все же, чтобы у читателя в итоге создавалось впечатление, что «цели» в языкоznании должны превалировать над проблемами. Смысл развивающей здесь концепции в том, что и «цели» должны иметь место. Вопрос о повышении научного престижа лингвистики на фоне естественных и технических наук, о расширении ее практического влияния на жизнь общества волнует языковедов всего мира. В литературе последних лет живо обсуждаются конкретные задачи, решение которых может поднять общественную значимость науки о языке<sup>32</sup>, намечаются пути превращения лингвистики в «освободительную» (emancipatorisch) науку<sup>33</sup>, всерьез делается попытка ответить на вопрос, может ли лингвист получить Нобелевскую премию, и если да, то за какого рода работу<sup>34</sup>.

Но при всем этом не надо забывать, что лингвистика — наука гуманитарная, более того — филологическая. А общественно-полезный эффект от филологии не всегда может проявиться по принципу «мне — здесь — сейчас», его характер кумулятивный по преимуществу. Кроме того, у лингвистики есть и другие, не менее важные задачи, которые она выполняла и будет выполнять всегда. Мы называем лингвистику наукой исторической не только потому, что она рассматривает свой объект в историческом развитии, в непрерывном изменении. Для лингвистики, как и для других наук гуманитарного цикла, характерно особое соотношение между теорией и историей самой науки. Одна из важнейших задач лингвистической теории — быть арсеналом исторических знаний не только о своем объекте, но и о своей науке, т. е. теория здесь как бы фильтрует историю, выбирает ее в себя. Если в ряде естественнонаучных и технических дисциплин история

науки отделена от теории и в известном смысле противопоставлена ей, то для лингвистической теории всегда актуальна, постоянно в ней незримо присутствует история лингвистики. И эта задача — быть арсеналом исторических знаний, а значит — быть современным идеологическим оружием — остается для языкоznания самой важной и всегда актуальной.

Последующее изложение, как хотелось бы думать автору, представляет собой развитие и систематическое приложение принципов лингвистического конструирования к созданию «нового лингвистического объекта» — построению с помощью ЭВМ тезауруса русского литературного языка и решению возникающих при этом научных проблем.

- 
- <sup>31</sup> Налимов В. В. Теория эксперимента. М., 1971, с. 13 и след.
- <sup>32</sup> Ferguson Ch. A. Applications of linguistics. — In: The scope of American linguistics. Lisse, 1975.
- <sup>33</sup> Januschek F. Sprache als Objekt. «Sprechhandlungen» in Werbung, Kunst und Linguistik. Kronberg, 1976.
- <sup>34</sup> Di Pietro R. J. The need to be practical. — In: Linguistics at the crossroads. Padova—Illinois, 1977.

Ю. Н. Караполов

## Глава I

# ПРИНЦИПЫ ЛИНГВИСТИЧЕСКОГО КОНСТРУИРОВАНИЯ ПРИМЕНИТЕЛЬНО К СЛОВАРОСТРОЕНИЮ

---

## § 1. ТЕНДЕНЦИЯ К ЛЕКСИКОГРАФИЧЕСКОЙ ПАРАМЕТРИЗАЦИИ ЯЗЫКА

Расчлените каждую изучаемую вами задачу на столько частей, на сколько сможете и на сколько это потребуется вам, чтобы их было легко решить.

*P. Декарт. Рассуждение о методе*

### Общая характеристика тенденции

В теории и практике составления словарей сегодня можно выделить четыре относительно самостоятельные отрасли:

- терминологическая,
- вычислительная,
- учебная и
- общая лексикография.

Различия между ними определяются прежде всего предназначенностю описания и, следовательно, специфичностью самого описываемого массива лексики. Это свойство каждой группы словарей очевидно, оно задано в самом названии отрасли и не требует специальных пояснений. Наряду с признаком «для чего делается» для нынешнего уровня словарного дела существенным оказывается признак «как делается». В области терминологической отечественной лексикографии все большее распространение получает автоматическая обработка данных и машинное составление словарей<sup>1</sup>. Вычислительная лексикография на современном этапе характеризуется ручным способом составления словарей естественных языков для использования их в системах автоматической обработки текстов<sup>2</sup>. Наша учебная лексикография, сделавшая в последние годы серьезные шаги как в теории<sup>3</sup>, так и в практике создания словарей различных типов, пользуется пока традиционными приемами составления картотек и перехода от них к словарю. Наконец, для обширнейшей области

общей лексикографии характерны пока только первые шаги по пути автоматизации подготовки простейших типов слоила рей обратных и орфографических<sup>4</sup>, основная же масса словарных работ здесь осуществляется вручную.

Еще одно различие четырех названных отраслей, характерное для их нынешнего состояния<sup>5</sup>, связано с наличием в каждой из них стремления и реальных предпосылок к созданию универсального (многоцелевого) словаря данной отрасли. Как представляется, наиболее продвинулась по пути решения этой проблемы терминологическая лексикография. От опыта создания узкоспециальных информационно-поисковых тезаурусов<sup>6</sup> здесь осуществляется движение к построению политехнических (политематических)<sup>7</sup>, а затем — целой системы научно-технических тезаурусов с универсальной структурой<sup>8</sup>. Однако на этом пути серьезным препятствием оказывается теперь отсутствие общеязыкового тезауруса, фиксирующего самые общие смысловые связи слов естественного языка<sup>9</sup>.

<sup>1</sup> Зиман Ю.Л., Егорова Е. В. Использование ЦВМ для автоматизации составления и редактирования словарей. М., 1977.

<sup>2</sup> Марчук Ю. Н. Вычислительная лексикография. М., 1970.

<sup>3</sup> См., напр.: Вопросы учебной лексикографии. М., 1969; Проблемы учебной лексикографии. М., 1977; Актуальные проблемы учебной лексикографии. М., 1977.

<sup>4</sup> См., напр.: Обратный словарь русского языка. М., 1974; *Viks Ü. Õigekeelsussõnaraamatü arvutivariant. Verbid. — Keel ja kirjandus*, 1974, № 10.

<sup>5</sup> Данное различие, как, впрочем, и предыдущее, носит преходящий характер и по мере развития науки, естественно, должно потерять свою релевантность.

<sup>6</sup> См., напр.: Тезаурус по автомобилестроению. М., 1971; Информационно-поисковый тезаурус по информатике. М., 1973; Герд А. С. и др. Тезаурус информационно-поисковый по сталям. Л., 1975.

<sup>7</sup> См., напр.: Тезаурус научно-технических терминов. М., 1972.

<sup>8</sup> Соколов А. В. Об одном возможном подходе к обеспечению совместимости ИПТ (универсальная фабула информационно-поисковых тезаурусов). — НТИ, сер. 2, 1977, № 1.

<sup>9</sup> Опыт разработки вопросов лингвистического обеспечения автоматизированных информационных систем приводит к мысли, что принципиальный отказ от использования естественных языков в этих системах был несколько поспешным. На недостаточность применения только формальных языков в ИПС и преимущество, обусловленные контролируемым использованием «природной многозначности естественного языка», указывает ряд авторов: Арапов М. В., Шрейдер Ю. А. Интегральная система и диалог. — НТИ, сер. 2, 1977, № 1, с. 3—4; Соколов А. В. Указ. соч., с. 21; Виленская С. К. О совместимости информационно-поисковых языков (язык-посредник и язык-переключатель). — НТИ, сер. 2,

Именно эти связи имеют внетематический, т. е. универсальный, характер и образуют высший уровень в иерархии отношений в ИПС. В области вычислительной лексикографии (и машинного перевода) разработаны теоретические обоснования и сделаны определенные практические шаги по созданию своего рода универсального машинного словаря<sup>10</sup>. Однако в этих разработках вопрос об отражении в таком словаре семантических связей является, пожалуй, наименее исследованным. Рассматриваются возможности построения синтетического словаря и в сфере учебной лексикографии<sup>11</sup>. И хотя последняя в некоторых отношениях опережает общую лексикографию<sup>12</sup>, и в той и в другой областях до практического воплощения этих идей пока далеко.

Чтобы закончить беглую характеристику возможного деления современной лексикографии на сферы, следует упомянуть о большой области частотных словарей, которую иногда выделяют в качестве самостоятельной<sup>13</sup>. Это может быть и справедливо, но с позиции рассмотренных выше признаков вряд ли целесообразно, поскольку все частотные словари создаются во вспомогательных целях, а кроме того, по отношению к ним не имеет смысла говорить об универсальном "словаре". Очевидно, статистическую лексикографию в целом можно рассматривать как часть общей. То же относится и к такой специфической группе словарей, как словари писателей. Что касается словарей энциклопедического характера, то и по типу дефиниций, и по предназначенности их следует отнести скорее к сфере терминологической лексикографии.

Универсальный словарь, будучи ориентирован на решение специфических задач той лексикографической отрасли, к которой он относится, должен тем не менее обладать рядом общих для всех отраслей свойств. Прежде всего, он должен быть словарем тезаурусного типа. Представление о тезаурусе, его структуре, задачах и возможностях к настоящему моменту определилось в науке довольно однозначно (подробнее см. ниже, § 4), хотя иногда и имеют место попытки расширительного толкования его: производится замена основных его единиц — слов и словосочетаний — суждениями и текстами<sup>14</sup>, включение в тезаурус грамматической информации<sup>15</sup> и даже фактическое приравнивание тезауруса к информационно-поисковому языку в целом<sup>16</sup>. Эти попытки, как и постановка вопроса о создании универсального словаря,

т. о. словаря, способного вмѣстить максимум информаціи данной отрасли, имеют под собой определенныя основания, связанные и с уровнем развития мировой лексикографической науки, и с той тенденцией, которая наметилась п ней в последние годы особенно отчетливо. Речь идет о явлении, которое я называю лексикографической параметризацией языка и к характеристику которого теперь перехожу.

Нынешнее «словаростроение» характеризуется резко возросшим потоком всевозможных словарей и чрезвычайным увеличением их разнообразия. В том, что касается изучения и словарной фиксации собственно лексики, лексического состава языка достигнута исключительная детальность и высокая степень полноты. В словарях нашла отражение социальная (словари литературных языков<sup>17</sup>, жаргонов<sup>18</sup>, табуированных слов<sup>19</sup>, и наоборот, рекомен-

- 
- 1977, № 5, с. 19; *Маршакова И. В.* Построение информационно-поискового тезауруса методом дистрибутивно-статистического анализа. — Там же, с. 11.
- <sup>10</sup> *Шаляпина З. М.* Англо-русский многоаспектный автоматический словарь (АРМАС). — В кн.: Машинный перевод и прикладная лингвистика, вып. 17. М., 1974; *Беляева Л. Н., Крисевич В. С., Липницкий С. Ф., Пионтровский Р. Г.* О многоцелевом автоматическом словаре русского языка (МАРС). — В кн.: Вопросы общей и прикладной лингвистики. Минск, 1975.
- <sup>11</sup> *Морковкин В. В.* Об универсальном учебном словаре русского языка. — В кн.: Международная ассоциация преподавателей русского языка и литературы. Конгресс 2-й. Варшава, 1973. Избранные доклады и сообщения. М., 1975, с. 365—373; *Wahrig G.* Anleitung zur grammatisch-semantischen Beschreibung lexikalischer Einheiten. Versuch eines Modells. Tübingen, 1973.
- <sup>12</sup> Именно в рамках учебной лексикографии впервые создан «Учебный словарь сочетаемости слов русского языка» (М., 1978).
- <sup>13</sup> *Алексеев П. М.* Статистическая лексикография. (Типология, составление и применение частотных словарей). Л., 1975.
- <sup>14</sup> *Сохор А. М.* Об определении семантической информации. — НТИ, сер. 2, 1977, № 1.
- <sup>15</sup> *Станчев Н. Т.* Об одном подходе к созданию дескрипторного языка с элементами грамматики. — НТИ, сер. 2, 1976, № 7.
- <sup>16</sup> Ср., напр., схему ИПЯ, разработанную в рамках BSO (Broad System of Ordering), со структурой обычного тезауруса (*Виленская С. К.* Указ. соч., с. 20).
- <sup>17</sup> Напр.: Словарь современного русского литературного языка. Т. 1—17. М.—Л., 1948—1965.
- <sup>18</sup> Напр.: *Бондалетов В. Д.* Условные языки русских ремесленников и торговцев. Вып. 1. Условные языки как особый тип социальных диалектов. Рязань, 1974; *Nentworth P., Flexner S. B.* Dictionary of American slang. N. Y., 1975.
- <sup>19</sup> *Martin J.* Diccionario de expresiones malsonantes del español. Madrid, 1974.

дуемых нормой выражений <sup>20</sup>), территориальная (от диалектного словаря индивидуума <sup>21</sup>, отдельного села <sup>22</sup> до словарей как небольших районов, так и крупных областей — смоленских, донских, псковских, старожильческих говоров Сибири — и, наконец, до сводного словаря всех диалектов <sup>23</sup>), профессиональная дифференциация языка (это не только необозримое море терминологических словарей по всем отраслям науки и техники, но и специальные профессиональные словари по отдельным подразделам тех или иных областей знаний и культуры <sup>24</sup>). Словари диахронического плана — исторические и этимологические разной глубины — позволяют проследить изменение значений во времени, установить характерные для того или иного языка линии семантического развития, типичные семантические процессы <sup>25</sup>. Наряду с такими ретроспективными словарями для языков с богатыми лексикографическими традициями характерно наличие пророспективных словарей, оперативно фиксирующих появление в языке новых слов и значений <sup>26</sup>. Системные отношения в лексике закрепляются в разнообразнейших словарях синонимов, антонимов, паронимов (в том числе и межъязыковых), омонимов <sup>27</sup>, а также в тематических, ассоциативных и идеологических <sup>28</sup>. Таким образом, лексический уровень языка представлен в словарях с достаточной полнотой и точностью.

Но возможности лексикографии, простота и исчерпываемость словарного описания, удобство пользования им обусловили стремление в современной описательной лингвистике к словарному представлению результатов исследования и всех других, помимо лексического, уровней языковой структуры, к «ословариванию» языкового описания. На фонетико-графемном уровне, кроме широко известных прямых и обратных орфографических словарей, значительное распространение получили орфоэпические <sup>29</sup> и словари ударений <sup>30</sup>. В принципе на этом уровне языковой структуры возможным представляется и построение интонационных словарей, или словарей интонем, которые могут иметь смысл не только для тоновых, но, очевидно, и для любых языков. На уровне морфологии лексикография представлена словарями служебных морфем, корней, основ, гнезд, словообразовательных элементов <sup>31</sup>, словоизменительными по частям речи <sup>32</sup> и, наконец, грамматическими, т. е. охватывающими словоизменение всех зна-

- 20 Напр.: *Граудингл. К., Ицкович А., Катлинская. П.* Грамматическая правильность русской речи. Опыт стилистического словаря вариантов. М., 1976.
- 21 *Тимофеев В. П.* Диалектный словарь личности. Шадринск, 1971.
- 22 Словарь современного русского народного говора (д. Деулино Рязанского района Рязанской области). М., 1969.
- 23 Словарь русских народных говоров. Вып. 1—14. Л., 1965—1978.
- 24 Напр.: *Граллерт В., Грушке В.* Филателистический словарь. М., 1977; *Васильев Г. М.* Русско-якутский поэтический словарь. Якутск, 1976; *Иванов В. В., Палькин В. М., Филиппов А. В., Шанский Н. М.* Краткий словарь традиционных символов русской поэзии. — Русский язык в школе, 1977, № 4—5.
- 25 Словарь русского языка XI—XVII вв. Вып. 1—4. М., 1975—1977; Словарь русского языка XVIII века. Проект. Л., 1977; Erdélyi magyar szótörténeti társ. I. Bukarest, 1975; Schröer M. M. A. *Englisches Handwörterbuch in genetischer Darstellung auf Grund der Etymologien und Bedeutungsentwicklungen, rait phonetischer Aussprachebezeichnung und Berücksichtigung des Amerikanischen und der Eigennamen*. Heidelberg, 1970; Этимологический словарь русского языка, т. I, вып. 1—5. М., 1963—1973; т. II, вып. 6. М., 1975; Этимологический словарь славянских языков. Нраславянский лексический фонд. Вып. 1—6. М., 1974—1979; Suomen kielen etymologinen sanakirja, т. I—V. Helsinki, 1955—1975.
- 26 Новые слова и значения. Словарь-справочник по материалам прессы и литературы 60-х годов. М., 1971; Gilbert P. *Dictionnaire des mots nouveaux*. Paris, 1971.
- 27 Словарь синонимов русского языка, т. 1—2. Л., 1970—1971; Boussinot R. *Dictionnaire des synonymes, analogies et antonymes*. Paris, 1973; Клышка М. К. Слóўші сінонімаў і близказначычаў слоў. Минск, 1976; Колесников И. Л. Словарь паронимов. Тбилиси, 1971; Гомлиб К. Г. М. Немецко-русский и русско-немецкий словарь «ложных друзей переводчика». М., 1972; Ахматова О. С. Словарь омонимов русского языка. М., 1974.
- 28 Русско-узбекский тематический словарь. Ташкент, 1975; Букэ М., Ессеев И. Русско-румынский тематический словарь! Тимишоара, 1975; Casares J. *Diccionario ideológico de la lengua española*. Barcelona, 1951; Титова Л. Н. Киргизско-русский ассоциативный словарь. Фруззе, 1975; Словарь ассоциативных норм русского языка. Под ред. А. А. Леонтьева. М., 1977.
- 29 Русское литературное произношение и ударение. Словарь-справочник. М., 1959; Martinet A., Walter P. *Dictionnaire de la prononciation française dans son usage réel*. Paris, 1973; Wörterbuch der deutschen Aussprache. Leipzig, 1974.
- 30 Словарь ударений для работников радио и телевидения. М., 1967.
- 31 Беляева Й. Н. Алгоритмизация составления словаря основ. — В кн.: XXVII Герценовские чтения. Иностранные языки. Л., 1974, с. 140—143; Словарь-справочник по словообразованию! Алма-Ата, 1975; Тихонов А. Н. Школьный словообразовательный словарь русского языка. М., 1978; Словарь словообразовательных элементов немецкого языка. Под рук. М. Д. Степановой. М., 1979; ср. также идею создания «словаря морфемного членения», «толкового словаря морфем» и «идеографического морфологического словаря»: Милославский И. Г. Вопросы словообразовательного синтеза. М., 1980.
- 32 Напр.: Daum E., Schenk W. *Die Russischen Verben*. Leipzig, 1966,

менательных частей речи<sup>33</sup>. Для полноты морфологической картины этот перечень следовало бы дополнить пока не существующими, но мыслимыми словарями морфонологических синхронических чередований (или, что то же, — словарями диахронических фонетических законов) и словарем грамматических категорий (очевидно — всех языков мира). Распространение лексикографических приемов описания на область синтаксиса шло от фиксации крупных устойчивых сочетаний типа пословиц, афоризмов, фразеологизмов и речевых штампов<sup>34</sup> — к составлению словарей синтаксической сочетаемости<sup>35</sup>, а затем и постановке вопроса о создании синтаксического словаря вообще<sup>36</sup>. Таким образом, в словарях зафиксированы практически все уровни языковой структуры, все языковые единицы, причем не только в статике, но и в динамике — эволюционной (диахронические словари) и актуальной (словари употреблений и вариативности).

Словарь, обобщая наши знания о том или ином аспекте языковой структуры, позволяет представить их в наиболее удобном для использования в определенных целях виде. Так, осознание необходимости для целей машинного перевода составления разнообразных словарей — в качестве альтернативы попыткам построения универсального грамматического описания, ориентированного на язык-посредник, — знаменовало по сути дела качественно новый этап в развитии этой прикладной отрасли<sup>37</sup>. Созданный к настоящему времени в рамках вычислительной лингвистики большой корпус словарей для машинного перевода вносит свои особые краски в общую картину лексикографической параметризации языка. Принципиальное отличие этих словарей от обычных в том, что они зачастую используют иные, чем общепринято, единицы, или параметры, т. е. по-иному категоризуют языковую материю. Рассмотрим в качестве примера комплекс словарей, работающих в системе машинного перевода с английского языка на русский<sup>38</sup>. Всего здесь имеется семь словарей. Английский словарь первичной обработки предназначен для идентификации словоформ входного текста с основами, содержащимися в алгоритме морфологического анализа. Слова в словаре для первичной обработки расположены по убывающим длинам, а внутри групп одной длины — по алфавиту. Таким образом, важным параметром здесь оказывается длина слова. Словарь оборотов с переводами служит для идентификации устойчивых словосочетаний

в тексте. Затем в работу вступает словарь омонимов и алгоритм разрешения омографии. Однозначные и многозначные английские слова сгруппированы в два разных словаря: перевод однозначных осуществляется по таблице соответствий, при переводе многозначных слов используются специальные алгоритмы поиска контекстных признаков, поэтому последний словарь называют еще контекстологическим. Словари оборотов и контекстологический являются специфическими разновидностями синтаксического словаря. В тех случаях, когда алгоритм перевода многозначного слова не дает результата, в работу вступает шестой — контрольный словарь однозначного перевода многозначных слов. И, наконец, словарь русского языка для синтеза воссоздает русскую словоформу. Свообразие категоризации языкового материала в этой системе словарей проявляется и в особом способе трактовки и разрешения полисемии<sup>39</sup>, и в специфическом представлении результатов морфологического анализа и классификации слов в выходном русском словаре<sup>40</sup>.

Содержательный смысл рассматриваемой тенденции не только в том, что в совокупности словарей могут быть

- 
- <sup>33</sup> Зализняк А. А. Грамматический словарь русского языка. Словоизменение. М., 1977.
- <sup>34</sup> Даль В. И. Пословицы русского народа. М., 1957; Абдурахманов М. Узбекско-русский словарь афоризмов. Ташкент, 1970; Фразеологический словарь русского языка. М., 1967; Herfadi M. Közhelyszótár. Budapest, 1976.
- <sup>35</sup> Деривас В. М. Устойчивые глагольно-именные словосочетания русского языка. М., 1975; Убин И. И. Словарь усилительных словосочетаний для русского и английского языков. Ч. 1. Прямой словарь. Ч. 2. Реверсивный словарь. М., 1979; Учебный словарь сочетаемости слов русского языка. М., 1978; Горбачевич К. С., Хабло Е. П. Словарь эпитетов современного русского литературного языка. М., 1979; The word finder. Ed. by J. I. Roldale. Emmaus, Pennsylvania, 1964.
- <sup>36</sup> Золотова Г. А. О принципах построения синтаксического словаря русского языка. Доклад на заседании Бюро Отделения литературы и языка АН СССР 16 января 1979 г.
- <sup>37</sup> Герасимов В. И., Марчук Ю. И. Современное состояние машинного перевода. — В кн.: Машинный перевод и автоматизация информационных процессов. М., 1975, с. 6—12.
- <sup>38</sup> Марчук Ю. И., Тихомиров Б. Д., Шербинин В. И. Система машинного перевода с английского языка на русский. — В кн.: Машинный перевод и автоматизация информационных процессов. М., 1975, с. 19—24.
- <sup>39</sup> Там же, с. 26—28.
- <sup>40</sup> Там же, с. 29—30.

отражены и в какой-то степени отражаются все уровни языковой структуры, но и в том, что к настоящему моменту в лингвистике пафос описания любого уровня или фрагмента языка видится в создании словаря его элементов, процедур, отношений. Словарь становится целью и итогом исследовательской работы над той или иной проблемой, как бы «ответом» к определенной лингвистической задаче. В связи с этим можно выдвинуть гипотезу «полноты лексикографического описания языка», предположив, что языковая структура — не только потенциально может быть, но и реально представлена в словарях полностью. Проверку этой гипотезы мы попытаемся осуществить в итоговой матрице для анализа современных словарей, после рассмотрения самого понятия лексикографического параметра, его статуса и свойств.

Словарь оказывается самой удобной формой обобщения и фиксации наших знаний не только о языке: именно в лексикографической форме созданы и создаются Шевченковская, Лермонтовская, Пушкинская энциклопедии, конкордансы по отдельным произведениям<sup>41</sup> и по всему творчеству писателя<sup>42</sup>, дающие исследователю кроме чисто лингвистической информации представление о мировоззренческой «картине мира» автора. В таких своих формах словарь уже перестает быть перечнем слов, так или иначе охарактеризованных со стороны своего содержания, а становится систематическим собранием сведений по определенному вопросу<sup>43</sup>. Этой характеристике отвечают и некоторые из упомянутых выше чисто лингвистических словарей — грамматический (словоизменительный), словообразовательный, стилистические<sup>44</sup>, синтаксический.

### Понятие лексикографического параметра и способы его выделения

В итоге общего рассмотрения тенденции к лексикографической параметризации языка можно констатировать, что первоначальный смысл, вкладываемый в понятие «словарь», значительно расширился, привычное определение словаря как произведения, трактующего значение слова, размыается. Вероятно, этим и объясняется известная недостаточность существующих словарных типологий<sup>45</sup> и не всегда оправданное стремление исследователей воспринимать появление несколько отличного от уже известных словаря как создание словаря «нового типа»<sup>46</sup>.

Вообще же содержание понятия «тип словаря» тоже остается по вполне ясным. Тип определяется по некоторой доминантной характеристике словаря — толковый, фразеологический, антонимов и т. п., которая, как правило, совпадает с его названием, т. е. существующая типология словарей — это в большой степени типология их названий. В то же время, как легко убедиться, информация, содержащаяся даже в перечисленных типах словарей, частично пересекается: так, толковый словарь включает и фразеологию, а фразеологический дает толкования фразеологизмов. Если с этой точки зрения рассмотреть всю сеть словарей одного языка, например русского, то можно увидеть, что зоны пересечения информации являются довольно значительными. В толковых словарях помимо собственно семантической широко представлена также грамматическая, фонетическая (орфоэпическая), стилистическая и частично этимологическая информация. Двуязычный словарь<sup>47</sup> снабжает читателя сведениями не только о склонении имен, спряжении глаголов и грамматических категориях (*Biigel m—s; einfahren \*I vt*, где астериск означает, что слово относится к сильному спряжению и его формы можно найти в грамматической таблице в конце словаря), но и о словообразовании и морфемном членении (*Erholungs |l ~ heim, ~ kur, — pause, — platz, ~ reise, ~ stiitle; be " ; rait " и т. п.*), об особенностях употребления данного выражения (например: *druff* разг.; *dun* диал.; *Doktor Eisenbart* шутл.; *erschinden* неодобр.,

<sup>41</sup> Словарь-каталог философских терминов в произведении В. И. Ленина «Материализм и эмпириокритицизм», вып. 1. Горький, 1970; Словарь автобиографической трилогии М. Горького, вып. 1—2. Л., 1975—1977.

<sup>42</sup> Словарь языка А. С. Пушкина, т. I—IV. М., 1956—1961; Goethe Wörterbuch, Bd. 1, Lfg. 1—10. Berlin, 1966—1978.

<sup>43</sup> Ср., напр.: ЭткундП. Кванты. Справочник концепций. М., 1977.

<sup>44</sup> Кроме названного в сноске 20 см. также: Трудности словоупотребления и варианты норм русского литературного языка. Словарь-справочник. Л., 1973; РозентальД. Э., ТеленковаМ. А. Словарь трудностей русского языка. М., 1976.

<sup>45</sup> Ср., напр., критический обзор точек зрения по этому вопросу: Денисов Н. И. Типология учебных словарей. — В кн.: Проблемы учебной лексикографии. М., 1977; см. также: Пывин А. М. К вопросу о классификации русских словарей. — ВЯ, 1978, № 1.

<sup>46</sup> Ср., напр.: Иванова Т. Ф. «Частичное гнездование» слов в учебном толковом словаре для иностранцев. — В кн.: Проблемы учебной лексикографии. М., 1977, с. 75.

<sup>47</sup> Немецко-русский словарь. Под ред. А. А. Лепинга и Н. П. Стравовой. М., 1958.

и т. п.). В однотомном толковом словаре С. И. Ожегова довольно широко представлена синтаксическая сочетаемость (ОРУЖИЕ . . . Холодное о. Стрелковое о. Артиллерийское о. Огнестрельное о. Защищаться с оружием в руках. Немеркнущая слава советского оружия. Род оружия. Печать могучее о. Сложить или положить о. Бросать о. Поднять о.)<sup>48</sup>, а в семнадцатитомном Академическом словаре вводится опыт установления и лексической сочетаемости<sup>49</sup>.

В этих условиях уместно поставить вопрос о минимальной единице языковой информации или параметре языковой структуры, находящем отражение в словаре.

Само слово «параметр» в лингвистической литературе стало теперь одним из самых распространенных и по употребительности приближается, вероятно, к термину «система». Однако смысл его не однозначен, довольно широк и не терминологичен, а скорее метафоричен. Конкретное его наполнение определяется контекстом, в котором он употребляется, но в общем случае им называют некоторую величину, характеристику, меньшую, как правило, лингвистической единицы соответствующего описания и по функции совпадающую с дифференциальным признаком<sup>50</sup>. Не думаю, что его распространение можно связывать непосредственно с появлением книги «Параметрическая лингвистика», хотя ее авторы и вкладывают в этот термин смысл, близкий к охарактеризованному нами: «Лингвистический параметр, как он понимается в этой монографии, представляет собой некоторую переменную, дифференцирующую любые две категории в языке либо (1) в функциональном плане (и тогда это параметр функционального разграничения), либо (2) в плане обозначения (и тогда это параметр манифестации). Мы избегаем термина «различительный признак», поскольку его употребление связывается с фонологическим, в крайнем случае — морфологическим, уровнем. Параметр же есть демаркационная переменная на любом уровне»<sup>51</sup>.

Использование термина «параметр» применительно к описанию и классификации лексикографической информации представляется оправданным, поскольку будучи, с одной стороны, характеристикой языка, он может находить разное выражение в словарях в зависимости от целей, которые словарь преследует, и специфики языковой структуры, которую он отражает, служа тем самым различию самим словарей.

Коль скоро статус лексикографического параметра будет определен и будет установлен набор таких параметров в словарях одного, например русского, языка, закономерно далее рассмотреть вопрос о том, насколько полно и точно эта совокупность параметров отражает языковую структуру, т. е. подтверждается ли наша гипотеза о лексикографической параметризации языка. Решение этих вопросов позволит подвести содержательную базу под словарную типологию и строить ее на основе понимания типа словаря как сочетания определенных параметров, т. е. как исчисление лексикографических возможностей. Тогда с помощью специальных приемов («морфологический ящик», «матрицы открытия» — см. здесь же ниже) мы можем конструировать новые типы словарей, отбирая из всей совокупности параметров такие их сочетания, которые будут оптимально отвечать заранее заданным целям словаря. Для конкретной задачи, рассматриваемой в настоящей работе, — задачи конструирования тезауруса русского литературного языка — решение указанных вопросов важно именно потому, что дает основания для включения в тезаурус максимального числа совместимых друг с другом параметров, т. е. для наполнения его максимально возможной языковой информацией.

В лексикографии известны попытки эмпирических поисков наилучшего сочетания максимального количества

---

<sup>48</sup> Ожегов С. И. Словарь русского языка. 2-е изд. М., 1952; 9-е изд. М., 1972.

<sup>49</sup> Котелова Н. З. Лексическая сочетаемость слова в словаре. — В кн.: Современная русская лексикография. 1976. Л., 1977.

<sup>50</sup> Ср.: «Параметр (от греч. *parametron* — отмеривающий) в математике величина, числовые значения к-рой позволяют выделить определ. элемент (напр., кривую) из множества элементов (кривых) того же рода. Напр., в ур-нии  $x^2+y^2=r^2$  величина  $r$  является П. окружности».

«Параметр (техн.) величина, характеризующая к.-л. осн. свойство процесса, явления или системы, машины, прибора (напр., электрич. сопротивление, теплоемкость, быстродействие, масса, коэф. трения и др.). П. могут быть сосредоточенными (напр., емкость электрич. конденсатора, масса подвешенного к балке груза) и распределенными в пространстве (напр., индуктивность линии электропередачи). — Советский энциклопедический словарь. М., 1979.

<sup>51</sup> Heller L. G., Macris J. Parametric linguistics. The Hague—Paris, 1967, p. 12.

параметров в словаре. Рассмотрим в качестве примера две таких попытки — словарь Г. Варига<sup>52</sup> и «Толково-комбинаторный словарь русского языка»<sup>53</sup>. Оговоримся сразу, что эмпирическими я называю эти попытки только потому, что в том и другом случае авторы не оперировали понятием параметра, а исходили из функциональной заданности, из необходимости построения словаря, с наибольшим удобством и эффективностью обеспечивающего овладение и пользование языком — в одном случае немецким, в другом — русским. На деле же в основе каждого из этих словарей лежит вполне определенная теоретическая концепция. Так, Г. Вариг, обобщая и осмысливая опыт подготовки трех изданий словаря в своей докторской диссертации, пишет: «Подобно тому, как трансформационная модель исходит из трансформ ядерного предложения, таким же образом можно подходить и к описанию лексем. Предлагаемая модель не претендует на содержательный анализ и синтез языка, а только обеспечивает метод внесения грамматических правил в выражения для отдельных словарных статей»<sup>54</sup>. В соответствии с принципами трансформационной грамматики основным критерием для описания лексемы в ее контексте и для внесения в словарную статью тех или иных словосочетаний, значений и употреблений слова является критерий синтаксический (с. 60—63). Далее Г. Вариг постулирует ряд отношений, или оппозиций, которые должны найти отражение в словарной статье и тем самым определить ее структуру (с. 85—86) и объем информации, в ней содержащейся. Это прежде всего грамматическая характеристика слова, которая выражается тремя рядами различий:

- морфо-парадигматическими (Mann — Männer : Männer; Druck — Drücke : Drücke; hängen — hing — gehangen : hängte — gehängt);
- интерлексемными, т. е. категориальными или принадлежностью к классу слов (gehen : Gehen);
- синтаксическими, т. е. принадлежностью к различным поверхностным и ядерным структурам (das Bild hing an der Wand : er hängte das Bild an die Wand).

Прагматический аспект находит выражение, по мнению автора, в двух рядах оппозиций:

- имманентных языковой системе и используемых для обозначения ситуации коммуникации (ich : du; ich — du : er; Affirmation : Negation);

— характеризующих намерение говорящего или акцентирующих внимание слушающего (Mann — Männer *немаркированная*: Mannen *поэт.* или *ирон.*).

Затем вводятся признаки внеязыковых ситуаций: географические или диалектные (Bickbeere *сев.* : Heidelbeere *немарк.*), социальные (besoffen *груб.* : betrunken *немарк.*) и материальные или денотативные различия (Schloß : Schloß, т. е. замок и замок). Наконец, последний аспект «семантической модели» словарного описания лексемы касается профессиональной дифференциации лексики (den Blinddarm herausnehmen *немарк.* и фактически неверно: dem Wurmfortsatz des Blinddarms entfernen *мед.*).

Эти теоретические принципы находят своеобразное воплощение в организации словарной дефиниции. Авторы отказываются от обоих традиционных путей упорядочения словарных значений — • этимологического (или исторического), когда на первом месте располагается наиболее древнее значение, а остальные — по мере их появления и фиксации в литературном языке, и семантико-деривационного, согласно которому различают главное (ближайшее) значение и значения второстепенные ( дальнейшие). В качестве принципа разграничения значений здесь принято описание смысла слова сначала в изолированном употреблении, т. е. без контекста («нулевая позиция», например «глаз» как орган зрения), а затем в наборе контекстов:

- в сочетании с существительными — das Auge des Gesetzes . . .
- с глаголами — die Augen anstrengen . . .
- с прилагательными и наречиями . . .
- с частицами (наречиями, предлогами, союзами) . . .
- с вопросительными словами . . .
- в пассиве . . .
- как participle презенса или перфекта и т. д.

При этом объяснения значений каждого (типа) сочетания максимально упрощены и «ни в какой мере не заменяют специальной научной дефиниции» (с. 24—25 «Введения к словарю»). Иными словами, в рассматриваемом словаре максимально усилено свойство «комплементарности», ко-

<sup>52</sup> Wahrig G. Das grosse deutsche Wörterbuch. Gütersloh, 19GG (2-е изд. — 1968; 3-е — 1972).

<sup>53</sup> См. препринты «Материалы к Толково-комбинаторному словарю русского языка», вып. 23, 29, 80 и др.

<sup>54</sup> Wahrig G. Anleitung . . ., S. XV—XVI.

торым характеризуются обычные толковые словари /в противовес словарям для машины<sup>55</sup>/ и которое основывается на предпосылке наличия у пользователя «фоновых знаний», обязательной языковой и экстралингвистической информации большего или меньшего объема. За счет сжатия, сокращения семантической и энциклопедической информации в словаре Варига максимально представлена синтаксический аспект значения: круг приводимых в словарной статье контекстов охватывает и устойчивые сочетания, и модели свободных сочетаний. Это избавляет составителей как от систематического использования иллюстраций, поскольку широкий набор образцов употреблений представлен в самой статье, так и от обращения к синонимам. Если, например, в американских словарях Уэбстера определение слова дано как бы трижды — объяснительным текстом дефиниции, синонимами и иллюстративным материалом, то в словаре Варига почти все значения и оттенки задаются своеобразным исчислением контекстов.

Эта особенность рассматриваемого словаря роднит его с упоминавшимся выше словарем сочетаемости слов русского языка. Однако от последнего он отличается исключительной полнотой сообщаемой грамматической информации и вообще стремлением к универсальности. Во введении автор пишет, что составители ставили своей целью дать в руки пользователя источник, где он может почерпнуть сведения обо всех аспектах немецкого языка, подобно Petit Larousse для французского или Concise Oxford Dictionary и Webster's Collegiate Dictionary для английского<sup>56</sup>. Универсальным, или во всяком случае претендующим на универсальный характер, следует называть, очевидно, такой словарь, в котором в равной мере, на достаточную и одинаковую глубину эксплицируются все параметры данной языковой структуры<sup>57</sup>. Словарь Варига приближается к такому качеству, но, как и вообще ни один словарь в мировой лексикографической практике, не воплощает его полностью. Здесь находят отражение такие параметры: орфография; слогоделение; ударение; произношение (этот параметр представлен не на полную глубину: у Варига зафиксировано с помощью Международной фонетической транскрипции произношение звуков и их сочетаний только для тех случаев, которые являются отклонениями от правил немецкой орфоэпии. Ср., например, этот параметр в словарях Уэбстера<sup>58</sup>, где под-

робнейшим образом представлено произношение всего слова, правила чтения каждой буквы, а также и частичное произношение — особые случаи произнесения частей слова в словоизменительных и словообразовательных формах). Далее — грамматические параметры у Варига: для существительных их, как правило, два — род, который может принимать одно из трех значений (т., ф., п.) и одновременно является показателем категориальной принадлежности слова, и номер образца склонения, например *Schütze* <m. 17>. Последний параметр тоже принимает несколько значений: помимо номера таблицы склонения, его значением может быть либо указание на формы Gen. и PL для слов, не вошедших ни в одну из таблиц, либо указание на отсутствие множественного числа — unz., либо, наоборот, — nur PL. Два параметра характеризуют здесь и глагол: первый принимает одно из трех значений (V. t., V. i. или V. refl.), второй либо указывает номер таблицы спряжения, если глагол неправильный, либо принимает нулевое значение — в случае слабого глагола. Параметр глагольного управления в силу отмеченных выше особенностей построения словарной статьи включается здесь непосредственно в дефиницию. Для других частей речи важен один параметр — указание на категориальную принадлежность — <Adj.>, <Adv.> и т. п.

Этимологические параметры разработаны в словаре Варига настолько глубоко, что в этом отношении он соизмерим со специальными этимологическими и историческими словарями: в словарной статье прослеживается развитие немецкого слова, т. е. указываются исторические формы, зафиксированные в ходе эволюции немецкого языка, отмечается изменение значений, приводятся его «родственники» — параллели в других германских и indoевропейских языках (*bauen*. . . <*ahd. buan* «wohnen, bewohnen. . .», *got. bauan* <*idg. \*bhu* «sein, werden, entstehen. . .»); широко используется также обращение к классическим языкам и языкам романской семьи. В словарях

---

<sup>55</sup> Марчук Ю. Н. Указ. соч.

<sup>56</sup> Wahrig G. Das grosse deutsche Wörterbuch, S. 9—10.

<sup>57</sup> Речь должна идти в каждом случае именно о конкретном языке, потому что отсутствие в немецком словаре указания на вид глагола или в английском — на род существительного нельзя считать минусом: параметр вида не свойствен немецкому глаголу, так же как параметр рода — английскому имени.

<sup>58</sup> Webster's seventh new collegiate dictionary. (Ed. in chief P. B. Gove). Springfield, Mass., 1963.

Уэбстера, где этимология традиционно имеет большой вес и определяет по сути дела структуру всей словарной статьи, этот параметр тем не менее разработан не столь подробно, как у Варига.

Аналогичным образом очень скрупулезно отражены в рассматриваемом словаре стилистические параметры: по сфере употребления — *<umg., vulg., poet.>*, в том числе и ошибочное употребление, т. е. смешение стилей *<fälschl.>*; ареальный параметр; детализированно представлены параметр арго *<Kaufmannsspr., Schülerspr., Studentenspr., Jägerspr. . . .>* и параметр профессионализмов и терминов (*Fachsprachen*) — *<Arch., Math., Gramm. . . .>*.

Однако, если продолжить сравнение со словарями Уэбстера, то следует отметить у Варига отсутствие системных параметров: синонимы, которые занимают важное место в словарной статье американских словарей, здесь почти не представлены. Слабо разработан также словообразовательный параметр.

Статьи «Толково-комбинаторного словаря русского языка» (см. сноску 53) имеют определенное сходство с рассмотренным словарем прежде всего в том отношении, что в обоих случаях основу их составляет своеобразное «сингтагматическое поле»: у Варига оно задается набором контекстов, в «Толково-комбинаторном словаре» — исчислением так называемых «лексических функций»<sup>59</sup>. Перечень параметров ТКС в основном совпадает с их набором в словарях, обнаруживающих тенденцию к универсальности: орфография, ударение, управление, лексические функции (специфические для данного словаря, раскрывающие весь спектр синтаксической и лексической сочетаемости описываемой единицы), семантический эквивалент (собственно дефиниция), фразеология, иллюстрации. Из наиболее распространенных параметров здесь совершенно не представлены два — этимология и произношение. Зато исключительное развитие получают словообразовательный, системный и иллюстративный параметры. Первый из них раскрывается как достаточно полное словообразовательное гнездо, каждый член которого характеризуется своим набором лексических функций: *стыдиться — стыдно, стыд, устыдиться, постыдный, стыдный, стыдливый, бесстыдный, застыдиться, стыдить*<sup>60</sup>. Второй сведен по объему с семантическим полем и так же, как и сочетаемость, разрабатывается с помощью особых лексических функций: *Syn* — синоним, *Anti* — антоним, *Ge-*

ner — родовое понятие, типы «синтаксического словообразования», инструмент (способ, средство, место, результат), Sing. — элемент, Mult — совокупность и некоторых других. По разработанности системных лексико-семантических отношений ТКС приближается к идеографическим словарям. Что касается стилистических параметров, то помимо традиционно используемых помет — книжн., уст., прост. и т. п., их значения покрываются частично лексическими же функциями, частично — расширением иллюстративного параметра, который иногда выливается в тематическое развертывание целой типовой ситуации --. Грамматические параметры специально не выделены, как это делается в обычных толковых или двуязычных словарях, однако все словоизменительные формы так или иначе фигурируют в словарных иллюстрациях, а категориальные значения раскрываются через посредство лексических функций. Тем не менее формально грамматический параметр в этом словаре можно считать отсутствующим, в явной форме не выраженным.

Теперь самое время поставить вопрос о том, что же следует понимать под параметром в строгом смысле, т. е. каков статус лексикографического параметра, его величина (объем) и структура. Наше определение по необходимости будет носить pragматический характер. Итак, под параметром понимается некоторый квант информации о языковой структуре, который в экстремальном случае может представлять для пользователя самостоятельный интерес, но, как правило, выступает в сочетании с другими квантами (параметрами) и находит специфическое выражение в словарях; иными словами — это особое словарное представление структурных черт языка.

Вопрос выделения и классификации параметров оказывается не таким простым, как могло бы показаться на первый взгляд. Некоторые из них выделяются довольно легко, как, например, орфографический (т. е. побуквен-

---

<sup>59</sup> Список лексических функций (=словарь метаязыка) см., напр., и приложениях к вып. «10 словарных статей Толково-комбинаторного словаря русского языка», М., 1971, с. 53—61.

<sup>60</sup> Там же, с. 36—37. В этом словообразовательном гнезде недостает, возможно, лишь трех слов, в числе лексических функций которых есть такие, которые дифференцируют их от всех названных: *пристыдить*, *пристыженный*, *застыдить*.

<sup>61</sup> Ср., напр., тематические контексты для слова *сердце*. — Там же, с. 15—16.

ная передача слова) или параметр «длина слова»/ (количество букв). С другими дело обстоит сложнее. Так, в случае параметра «семантический эквивалент слова»/ определить его содержание без предварительного анализа представляется затруднительным: является ли здесь существенным число значений; следует ли включать в структуру параметра различие между самостоятельным значением и оттенком или брать всю дефиницию недифференцированно; представлять ли этот параметр как пословную передачу некоего объяснительного текста (дефиниции) или давать его как сумму семантических компонентов, набор дифференциальных семантических множителей, лексических функций, семантических маркеров<sup>62</sup>, сем<sup>63</sup>, семантических классификаторов<sup>64</sup> и т. п.? Подобные вопросы решаются тем или иным конструктивным образом для каждого параметра в отдельности. Одним из основных критерииев при этом является наличие в лексикографической практике так называемых «однопараметровых»<sup>65</sup> словарей — синонимов, ударений, морфем, словообразовательных, новых слов и т. п. В рассматриваемом случае ответы на поставленные вопросы будут таковы.

1) Число значений не является существенным и его нельзя считать параметром, во-первых, потому, что в данном языке оно колеблется от словаря к словарю<sup>66</sup>, во-вторых же вследствие того, что при передаче смысла многозначного слова средствами другого языка, например, в двуязычных словарях, число значений будет колебаться от языка к языку в зависимости от наличия в выходном языке тех или иных эквивалентов<sup>67</sup>. Однако существенным следует признать различие многозначных и однозначных слов, и основой для введения такого параметра — «однозначность—многозначность» — является как раз существование в лексикографической практике специальных словарей, противопоставленных только по этому признаку (см. сноску 38). Данное решение<sup>68</sup> вовсе не означает, что параметр устанавливается всегда по типу словаря, тем более, что далеко не по всем параметрам такие словари существуют: если есть, например, словарь глагольного управления<sup>69</sup>, то нет словаря безличных конструкций (естественно, что он может быть создан, поскольку объективных препятствий для его построения как будто нет). Все же имеющиеся типы словарей (понимая под типом доминирующий параметр) помогают уточнить структуру (градуирование) лексикографических параметров. Так,

комплекс лексико-системных отношений с учетом имеющихся словарей градуируется, с одной стороны, последовательным наращиванием системных содержательных параметров: синонимов<sup>60</sup>, сенсем<sup>71</sup>, антонимов<sup>72</sup>, синонимов и антонимов<sup>73</sup>, ассоциаций<sup>74</sup>, синонимов, антонимов и близких по значению слов<sup>75</sup>, идеографических групп

<sup>62</sup> Katz J., Fodor J. A. The structure of a semantic theory. *Language*, 1963, v. 39.

<sup>63</sup> Greimas A. *Sémantique structurale*. Paris, 1966.

<sup>64</sup> Wahrig G. The syntagma as a fundamental unit of lexicon entries. — *Zeitschrift für Dialektologie und Linguistik*, 1969, Bd. 36.

<sup>65</sup> Как будет показано ниже, это понятие в значительной степени условно. На деле в подобном словаре информация дана синкетическим сочетанием параметров, и его название отражает параметр, доминирующий в таком сочетании.

<sup>66</sup> См., напр.: Верков В. П. Слово в двуязычном словаре. Таллин, 1977, с. 50–66.

<sup>67</sup> Там же.

<sup>68</sup> Понятно, что его (это решение) нельзя считать и единственным возможным: для нашей задачи оно представляется конструктивно оправданным, поскольку подводит единую основу под принципы кодирования семантической информации (см. ниже, § 6). Альтернативный путь, предполагающий релевантность числа значений, принят, например, в работе: Scherman D. A common structure for lexicographic data. — In: *Computers in the humanities*. Minneapolis, 1974, p. 220, где для этого параметра вводится особое обозначение *§ a*.

<sup>69</sup> Сердюкова П. П. Словарь глагольного управления испанского языка. М., 1966.

<sup>70</sup> Chauzaud H. B. *du Nouveau dictionnaire des synonymes*. Paris, 1971; *The modern thesaurus of synonyms, formerly the comprehensive word guide*. N. Y., 1965.

<sup>71</sup> Каменская О. Л. Об одном принципе составления двуязычных словарей. — В кн.: Перевод научно-технической литературы. Материалы семинаров ВЦП. М., 1976. — Под сенсемой в этой работе понимается совокупность слов, которые могут заменять друг друга без изменения смысла всего предложения (контекста) в языке, с которого переводят, и в языке переводящем. Таким образом, сенсемы определяются на основе расширенного толкования синонимии, и словарь сенсем представляет собой переводной словарь, обладающий свойством полной обратимости.

<sup>72</sup> Львов М. Р. Словарь антонимов русского языка. М., 1978.

<sup>73</sup> Введенский Л. А., Колесников Н. 77. Антопимо-синонимический словарь и его значение для обучения русскому языку иностранцев. — В кн.: Актуальные проблемы учебной лексикографии. М., 1977; Hayakawa S. I. *Modern guide to synonyms and related words*. Darmstadt, S. a. — Под «относящимися к ним словами» (related words) в этом словаре понимаются как раз антоними.

<sup>74</sup> Титова Л. Н. Киргизско-русский ассоциативный словарь. Фрунзе, 1975.

<sup>75</sup> Boussinot R. *Dictionnaire des synonymes, analogies et antonymes*. Paris, 1973.

(семантическое поле) <sup>76</sup>; с другой стороны, — системных параметров плана выражения: омографов <sup>77</sup>, омофонов <sup>78</sup>, омонимов <sup>79</sup>, паронимов <sup>80</sup>.

2) Отрицательный ответ на второй вопрос является следствием предыдущего решения. В силу подвижности, неопределенности границ между отдельными значениями различение оттенка и значения тоже становится довольно неустойчивым <sup>81</sup>, и поэтому в качестве параметра целесообразно принять всю дефиницию целиком.

3) Наконец, решение третьего вопроса находится в зависимости от ответов на первые два. Словарная дефиниция представляет собой текст особого рода — объяснительный по своему характеру. Специфика таких текстов заключается в том, что они целиком состоят как бы из ключевых слов, а это значит, что всякое полнозначное слово (за исключением так называемых «нулевых компонентов» — подробнее см. ниже, § 6, 10) в правой части толкового словаря может расцениваться как отдельный семантический компонент. Коль скоро вся дефиниция целиком приравнивается одному параметру, то его содержанием и будет набор слов, составляющих эту дефиницию <sup>82</sup>. Такое решение имеет то конструктивное преимущество, что позволяет в одном ряду рассматривать и одноязычные (толковые) и двуязычные словари: в последних «дефиниция» как раз и дана набором семантических эквивалентов (компонентов) входному слову. Отказ от использования для наших целей сем, маркеров, дифференциальных множителей или элементарных значений объясняется, таким образом, тем, что эти семантические элементы разработаны не для всего лексического состава языка, а только для некоторых специальных областей (ср. «термины родства», «цветообозначения», «мебель» и др.). Формально же эта невозможность их использования подтверждается отсутствием общего словаря элементарных значений (сем, семантических маркеров, классификаторов. . .), хотя в принципе такой словарь представляется осуществимым, и первый этап работы по обобщению, сведению в одну систему всех элементарных значений, выделенных в разных тематических областях и на материале разных языков мира, дал бы, вероятно, весьма поучительный результат.

Далее, способы задания дефиниции, как известно, могут быть различными. Например, в Оксфордских и Уэбстерских словарях дефиниция дана как бы трижды — толкованием слова, его синонимами и иллюстрациями его

употребления<sup>83</sup>. Кроме того, в идеографических словарях представлен четвертый способ — «неявного определения», заданный перечнем элементов семантического поля, а в словарях сочетаемости — пятый — указанием его синтаксической валентности<sup>84</sup>. Ясно, что хотя все эти способы являются описанием одного и того же слова, они не могут быть признаны эквивалентными друг другу, так как осуществляют это описание, эту характеристику с разных сторон. Таким образом, во всех этих случаях мы имеем дело не с одним, а с разными параметрами. Обоснованность такого вывода подтверждается и наличием в лексикографической практике «однопараметровых» словарей по всем названным величинам: толковых, синонимических, идеографических, сочетательных. Что касается «словаря употреблений», то им является «тезаурус» в своем первом, исконном значении, как он понимался, например, в рабо-

<sup>76</sup> Saareste A. Eesti keele mõistiline sõnaraamat, I—IV. Stockholm, 1958—1963.

<sup>77</sup> Системная роль омографии, а следовательно, значимость омографического параметра особенно подчеркнута в организации словарной статьи в словарях Уэбстера. Например, в «Webster's seventh new collegiate dictionary» указываются не только омографы к опорному слову, но омографы и их значения также и для синонимов опорного слова. Ср. использование специальных обозначений для тех (\$t) и других (\$y) при машинной обработке данного названного словаря в работе: SchermanD. Op. cit., p. 220.

<sup>78</sup> Townsend W. C. A handbook of homophones of general American English. Arlington, 1975; Aschiero H. R. Diccionario de homófonos castellanos. Repertorio alfabetico de adjetivos, verbos y su correcta ortografia homófonos, homógrafos, parónimos. Buenos Aires, 1975.

<sup>79</sup> Тышлер И. С. Словарь лексических и лексико-грамматических омонимов современного английского языка. Саратов, 1975.

<sup>80</sup> Constantinescu S. Mic dicționar de cunoscere și perechi. București, 1976.

<sup>81</sup> Бабкин А. М. Словарь языка и язык словаря. — В кн.: Современная русская лексикография. 1975. Л., 1976; ГорбачеваК. С., СороколетовФ. П. Значение и оттенок в лексикографической практике. — Изв. АН СССР. Сер. лит. и яз., 1975, т. 34, № 6.

<sup>82</sup> Ср.: Карапузов Ю. Н. Общая и русская идеография. М., 1976, с. 181—200.

<sup>83</sup> Ср. также: LudwigR. R., Barret E. F. Words, words, words. Vocabularies and dictionaries. Rochelle Park, 1973, p. 157.

<sup>84</sup> Этот параметр может быть детально разработан и в обычных толковых словарях, особенно в тех случаях, когда их составители ставят задачу достичь максимально возможной универсальности сообщаемой словарем информации, как, например, в проекте однотомного словаря чешского литературного языка: Filipec J. Cesta k češkemu jednosvazkovému slovníku. — Naše řec, Praha, 1975, roč. 58, с 5, s. 229—231.

tax Б. А. Ларина и Л. В. Щербы<sup>85</sup>, т. е. словарь, фиксирующий абсолютно все контексты, в которых встречается слово данного языка.

Подобные словари возможны для мертвых языков, и особенно эффективно их составление в тех случаях, когда письменные свидетельства какого-то языка дошли до наших дней во фрагментарном виде. Как частный случай общеязыкового «словаря употреблений», «словаря контекстов» следует рассматривать и конкорданс к произведениям писателя. Кроме того, иллюстративный параметр обычно широко представлен в исторических словарях, где материал ограничен определенными историческими рамками. Здесь, как правило, семантическое толкование играет вспомогательную роль, а решающее значение принадлежит перечню контекстов, иллюстраций, в идеале стремящихся с абсолютной полнотой отразить лексику соответствующих памятников<sup>86</sup>.

### Устройство и статус параметров

По своей величине, емкости параметры, будучи отражением всей языковой структуры, оказываются совершенно разными, часто несопоставимыми друг с другом. С одной стороны, мы имеем такой относительно простой параметр, как ударение в современном русском языке, который находит формальное выражение в выделении буквы, соответствующей ударенному звуку (постановкой ли специального значка сверху, снизу или сбоку от буквы, напечатанием ли ее другим шрифтом и т. п.); с другой стороны, мы имеем такой сложный и громоздкий параметр, как словообразовательное гнездо, который тоже может выражаться разными средствами:

- перечислением составляющих его элементов, либо сопровождающимся семантическим толкованием<sup>87</sup>, либо без такого толкования<sup>88</sup>;
- описанием семантических связей составляющих гнездо слов, которое может осуществляться, например, в виде графа с вершинами, соответствующими элементам гнезда<sup>89</sup>;
- перечнем реализованных данным гнездом словообразовательных значений<sup>90</sup>;
- указанием типов лексических функций, **свойственных исходному слову**<sup>91</sup>;

— может, наконец, иметь частичное выражение, как, например, в большинстве толковых словарей, где распространены «словообразовательные» дефиниции (отсылачные определения) типа: *идеалистический* — относящийся к идеализму, *идейность* — свойство идейного, *казуист* — занимающийся казуистикой, *каление* — действие по знач. глаг. *калить, ручеек* — уменьш. к *ручей, избенка* — униж. к *изба, ковкость* — качество, свойство ковского, *измятость* — состояние измятого, *железнодорожный* — относящийся к железной дороге, *картофельный* — относящийся к картофелю, приготовленный из картофеля, *жабий* — принадлежащий или свойственный жабе, *желтоватый* — слегка, несколько желтый, *жадущий* — крайне жадный, *жизне.* . . . — первая часть сложных слов, вносящая значение сл. *жизнь, . . . значный* — вторая часть сложных прилагательных, указывающих на количество цифр в числе, *обосабливать* — то же, что *обосаблять*<sup>92</sup>.

<sup>85</sup> Ларин Б. А. Принципы составления словаря автобиографической трилогии М. Горького. — В кн.: Словоупотребление и стиль М. Горького. Л., 1962, с. 3; Щерба Л. В. Избранные работы по языкоznанию и фонетике. Л., 1958, с. 69.

<sup>86</sup> См., напр.: Словарь русского языка XI—XVII вв., вып. 5. М., 1978.

<sup>87</sup> Словарь современного русского литературного языка, т. I—III, М.—Л., 1948—1952.

<sup>88</sup> Словарь-справочник по словообразованию. Алма-Ата, 1975.

<sup>89</sup> Гинзбург Л. Словообразование и синтаксис. М., 1979, с. 27—32, 41—57.

<sup>90</sup> См., напр.: Улуханов И. С. Словообразовательная семантика в русском языке. М., 1977, с. 127—129; Милославский Г. Вопросы словообразовательного синтеза. М., 1980, с. 11—13.

<sup>91</sup> 10 словарных статей Толково-комбинаторного словаря русского языка. М., 1971, с. 54—55. — К числу словообразовательных лексических функций в этом перечне следует отнести, в частности, № 7—10 (*обвинять — обвинение, пассажир — пассажирский, быстрый — быстро, друг — дружить*), № 11 (*стрелять — стрелок*) и № 17 (*стрелять — выстрел, стрельнуть*).

<sup>92</sup> Словарь современного русского литературного языка, т. IV—XVII. М.—Л., 1953—1965. — По сути дела, те же определения используются и в других словарях современного русского языка — Д. Н. Ушакова и С. И. Ожегова, в словарях которых лишь не с такой степенью полноты, как в Большом академическом словаре, включены производные слова. Ср.: «В Словаре представлено все богатство словообразования русского литературного языка, в том числе вторичные словообразования типа «легонько», «крючочек» и т. п., многие сложносоставные слова, многие относительные прилагательные и т. д., которые в иных словарях, как правило, не даются». — Инструкция для составления «Словаря современного русского литературного языка». М.—Л., 1958, с. 7,

Все эти разные средства выражения передают, однако, один и тот же параметр: при их воплощении в перечень элементов гнезда, связанных с исходным словом набором «отсылаочных определений», лексических функций или словообразовательных значений, в принципе должен получиться один и тот же результат. Таким образом, параметр выступает как своего рода функция, которая может принимать разные значения из области ее существования. Так, параметр рода имен существительных принимает значения муж., ср., жен. р. для русского языка, т., п., ф. — для немецкого, т., ф. — для французского и т. д.; параметр словоизменения (склонения) существительных либо принимает значения номера типа парадигмы, как в словаре Варига (см. выше), либо выражается с помощью специальных индексов, полностью отражающих соответствующую парадигму, как в словаре Зализняка <sup>93</sup>, либо, наконец, при выражении этого параметра используется свойство словоизменительных парадигм — через одну или несколько ключевых форм задавать всю совокупность составляющих ее словоформ <sup>94</sup>. Так, в русском языке для восстановления всей парадигмы имени существительного в большинстве случаев достаточно бывает указать, помимо рода, формы род. пад. ед. ч. и им. пад. мн. ч. В случае каких-то особенностей склонения (чередования в корне), к названным формам добавляется дат. пад. ед. ч., реже — твор. или предложный <sup>95</sup>. Для эстонской четырнадцатипадежной именной парадигмы такими ключевыми формами являются генитив и партитив ед. ч., иногда и партитив мн. ч. (*kala* — *kala* — *kalasid* 'рыба'; *inimene* — • *inimest* — *inimesi* 'человек'). В глагольном словоизменении для русского показательны 1-е и 2-е лицо ед. ч., тогда как для немецкого — 3-е лицо презенса и 3-е лицо имперфекта (*schlagen* — *schlägt* — *schlug*; *weisen* — *weist* — *wies*), а для финского — 1-е лицо презенса и 1-е лицо имперфекта (*sanoa* — *sanon*—*sanoi* 'говорить', *tehda* — *teen* — *teki* 'делать'). Таким образом, сами параметры могут либо быть в языке, либо отсутствовать, тогда как ключевые формы, или обладающие способностью к развертыванию в целую парадигму значения параметров, характеризуются национальной спецификой, варьируются от языка к языку, от словаря к словарю.

Что касается формального выражения значений параметров, то они закрепляются, как правило, в системе словарных помет, хотя есть и параметры (например, дефи-

ниция, орфография), которые специальных помет не имеют. Пометы одинакового порядка могут быть и значениями одного параметра, допуская его большую или меньшую детализацию, как, например, пометы стилистические (ср. в словаре Варига: *dial.* — диалектное, *img.* — разговорное, *vulg.* — грубое или в словаре Ушакова: *бран.* — бранное, *вульг.* — вульгарно», *детск.* — детское, *просторен.* — просторечие и т. п.), но могут представлять собой и разные параметры, как это было показано выше в случае системных отношений (синонимы, антонимы и т. д.). Подобную способность к детализации, развитию в глубину можно предположить, вероятно, у каждого параметра. Так, этимологический параметр может для русского языка ограничиваться исторической глубиной до древнерусского периода, т. е. восточнославянской общности, или же доходить до праславянского языкового состояния, может, наконец, углубляться до общеиндоевропейского уровня. Параметр, получающий в словаре Ожегова значение *спец.*, у Варига, например, детализируется в пометах *Arch.*, *Math.*, *Gram.* и т. д., а *жарг.* — в пометах *Kaufmannsspr.*, *Schülerspr.*, *Studentenspr.*, *Jägerspr.* и т. д. Даже такой параметр, как орфография, в принципе поддается детализации: поскольку в основе его лежат разные формы алфавитов, можно, например, составить словарь форм букв, как это сделано для французского письма определенного исторического периода<sup>96</sup>.

Стремление к детализации параметров находит своеобразное отражение в лексикографической практике и проявляется в создании однопараметровых словарей. Можно сказать, что каждый параметр, а экстремально — каждое значение параметра стремится стать отдельным словарем. Эта тенденция, имеющая, так сказать, центробежную направленность, прямо противоположна отмечен-

<sup>93</sup> Зализняк А. А. Грамматический словарь русского языка. М., 1977.

<sup>94</sup> «Например, если мы знаем, что исландское существительное *nafn* «имя» среднего рода и имеет в род. п. ед. ч. окончание -s (*nafns*), то на основании этого факта мы определяем всю парадигму этого существительного. . . Таким образом, отдельные элементы парадигмы, предсказывающие остальные, могут использоваться в качестве сокращенной записи всей парадигмы» (Берков В. И. Указ. соч., с. 32).

<sup>95</sup> Именно эти формы используются в словаре Ушакова.

<sup>96</sup> Малое В. Н. Происхождение современного письма. (Палеография французских документов конца XV—XVIII вв.). М., 1975.

ной в начале данного параграфа центростремительной тенденции к универсальности, к объединению всех лексикографических параметров в одном словаре. Обе разнонаправленные тенденции вместе и создают характерный для лексикографической параметризации дух — закреплять результаты языковых изучений во всех областях лингвистики в виде словарей.

Центробежную тенденцию можно видеть, во-первых, в построении системы словарей одного языка по наиболее важным, но традиционным для лексикографии параметрам, как сделано, например, Кюппером для немецкого языка<sup>97</sup>, или в углубленной разработке в словаре единственного параметра, но таким образом, что он предстает как сочетание ряда отдельных величин. Так, в выполненнном в Лавальском университете (Квебек) машинном словаре французского произношения<sup>98</sup> этот единый параметр подвергается разбиению на составляющие, каждая из которых поддается в том числе и количественной оценке. Таких составляющих выделено 24, но могло быть меньше или больше, поскольку матрица разбиения и измерения произносительного параметра, составленная методом последовательных приближений, или «планируемого эксперимента», допускает перманентную перестройку и корректировку. Вместе с тем такая неопределенность числа составляющих одного параметра вовсе не означает, что глубина его разработки не имеет ограничений, что параметр неисчерпаем вглубь. В самом деле, величины его разбиения в данном машинном словаре таковы:

- орфографические символы и их частотность на материале словаря в 40 тыс. слов;
- фонемы, их частотность в зависимости от позиции в слове (1-я, 2-я, 3-я . . .) и принадлежности к открытому или закрытому слогу;
- распределение фонем, многократно появляющихся в одном слове, по убывающей частоте;
- распределение фонем по позиции в слоге и в зависимости от слоговой длины слова.

Далее следуют пары фонем, затем фонетические слоги по составу и частоте, по отношению к началу или концу слова, закрытости—открытости слога, распределенности в словах разных грамматических категорий, в их соответствии с орфографическими слогами, по структуре (CVCC и т. п.), по графической длине слова и т. д.

Как легко видеть из этого неполного перечня, увеличение глубины представления одного параметра с неизбежностью ведет к расширению описания его связей с другими параметрами языковой структуры, к вовлечению их в сферу представления данного параметра. В рассматриваемом нами случае вовлеченными оказываются орфографический, параметр слогоделения, длины слова, категориальный (принадлежность к той или иной части речи), статистический. Но в принципе этим дело не кончается: при учете текстовых особенностей произношения (а такое расширение исследования предусматривается) в матрицу его характеристик могут быть вовлечены словоизменительный, словообразовательный и ряд других параметров.

Таким образом, при достижении определенной глубины в разработке отдельного параметра центробежная тенденция превращается в свою противоположность, в сферу лексикографического представления одного параметра вовлекается все больше и больше других <sup>99</sup>, а в экстремальном случае могут быть использованы, вероятно, все параметры языковой структуры. То есть максимальная глубина разработки одного параметра приводит опять-таки к идею универсального словаря.

Во-вторых, центробежная тенденция проявляется в создании словарей или проектов словарей по принципиально новым параметрам языковой структуры, вскрываемым на базе современных теоретических взглядов. Таков, например, словарь ядерных структур языка, возможности построения которого рассматриваются в зависимости от грамматической модели (таксономической или генератив-

---

<sup>97</sup> Küpper II. Wörterbuch der deutschen Umgangssprache. Bd. 3. Hochdeutsch—Umgangssdeutsch. Hamburg, 1905; Bd. 5. 10000 neue Ausdrücke von A—Z (Sachschen). Hamburg, 1967; Bd. 6. Jugenddeutsch von A bis Z. Hamburg—Düsseldorf, 1970; *Idem*. dtv-Wörterbuch der deutschen Alltagssprache. Bd. 1—2. München, 1971.

<sup>98</sup> MephamM. Le dépouillement d'un dictionnaire de prononciation. Département de langues et linguistique, Université Laval, Québec, 1977 (машинопись).

<sup>99</sup> С этим явлением нам придется еще столкнуться в матрице параметрического анализа словарей (см. ниже, § 2), где наглядно, на самих размерах тех или иных клеток матрицы, можно видеть, что расширение в некоторых словарях информации о каком-то параметре всегда означает одновременно вторжение в область значений других параметров.

ной), в рамках которой производится анализ<sup>100</sup>. Таков далее лингвострановедческий словарь<sup>101</sup> и его разновидность, его частный случай — объяснительный словарь<sup>102</sup>, доминирующим в которых является исторический этно-культурный параметр языка. В этом ряду должен быть назван и контентный словарь, или словарь для контент-анализа<sup>103</sup>, который, будучи ориентированным на применение машинной техники, оперирует классификаторами, приближающимися по своему статусу и своей структуре к развивающей здесь трактовке параметров, и создаваемый в помощь переводчику словарь «оперативных единиц речевой деятельности», выделяемых на синтаксическом и фразеологическом уровне<sup>104</sup>.

Дальнейшее углубление центробежной тенденции, т. е. стремление лексикографировать даже отдельные значения параметров<sup>105</sup>, как и сама способность параметра принимать разные значения, допускать фиксацию в разной форме, могут приводить к размытию этого понятия. Создается впечатление, что параметр, понимаемый как функция, принимающая различные значения, не имеет четких границ, не выражается в раз и навсегда заданных формах и может даже переходить один в другой. Так, в ареальный параметр, получающий наиболее полную детализацию в диалектных словарях, включаются не только указания на территорию распространения слова (явления), но и временные характеристики (время первой записи), и индивидуальные, персональные, связанные либо с авторством дефиниции, либо с указанием па источник информации, информанта<sup>106</sup>. С другой стороны, как отмечалось выше, многие параметры выступают в синкретической форме: акцентологическая информация дана либо в сочетании с орфографической, как, например, в словаре ударений русского языка, либо может быть привязана к произносительному параметру; слогоделение предполагает сведения о произношении<sup>107</sup>; переходность, безличность, рефлексивность не могут существовать вне категориального параметра «глагол», так же как род — вне параметра «существительное»; с категориальным параметром N — существительное, V — глагол в ряде случаев соединяется словообразовательный параметр, указывающий па сложное слово данной части речи, — NG (noun combining form), VG (verb comb. form)<sup>108</sup>.

В этих условиях особую важность приобретает определение статуса параметра, установление его отличия от

единицы языковой структуры и от лингвистической категории. Главное отличие параметра от названных понятий — прежде всего в его глобальности, параметр всегда относится к слову в целом: это не слог, а слогоделение, не аффикс или суффикс, а морфемное членение, не отдельное словообразовательное значение, а комплекс словообразовательных отношений, или гнездо, не фонема и не звук, а произношение, и т. п. Что касается его величины, то параметр, как правило, больше единицы, состоит из нескольких единиц (произношение, словообразовательное гнездо, дефиниция); в ряде случаев он может соотноситься с отдельной единицей (ядерные структуры, словосочетания, фразеологизмы) или быть меньше ее, совпадая с грамматической категорией (род, переходность). Вместе с тем, значительно число случаев, когда параметр вообще несоизмерим с языковой единицей: таков частотный параметр, особенно в специфических условиях, когда, напри-

- 
- <sup>100</sup> *Вовшин Я.М.* К проблеме выделения ядра предложения.— В кн.: Вопросы общей и прикладной лингвистики. Минск, 1975.
- <sup>101</sup> *Верещагин Е. М., Костомаров В. Г., Прохоров Ю. Е.* Лингвострановедческий словарь: к вопросу об учебной паремиографии. — Русский язык за рубежом, 1975, № 1; *Грдличка М.* Лингвострановедение и безэквивалентная лексика. — В кн.: Международная конференция преподавателей русского языка. 1975. Прага, 1976; *Денисова М. А.* Лингвострановедческий словарь. Народное образование в СССР. М., 1978.
- <sup>102</sup> *Андреева И. В., Баско Н. В., Шанский В. П.* Объяснительный словарь как один из типов учебного словаря. — Русский язык в национальной школе, 1976, № 2.
- <sup>103</sup> *IJolstig O. R.* Content analysis for the social sciences and humanities. Addison-Wesley Publishing Company, 1969, p. 150—152, 168.
- <sup>104</sup> *Горбис Б. И.* Психолингвистика и порождающая лексикография. — В кн.: Тетради переводчика, вып. 14. М., 1977.
- <sup>105</sup> Ср., напр., идею построения словаря словообразовательных значений, или «идеографического словаря морфем», в кн.: *Милюсавский И. Г.* Вопросы словообразовательного синтеза. М., 1980.
- <sup>106</sup> Ср., напр.: «Каламá, ý, ж. Кошелек с деньгами, мошна. Волог., Баженов. Велика ли у тя калама-то? Волог., 1902. ...» — Словарь русских народных говоров, вып. 12. Л., 1977. См. также: *Блинова О. И., Палагина В. В., Федоров А. И.* Инструкция для составления словаря русских народных говоров Сибири. — Учен. зап. Томского ун-та, 1975, № 9. Вопросы русского языка и его говоров, вып. 3.
- <sup>107</sup> *Бардовіч А. М., Шакун Л. М.* Марфемны слоўпік беларускай мовы. МіНСК, 1974. — Здесь наряду с выделением условных графических морфем дается разбивка слова па слоги в упрощенной транскрипции.
- <sup>108</sup> *Sherman D.* Op. cit., p. 219,

мер, статистической обработке подвергается семантическая сторона слова<sup>109</sup>; таков параметр семантического пробела в двуязычном словаре лексических лакун<sup>110</sup>; таков, наконец, параметр интерференции в специальном интерференционном словаре для обучения неродному языку<sup>111</sup>.

Эта общая несоотносимость параметра с языковой единицей обуславливает и непригодность для параметров обычной классификации по координатам «выражение—содержание», «язык—речь», «лингвистическое—экстравалингвистическое», «парадигматика—синтагматика» и т. п. Такая многомерная классификация, конечно, может быть создана и в какой-то степени будет полезной, но на ее основе нельзя построить теорию, а кроме того, она бедна конструктивными следствиями. На этом пути мы в лучшем случае сумеем лишь обнаружить некоторые, но далеко не все «пустые клетки» в системе: например, выделив по координате «выражение—содержание» группу ономасиологических параметров, т. е. параметров плана содержания, мы можем констатировать их реализацию в соответствующих словарях — ономасиологических в обычном смысле<sup>112</sup>, контентных, словообразовательных значений, образов или метафор<sup>113</sup>, идеографических. В этом ряду обращает на себя внимание отсутствие словаря грамматических значений<sup>114</sup>, словаря семантических множителей, «обратного», т. е. от смысла (ономасиологического) словаря фразеологизмов<sup>115</sup>. Названные недостающие словари призваны решать вполне определенные задачи в лингвистической науке, и формулирование предпосылок и целей их разработки могло бы расцениваться как практический результат многомерной классификации параметров. Но этот результат, кажется, и был бы единственным. Уже проблему совместности параметров и оптимального их сочетания в одном словаре нельзя решить на основе указанной классификации. А важность этой проблемы обусловлена по меньшей мере двумя причинами: во-первых, если центробежная линия развития лексикографической параметризации языка помогает выделять отдельные параметры, устанавливая случаи их синкretического задания, то центростремительная тенденция как раз и требует решения проблемы сочетаемости, совместности—несовместности параметров; во-вторых, сами приемы выделения параметров базируются на использовании критерия сочетаемости, поскольку совместность их значений или

способов задания есть показатель различия, разности параметров, а взаимоисключаемость значений — показатель того, что мы имеем дело с одним и тем же параметром.

### «Закон свободы параметров» и пути их систематизации

Рассмотрение современных словарей разных языков, предпринятое в данном разделе, не дало пока оснований говорить о наличии каких-то объективных препятствий на пути построения любого сочетания параметров. Этот вывод подтверждается историческим движением лексикографии (даже в рамках одной центростремительной линии ее развития) в направлении наращивания числа параметров в словаре. Первые словари всех языков, как известно, были двуязычными, и сообщаемая ими информация была минимальной. Так, в словаре Палласа мы находим три параметра — приблизительное произношение, семантическое соответствие (эквивалент) и название языка

- 
- <sup>109</sup> Лексика английского языка. Семантический частотный словарь по автоматической обработке экономической информации. М., 1975.
- <sup>110</sup> Муравьев В. Л. Лексические лакуны. (На материале лексики французского и русского языков). Владимир, 1975.
- <sup>111</sup> Чантуриашвили С. Основные проблемы сопоставительно-типологических исследований неродственных языков. — В кн.: Русский язык — язык межнационального общения народов СССР. М., 1976, с. 213—217.
- <sup>112</sup> Baldinger K. Dictionnaire onomasiologique de l'ancien gascon, fasc. 1. Tübingen, 1975; fasc. 2—3. Tübingen, 1977; *Idem*. Dictionnaire onomasiologique de l'ancien occitan, fasc. 1. Tübingen, 1975.
- <sup>113</sup> Григорьев В. П. Поэтика слова (на материале советской поэзии). М., 1979.
- <sup>114</sup> Принципиальная возможность создания такого словаря обоснована, как представляется, значительным числом работ по изучению семантико-функциональных полей в разных языках. Подобные исследования резюмируются, как правило, перечнем («словарем») самих значений, актуализирующихся в соответствующем поле, и средств их выражения в данном языке.
- <sup>115</sup> Что касается ономасиологического словаря фразеологизмов, то идея его создания могла бы родиться и без помощи параметризации — в рамках самой фразеографии. Предпосылками для его разработки являются, во-первых, кардинальная идея о сопотесенности фразеологических единиц со словом, и во-вторых, довольно обширный уже к настоящему времени опыт изучения фразеологизмов, эквивалентных различным частям речи: фразеологизмы с адвербильным, адъективным, номинативным, глагольным значением.

или диалекта (т. е. опять же приблизительный ареальный параметр) <sup>116</sup>. В первом русском академическом толковом словаре <sup>117</sup>, составленном по гнездовому принципу в алфавите корней, уже широкое применение нашли грамматические параметры. Всего здесь представлено 15 параметров — орфография, ударение, словообразовательный параметр, часть речи, род, склонение имен (указан род. п. ед. ч.), родовые окончания прилагательных, спряжение глаголов <sup>118</sup>, управление, вид и способ глагольного действия, залог, стилистический параметр, фразеологический, дефиниция, иллюстрации. В словаре Даля <sup>119</sup> только 12 параметров, здесь сильно сокращена грамматическая информация, зато большой глубины достигают ареальный, иллюстративный и словообразовательный параметры. В неизданном академическом словаре русского языка, большая часть выпусков которого издавалась под руководством Шахматова <sup>120</sup>, число параметров почти не увеличилось по сравнению с первым толковым словарем (добавился только этимологический), но все они получили более углубленную разработку. В частности, этимологический параметр представлен с такой же полнотой, как в этимологических словарях, а способ раскрытия иллюстративного параметра приближается к тезаурусному — в первом, исконном его смысле: ни в одном словаре русского языка нет такого количества и разнообразия контекстов употребления слова. Кроме того, были усовершенствованы и экономизированы способы задания некоторых параметров, что при том же количестве позволяет передать большую информацию. В Большом академическом словаре число параметров достигает уже трех десятков. Таким образом, факты как будто говорят в пользу того, что параметры могут совмещаться друг с другом в любых сочетаниях, а значит, в экстремальном случае соединение абсолютно всех известных языковых параметров в одном словаре представляется принципиально возможным. Иными словами, в лингвистическом конструировании словарей, так же как в художественном конструировании, или дизайне, действует своеобразный «закон свободы параметров» <sup>121</sup>.

Тем не менее мировая лексикографическая практика не знает универсального словаря, который объединил бы в своем составе все параметры, всю информацию о данной языковой структуре и ее функционировании. Едва ли препятствием здесь можно считать громоздкость подоб-

ного словаря и затрудненность восприятия большого объема информации пользователем, т. е. его трудную читабельность: ведь словарь — это произведение, которое не предназначено для сквозного прочтения и воспринимается всегда по частям, фрагментарно. Вполне преодолимыми представляются и всякого рода технические трудности полиграфического оформления такого издания. Дело здесь, очевидно, в чем-то другом, в каких-то объективных ограничениях, накладываемых на закон свободы параметров в словаре самими свойствами языковой структуры.

Чтобы подойти к разрешению этого парадокса (параметры сочетаются свободно, а универсальный словарь не создан), рассмотрим подробнее, каким образом реализуется в словаре закон свободы параметров. Во-первых, он проявляется варьировании их числа — от одного до нескольких десятков, причем эти колебания определяются не только целевой установкой словаря, но и его ориентацией на того или иного пользователя. Например, такой важный параметр структуры немецкого языка, как указание на то, с *haben* или *sein* спрягается в сложных временах данный глагол, является излишним в словаре, предназначенном для немецкого читателя, поскольку у последнего алгоритм употребления того или другого вспомогательного глагола заложен в сознании с детства. И наоборот, эта информация в виде пометки (h) или (s) после соответствующего глагола оказалась бы очень полезной в немецком словаре для иноязычного пользователя.

Во-вторых, закон свободы параметров проявляется в возможности выбора между разными способами задания одного и того же. Так, для словоизменительного параметра помимо упомянутых выше способов задания — указанием

---

<sup>116</sup> *Pallas P. S. Reise durch verschiedene Provinzen des Russischen Reichs*, Bd. III. SPb., 1776, напр., с. 373—376.

<sup>117</sup> Словарь Академии Российской, ч. I—VI. СПб., 1789—1794.

<sup>118</sup> Если на входе дан инфинитив, то к нему указаны формы 1-го лица ед. ч. наст. вр., 3-го лица ед. ч. прош. вр. и 1-го лица ед. ч. наст. вр. противоположного вида; если на входе дана форма 1-го лица ед. ч. наст. вр., то к нему указаны 2-е лицо наст. вр. и инфинитив.

<sup>119</sup> *Даль В. И. Толковый словарь живого великорусского языка*, ч. 1. М., 1863.

<sup>120</sup> Словарь русского языка, составленный Вторым отд. имп. Академии наук, т. I, вып. 1. СПб., 1891; т. III, вып. 1. Пг., 1922.

<sup>121</sup> *Селезнев И. Ф. Дизайн. Проблемы материально-художественной культуры*. Минск, 1978, с. 109—119.

номера типовой парадигмы или указанием ключевых форм — можно назвать еще два: агглютинативный, при котором каждой основе соответствует полный список ее флексий, и супплетивный, когда фиксируются все формы, составляющие данную парадигму<sup>122</sup>.

Наконец, в-третьих, действием рассматриваемого закона обусловлено потенциальное изменение глубины раскрытия ряда параметров — таких, как этимологический (его удобнее называть параметром родства, чтобы четче дифференцировать от заимствований, которые представляют собой самостоятельный параметр, тоже характеризующийся глубиной), стилистический, ареальный, дефинитивный и некоторые другие. Глубина, или объем дефинитивного параметра, например, колеблется от «филологического» определения, которое должно помочь только опознать «вещь», т. е. при минимуме средств дать ответ на вопрос «что это такое?», до энциклопедического определения, которое отвечает на вопросы «как это сделано или как сделать?» и «как функционирует или используется?». Причем уже само филологическое определение может быть разного объема — от предельно лаконичного, как в упоминавшемся выше словаре Варига, до максимально развернутого, включающего лингвострановедческий момент и исторические сведения, вплоть до указания на место и время первой фиксации данного слова, как в словаре современного французского языка<sup>123</sup>.

Что касается параметра заимствования, то помимо самого факта существования для каждого языка специальных словарей или хотя бы списков заимствованных им слов («однопараметровый словарь»), методологические аргументы в пользу его самостоятельности приводятся в статье В. Т. Коломиец, которая обращает внимание, в частности, на принципиально разные основы установления и анализа этимологии унаследованного слова, с одной стороны, и происхождения и условий освоения языком слова заимствованного, с другой<sup>124</sup>. Градуальный же характер параметра заимствования обусловлен тем, что он может варьироваться и предусматривать описание фонетических и семантических процессов, сопутствующих заимствованию, характеристику особенностей контактов взаимодействующих языков и народов их носителей, раскрытие роли языков-посредников<sup>125</sup>, наконец, оценку лексики всего языка по совокупности заимствованных им и заимствованных из него слов. Последний критерий практи-

Чески не разрабатывался в языкоznании, хотя мысль о том, что объем и характер лексики, заимствованной из данного языка, может стать социолингвистическим показателем в сравнительной оценке языковых структур, в той или иной форме высказывалась в работах последнего времени<sup>126</sup>. В этой связи возникает идея создания наряду со словарями заимствованных данным языком слов также особого словаря слов, заимствованных из данного языка. Так, для русского языка в условиях, когда возросла его роль как средства межнационального общения, когда он выдвинулся в число основных мировых языков, сводный «Словарь слов, заимствованных из русского языка» помимо его безусловной научной ценности, широкого применения в учебных целях, имел бы и большое общественно-политическое значение<sup>127</sup>.

---

<sup>122</sup> Агглютинативный, супплетивный и изолирующий (т. е. ссылкой на номер типовой парадигмы) способы фиксации гнезда слов в словаре различают авторы статьи: Беляева Л. И. и др. О многоцелевом автоматическом словаре русского языка (МАРС). — В кн.: Вопросы общей и прикладной лингвистики. Минск, 1975.

<sup>123</sup> *Trésor de la langue française. Dictionnaire alphabétique de la langue des XIX-e et XX-e siècles* (1789—1960), t. 1—5. Paris, 1971—1976.

<sup>124</sup> Коломієць В. Т. Основні аспекти етимологічних досліджень. — Мовознавство, 1977, № 6.

<sup>125</sup> См., напр.: *Leeming H. Rola języka polskiego w rozwoju leksyki rosyjskiej do roku 1696. Wyrazy pochodzenia łacińskiego i romańskiego*. Wrocław, 1976, где в качестве приложения к исследованию дан словарь латинских и романских слов, вошедших в русский через польский.

<sup>126</sup> Так, Дж. Бонфанте, анализируя заимствования в итальянском языке, отмечает, в частности, что германский пласт обогатил лексику итальянского языка значительным числом слов, обозначающих грубые, свойственные варварам действия, и характеризующихся зачастую яркой экспрессивной окраской (*Bonfante G. Latini e Germani in Italia*. Bologna, 1977, p. 48—52). С другой стороны, в анализе заимствований, обозначающих речь в английском языке, наряду с установлением постоянных однозначных соответствий между лексикой, заимствованной из одного языка-источника, и определенной тематической сферой, вскрывается тенденция к отчетливо выраженной и устойчивой эмоционально-экспрессивной оценке лексики из одного и того же языка-источника: германизмы описывают бранную коммуникативную ситуацию, характеризуют низкие формы речевого общения, галлизмы классифицируются как принадлежность светской беседы, несколько манерного разговора, но подчас с определенным подтекстом. См.: Олдырева Л. И. Именование речи (на материале американских словарей компании Мерриам—Уэбстер). Канд. дис. М., 1978, с. 34—85.

Итак, рассмотрение закона свободы параметров в словаре показывает, что по характеру его действия все параметры делятся на две группы. В одну из них входят собственно языковые, т. е. структурогенные параметры — ударения, орфографический, произносительный, словоизменительный, словообразовательный, категориальные, сочетаемости и др. Их значения дискретны, и возможность варьирования распространяется только на способы их задания, фиксации в словаре. Другую группу составляют параметры, содержание которых по необходимости включает и экстралингвистический фактор — денотативный, историко-культурный, pragматический, т. е. моменты, вторичные по отношению к языку, связанные не только с самим языком, но и с его изучением, т. е. с языкознанием. Эти параметры недискретны по своей природе, характеризуются изменяющейся глубиной раскрытия и отражают не собственно структурные отношения, а процессы — диахронические, синтагматические, взаимодействия языков, интерференции и т. п. В противоположность собственно языковым, или структурогенным, назовем их лингвистическими.

Не углубляя дальше характеристики этих двух групп параметров, можно высказать следующие предположения относительно ограничений, накладываемых на закон свободы параметров.

Первое из них заключается в том, что, как показывает лексикографическая практика, возможно построить словари, содержащие исключительно структурогенные параметры, но невозможны словари из одних лингвистических параметров: последние обязательно должны дополняться некоторыми структурогенными.

Второе ограничение касается сочетаемости: хотя сами параметры могут соединяться в любых комбинациях, далеко не все способы их задания и степень глубины оказываются совместимыми друг с другом. Таким образом, одна из трудностей на пути построения универсального словаря лежит, как представляется, в отыскании оптимального соотношения между способами задания структурогенных и степенью глубины лингвистических параметров в одном словаре.

Наконец, третье ограничение обусловлено спецификой входа в словарь. Во всех известных автору этих строк словарях вход осуществляется через один или несколько параметров, но обязательным условием при этом является

то, что эти параметры могут быть только структурогенными (орфографический, т. е. побуквенный — прямой или обратный алфавитный с учетом длины слова или без него; частотный; смысловой, дополненный алфавитным; морфологический, т. е. по алфавиту корней и др. — ср. ниже соответствующую графу параметрической словарной матрицы) и не могут быть лингвистическими. Более того, число входных параметров оказывается довольно ограниченным, т. е. далеко не все из структурогенных параметров могут реально использоваться на входе. Универсальный же словарь предполагает, очевидно, не один, а несколько входов, в том числе не только от основных структурогенных, но и от ряда лингвистических параметров. Эта трудность при современном уровне словаростроения представляется неразрешимой, поскольку возможности увеличения входов при сохранении словарей в традиционном виде — как книг, весьма ограничены. Наибольшим числом входов сейчас характеризуются словари-тезаурусы, но и здесь их все-таки безнадежно мало. Так, в словаре Касареса три входа — два смысловых (от дескриптора и от дескрипторной области) и один орфографический, т. е. алфавитный — от слова. Увеличение еще на один вход, скажем, частотный, сразу резко увеличит объем всего словаря и тем самым затруднит пользование им, снизит его читабельность.

Преодоление этой трудности, т. е. снятие или хотя бы сокращение ограничений на входы, видится в переходе к словарям на принципиально иной основе — машинной. В этом случае в память ЭВМ вводится некоторый список слов, каждое из которых снабжено набором параметров и их значений, и по желанию потребителя можно получить любую информацию, т. о. практически вход в словарь осуществляется от каждого явным образом заданного параметра. Применительно к русскому языку, например, такой словарь на 4000 слов создан в учебных целях в университете Кардиффа<sup>128</sup>. Здесь использовано

---

<sup>127</sup> См. подробнее: *Караулов Ю. Н.* Обратный словарь заимствований как способ изучения лингвоэкологии. — Изв. АН СССР. Сер. лит. и яз., 1979, т. 38, № 6.

<sup>128</sup> *Shibayev V.* The Carlex computerized analytical Russian dictionary. The computerization of complete complex grammar structures into dictionaries of highly inflected languages — Carlex results. — In: The computer in literary and linguistic studies. Cardiff, 1976.

всего 26 параметров, и только структурогенных, и обучающийся может сформулировать свой запрос в нужной ему форме: например, вывести на дисплей всю парадигму данного прилагательного или только его степени сравнения; найти форму прошедшего времени несовершенного вида от глагола *купить* (= *покупал*, -а, -о, -и); выяснить, какими падежами управляет предлог *под*, или найти все соответствия английскому *stir*. Более серьезная и глубокая исследовательская программа осуществлена на материале эстонского языка в Институте языка и литературы АН Эстонской ССР<sup>129</sup>. Построенный здесь на базе стотысячного ортологического словаря машинный словарь глаголов и существительных (т. е. по сути действующая модель спряжения и склонения) позволяет анализировать и синтезировать любую форму этого сильнофлективного языка. Понятно, что и число параметров здесь оказывается существенно большим: их выделено 88 только для одних глаголов<sup>130</sup>. В этом автоматическом словаре использованы также только структурогенные параметры. Примером словаря, сочетающего параметры обоих типов, может быть машинный архив финских диалектов, действующий в университете г. Турку. Его нельзя назвать в полном смысле универсальным, поскольку основан он на диалектных данных, хотя и привлекает для сравнения соответствующие конструкции литературного языка, но зато входы в этот словарь возможны и от многих грамматических параметров (как морфологических — общим числом 52, так и синтаксических — общим числом 33), и от параметров лингвистических, в частности — от названия диалекта и ареала его распространения, от имени (или возраста) информанта, от времени, когда была сделана магнитофонная запись<sup>131</sup>. Отличительной особенностью этого словаря, выделяющей его из всех других работ подобного рода, является исключительно широкая представленность в нем синтаксических параметров. Это и понятно: хотя по своему устройству данный машинный архив представляет собой словарь (перед вводом в машину каждое слово было занесено на отдельную карточку, и ему была приписана вся связанная с ним морфологическая и синтаксическая информация, а сами слова упорядочены как в прямом, таки в обратном алфавитном порядке), но он имеет дело с текстом — текстом финских диалектных записей, поэтому исключительное значение приобретают здесь синтаксические параметры. Среди последних тоже можно

йыделить как структурогенные (члены предложения; главное и подчиненное предложения; порядок следования главного и подчиненных; повествовательное, вопросительное или восклицательное предложение; прямая или косвенная речь), так и лингвистические (число предложений в тексте, порядковый номер предложения в тексте, число слов в предложении). Понятно, что почти все названные параметры, за исключением, может быть, первого — «члены предложения», получают определенные значения только в конкретном тексте, т. е. являются иными словами «текстовыми», и поэтому не могут быть включены в общий перечень «лексикографических», по крайней мере до появления собственно синтаксического словаря (ср. выше, сноска 36), который поможет решить вопрос об их квалификации.

Рассмотрение тенденции к лексикографической параметризации языка целесообразно завершить обобщением, которое подвело бы новую, параметрическую базу под словарную типологию, могло бы служить источником открытия и построения новых типов словарей и основой для выбора параметров при конструировании русского тезауруса. Путей для подобного обобщения существует много. Одним из самых привычных в языкознании была бы многомерная классификация, но от нее мы отказались по некоторым соображениям уже раньше. Весьма эффективным является метод морфологического ящика, предполагающий такое представление данных, которое позволяет анализировать все возможные изменения параметров и отражение этих изменений на конечном результате. «Суть метода состоит в систематическом исследовании всех мыслимых вариантов, вытекающих из закономерностей строения (морфологии) усовершенствованного объекта. При этом синтезируются как известные, так и новые необычные варианты, которые при простом переборе могли быть

---

<sup>129</sup> Remmel M., Viks O. Eesti verbid. Morfoloogiline klassifikatsioon. Loendid. (Estonian verbs. Morphological classification. Lists). Preprint KKI-3. Tallinn, 1974.

<sup>130</sup> Викс Ю. А. Классификаторная модель эстонской морфологии (автоматический синтез глагольных словоформ). Автореф. канд. дис. Таллин, 1978, с. 14.

<sup>131</sup> Ikola O., Karjalainen Y. Syntax archives of Finnish dialects for computer work. — In: Computational and mathematical linguistics, I. Firenze, 1977.

упущены. Идея метода — уйти в зону, далекую от того, что лежит на виду»<sup>132</sup>. Раскрытие возможных вариантов по каждому параметру осуществляется путем составления матрицы, карты или ящика. Однако при всех достоинствах этого метода он фактически оставляет в стороне проблему сочетаемости, совместимости различных характеристик. Поэтому его целесообразно дополнить приемами другого метода — метода «матриц открытия», который позволяет находить наиболее целесообразные комбинации параметров и включает такие этапы:

- 1) Составление перечня элементов, свойств, объектов, фактов, идей и т. п.
- 2) Выработка поля анализа и его структурирование, т. е. размещение характеристик выбранных элементов, свойств и т. п. по рядам и столбцам таблицы (матрицы).
- 3) Определение пересечения рядов и столбцов, обнаружение возможных комбинаций. Выясняется поле возможных решений, что является целью исследования. Каждая ячейка таблицы представляет собой связь двух характеристик.
- 4) Рассмотрение всех возможных решений с целью обнаружения новых допустимых комбинаций.
- 5) Изучение различных комбинаций и выбор рациональных решений<sup>133</sup>.

В этой главе мы ограничимся первыми тремя этапами охарактеризованного метода, посвятив им следующий параграф и резервировав за собой возможность обращаться к построенной матрице при решении практических вопросов по ходу изложения последующих глав.

## § 2. ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ СЛОВАРЕЙ

Весь метод состоит в размещении и упорядочении того, на что должно быть направлено острие ума в целях открытия какой-либо истины.

*P. Декарт. Правила для руководства ума*

Содержание этого раздела составит собственно матрица лексикографических параметров («матрица открытия»), материал для которой подготовлен предыдущим параграфом. Что касается (1) «составления перечня элементов» (в данном случае «параметров»), то эта операция требует лишь обобщения и упорядочения изложенных выше расуждений. **Лексикографические параметры, каждый из**

которых снабжен порядковым номером, откладываятся по горизонтали, тогда как по вертикали расположены различные словари. «Поле анализа» (2), таким образом, представлят собой, с одной стороны, столбцы матрицы, включающие варьирование значений параметров от языка к языку и способов их задания и глубины раскрытия — от словаря к словарю, а с другой стороны, ряды матрицы, объединяющие определенные значения параметров в конкретном языке и показывающие их глубину в данном словаре. «Пересечение рядов и столбцов» (3) демонстрирует «поле возможных решений» и дает материал для комбинирования, экстраполяции и обнаружения иных комбинаций параметров при конструировании новых лексикографических объектов.

По техническим причинам эта единая матрица передана в книге не в виде непрерывной таблицы, а разбита на фрагменты и напечатана постранично, что не снижает, на наш взгляд, ее эффективности, поскольку главный интерес для читателя представляют «поле анализа» и «поле возможных решений», сконцентрированные в ее отдельных ячейках. Для облегчения ориентировки читателя при пользовании матрицей ниже приводится полный перечень ее столбцов и рядов. Следовательно, чтобы обнаружить интересующую нас ячейку, например, значение, способ задания и глубину раскрытия хронологического параметра (столбец 3) в словаре «Trésor» (IV ряд), нужно раскрыть книгу на странице, которая является общей для указанного столбца и ряда (в данном случае — с. 88).

#### Перечень лексикографических параметров — по горизонтали матрицы

- |  |   |
|--|---|
| 1. Язык (языки)                                      | 5. Орфографический параметр                     |
| 2. Вход(ы) Б словарь: количество и характер          | 6. Графическая длина слова<br>(количество букв) |
| 3. Хронологический параметр                          | 7. Ударение                                     |
| 4. Количественный параметр<br>(число слов в словаре) | 8. Фонетический параметр<br>(произношение)      |

---

<sup>132</sup> Методы поиска новых технических решений. Йошкар-Ола, 1976, с. 10; подробнее см.: Zwicky F. The morphological approach to discovery, invention, research and construction. — In: Zwicky P., Wilson A.G. New methods of thought and procedure. Berlin, 1967.

<sup>133</sup> Там же, с 15. Подробнее см.: Moles A., Claude D. Créativité et méthodes d'innovat. Paris, 1970.

9. Слогоделение (слоговая длина слова, число слогов)
10. Часть речи (категориальный параметр)
11. Грамматический род имени существительного
12. Число
13. Род имени прилагательного (родовые окончания)
14. Краткие формы прилагательных
15. Склонение прилагательных
16. Степени сравнения прилагательных и наречий
- 17. Вид глагола и способ глагольного действия**
18. Переходность
19. Безличность
- 20. Управление глагола**
21. Возвратность (возвратный залог)
22. Залог(и) (прочие)
23. Сослагательное наклонение
24. Повелительное наклонение
25. Инфинитив(ы)
26. Прочие наклонения
27. Спряжение глагола (в настоящем времени)
28. Отделяемые—неотделяемые приставки
29. Вспомогательные глаголы в глагольной парадигме
30. Причастие (или основа прошедшего времени)
31. Будущее время (будущие времена)
32. Прошедшее время (прошедшие времена)
33. Склонение имен существительных
34. Морфологическое членение слова
35. Словообразовательный параметр
36. Многозначность—однозначность
37. Семантема (семантический эквивалент, словарное толкование, дефиниция)
38. Параметр транспозиции (общности) значения
39. Ареальный параметр
40. Синтагматический параметр (свободная сочетаемость)
41. Фразеологический параметр (несвободная сочетаемость)
42. Экземплярно-иллюстративный параметр (вербальные иллюстрации, примеры употребления)
43. Стилистический параметр
44. Эмоционально-оценочный параметр
45. Нормативный параметр
46. Статистический (частотный) параметр
47. Лингвострановедческий (этнокультурно-исторический) параметр
48. Родство (генетическая общность, этимологический параметр)
49. Заемствования
50. Синонимы
51. Антонимы
52. Омонимы
53. Ассоциативный параметр
54. Семантическое поле
55. Тематическая группа
56. Лексический класс
57. Ономастический параметр
58. Терминологический параметр
59. (Лингво-)исторический параметр (история слова)
60. Библиографический параметр
61. Лексикографический параметр (указание «лексикографических адресов», т. е. других словарей, в которые включено данное слово)
62. Лексическая сочетаемость
63. Аббревиатуры
64. Рифма (совпадающие финалы слов)
65. Паронимы
66. Картинный параметр (иллюстрации рисунком как эквивалент остативного определения)
67. Общее число параметров:
  - а) лексикографируемых в данном языке;
  - б) представленных в данном словаре полностью;
  - в) представленных частично

## Перечень анализируемых лексикографических объектов (словарей) — по вертикали матрицы

- I. Ушаков — Толковый словарь русского языка. Под ред. Д. Н. Ушакова. Т. 1-4. М., 1934-1940.
- II. Competition Dictionary — The indispensable competition dictionary. Bombay, the Times of India press, S. a.
- III. Webster-7 — Webster's seventh new collegiate dictionary. Springfield, Mass., 1963.
- IV. Trésor — *Trésor de la langue française. Dictionnaire alphabétique de la langue des XIX-e et XX-e siècle* (1789—1960). T. 1—5. Paris, 1971—1976.
- V. Oxford Dictionary — The Oxford English dictionary, v. 1—12. Oxford, at the Clarendon press, 1933.
- VI. Grand Larousse — *Grand Larousse de la langue française, en six volumes*. Paris, Librairie Larousse, 1971—1978.

В матрице использованы только два условных обозначения: 1) «0» — свидетельствует, что соответствующий параметр отсутствует, так как не является пертинентным в данном языке (ср., например, в ячейке 17-V, где должен быть указан способ подачи глагольного вида в Оксфордском словаре);

2) «—» — прочерк означает, что параметр, будучи представленным в соответствующем языке, не имеет специального выражения (отсутствует) в данном словаре, т. е. не является для него релевантным (ср., например, в ячейке 9-V — «слогоделение в Оксфордском словаре»).

Наличие текста в ячейке есть показатель релевантности параметра, а само содержание текста раскрывает значение параметра и глубину его разработки в словаре. Возможны переходные случаи, когда ячейка содержит и прочерк, и текст после прочерка. Эта ситуация характеризует частичное выражение параметра в соответствующем словаре, что явствует и из содержания текста (см., например, 50-I «синонимия в словаре Ушакова» или 39-VI «ареальный параметр в словаре Larousse»).

Таблица 1. Матрица параметрического анализа словарей

Словарь	Параметр, его значение, способ задания или глубина			
	(Язык (языки)	Входы в словарь: количество и характер	Хронологический параметр	Количественный параметр (число слов в словаре)
0	1	2	3	4
I. Ушаков	Русский	Один: орфографический, прямой алфавитный	Рамки 1800—1930	100 тыс.
II. Competition Dictionary	Английский	От любого параметра, пред- ставленного в словаре — с учетом «синонимии вхо- дов» — 177 входов (по алфа- виту их названий)	—	400 тыс.
	Орфографический параметр	Графическая длина слова (количество букв)	Ударение	Фонетический параметр (произношение)
0	5	6	7	8
I	Побуквенная передача	—	1. На входе: знак ' после бук- вы, обозначающей ударенный звук; 2. в словоформах: знак ' над буквой	Частичное, для отдельных сло- гов заимствованных слов: гал- люцинация [алю]
II	Побуквенно	Внутри каждого параметра словарные статьи располо- жены по возрастанию числа букв	Только в толковой части — выделением ударного слога: Comfort, flum-fort	Только в толковой части: тран- скрипция упрощенная «англий- ская» — с помощью названий букв: Commonly, kom-mon- lē . . ; Lazy, /ay-ze. . .

Таблица 1 (продолжение)

	Слогоделение (число слогов)	Часть речи	Род имени существительного	Число	Род имени прилагательного
0	9	10	11	12	13
I	—	Суш. — синкетично с указанием рода (см.), прил. — синкетично с указанием родовых окончаний (см.), глаг. — синкетично с указанием 1 и 2 лица наст. вр. <см.>, нареч., прич., предлог, частица, союз, междом., местоим., числิต.	1. м., сп., ж. 2. ж. к: <i>готентотка</i> — ж. к <i>го- тентот</i> ; <i>гостя</i> — ж. к <i>гость</i>	Только в отклонениях: <i>вред</i> — мн. нет, <i>врата</i> — ед. нет, <i>люди</i> — мн. к <i>человек</i> , <i>весло</i> — мн. <i>весла</i> , <i>выдрята</i> — мн. от <i>выдрёно</i> , <i>выгреб</i> — только ед.	<i>Веселый, ая, ое</i>
II	Только в толковой части — постанов- кой дефиса между слогами: <i>Indeciphe- rable, in-de-sy-fer- a-ble</i> ...	—	0	—	0

Таблица 1 (продолжение)

	Краткие формы прилагательных	Склонение прилагательных	Степени сравнения прилагательных и наречий	Вид глагола	Переходность	Безличность	Управление глагола	
0	14	15	16	17	18	19	1	20
I	<i>Веселый</i> — весел, весела, <i>весело</i> ; <i>зажиточный</i> — чен, чна, <i>что</i>	—	На входе — в случае отклонений от правил: <i>больше</i> <i>j</i> — больший <i>j</i> — ср. ст. от <i>большой</i> , <i>гуще</i> — ср. ст. прил. от <i>густой</i>	Сов., несов., многокр., однокр.	Кого-что, что	Безл.	Звонить — во что и без доп., доплю- нуть — до чего; взыскать — с кого- чего; <i>взвывать</i> — к кому-чему, о чем	
II	ø	ø	—	0	—	—	—	—
	Возвратность (возвратный залог)	Залог(и) (прочие)	Сослагательное наклонение	Повелительное наклонение	Инфинитив(ы)	Прочие наклонения		
0	21	22	23	24	25		26	
I	<b>Возвр.: мыть-ся...</b> — возвр. К <b>мыть</b>	В глаголе и причастии (см. колонку 30) действ. залог специального обозначения не имеет, т. е. отсутствие пометы соотв. действ. залогу; страдательный обозначается «страд. к ...»; <i>взыхляться</i> — страд. К <i>взыхлять</i>	—	Флексией на выходе: <i>взъестся... пов. -вшись</i>	Глагол в форме инфинитива подается на вход слова; обозначение — инф. в описании сочетаемости: <i>готовый</i> к чему с инф. и без доп.		0	
II	—	—	—	—	—	На вход, как исходная форма глагола; специально не обозначен	—	

Таблица 1 *(продолжение)*

	Спряжение глагола (в настоящем времени)	Отделяемые—неот- деляемые пристав- ки	Вспомогательные глаголы в глаг. парадигме	Причастие (или основа прош. вр.)	Будущее время	Прошедшее время
0	27	28	29	30	31	32
I	Обозначается на выходе: 1) флексиями или формами 1 и 2 лица ед. ч.: <i>иссушить, ушу, ушишь; иста- ять, аю, ешь;</i> 2) флексиями или формами 1, 2 и 3 лица ед. и мн. ч. в случаях морфонологических изменений: <i>взъестись, ёмся, ёсься, ёстся, едамся, едитесь, едатся;</i> 3) при отсутствии наст.: <i>быть, наст.</i> нет, кроме 3 л. ед. ч. <i>есть</i> и устар. 3 л. мн. ч. <i>суть, едатъ,</i> <i>наст.</i> нет	0	быть для образования буд. и прош. времен ( <i>будет играть, будет, была сыграна</i> ) определяется не как вспомогательный глагол, а как связка	Причастие подается на вход самостоятельно: <i>взятый, ая, ое; взят, ята, ято.</i> Прич. страд. прош. вр. от <i>взять</i> , <i>воплощенный, ая, ое; щен, щена, щен.</i> Прич. страд. прош. вр. от <i>воплотить; истек- ший — прич. действ. прош. вр.; выдающийся — прич. наст. вр. от выдавать</i>	На вход дополнительно к словнику (мелким шрифтом): <i>истешу, ешешь — буд. вр. от истес- тать, выдам, аша, ост Ед. ч. буд. вр. от выдать</i>	1. На выходе при обозначении словоформ — прош.: <i>до- плести... прош. ел., ела', взятыся... прош. ялся, ялась, ялось; истереть... прош. истер, ла. 2. На вход подаются отдельные словоформы: <i>истёп, екло — прош. вр. от истечь; истёр, ла — прош. вр. от исте- реть</i></i>
II	—	0	—	—	—	—

Таблица 1 (продолжение)

	Склонение имен существительных	Морфологическое членение слова	Словообразовательный параметр	Многозначность-однозначность	Семантича семантический эквивалент, дефиниция)
0	33	34	35	36	37
I	<p>1. Род. п. ед. ч. <i>стол</i>, <i>а</i>      2. Род. п. ед. ч., предл. п. ед. ч.: <i>лес</i>, <i>а</i>, <i>о лесе</i>, <i>в лесу</i>.      3. Род. п. ед. ч., им. п. мн. ч., род. п. мн. ч.: <i>друг</i>, <i>а</i>, <i>брата</i>, <i>зей</i>.      4. Род. п. ед. ч., им. п. мн. ч., род. п. мн. ч., дат. п. мн. ч.: <i>кольцо</i>, <i>а</i>, мн. <i>кольца</i>, <i>колец</i>, <i>кольцам</i>.      5. Род. п. ед. ч., вин. п. ед. ч.: <i>зима</i>, <i>ы</i>, <i>вин. зиму</i></p>	<p>—</p> <p>Частично:</p> <p>1. Префиксы: <i>вес-</i>, <i>анти-</i>, <i>вы-</i>.      2. Композиты — 1-е и 2-е части: <i>дву-</i>, <i>макро-</i>. . .      3. Суффиксы: <i>-ификация</i>. . .</p>	<p>Отсылочными определениями: <i>всё</i> —      1. только ед. Действие по глаг. <i>всеть</i> — 2 знач.; <i>вещь</i> — прил. к <i>вещь</i>; <i>вещица</i> — уменьш.-ласк. к <i>вещь</i>; <i>вещунья</i> — ж. к <i>вещун</i>; и т. п.</p>	<p>Многозначность выражается во введении нумерации значений, однозначность — в отсутствии нумерации</p>	<p>Филологическое определение</p>
II	0	<p>1) Суффиксы,      2) префиксы,      3) вторые части композитов (вход — по алфавиту названий групп): Suffixes: ac, al, an, ar. . . ; ble, cle. . . , ation. . .</p>	—	—	<p>Предельно сжатое определение (Concise Dictionary):      Line, to place in a line; to mark out; to cover on the inside.      Look, to direct the eye at an object; to examine; to seem; to expect</p>

Таблица 1 (продолжение)

	Параметр транспозиции (образности) значения	Ареальный параметр	Синтагматический параметр (свободная сочетаемость)	Фразеологический параметр (несвободная сочетаемость)	Экземплярно-иллюстративный параметр (вербальные иллюстрации, примеры употребления)
0	38	39	40	41	42
I	Перен. букв.	Частично: 1. обл. (напр. <i>волок, галушка</i> ) 2. моск.	—	— В «заромбовой» части словной статьи на ключевое слово: о иметь зуб (против кого-н.), зуб на зуб не попадает. Т. о. «ходом» во фразеологии служит ключевое слово (в данном случае — зуб)	1. Литературные цитаты. 2. Речения. 3. Словосочетания
II	—	«Индийский глоссарий» — список слов, употребляющихся в речи англичан в Индии: Baba—Lit. «Father», a respectful «Mr.», Irish «Your Honour». Zanana — of women. Women's apartment, harem	—	Устойчивые выражения (отдельный список) — Triple Phrases: By and large. As dead as mutton. Pull devil, pull baker	—

Таблица 1 (*продолжение*)

	Стилистический параметр	Эмоционально-оценочный параметр	Нормативный параметр	Статистический (частотный) параметр	Лингвострановедческий (этно-культурно-исторический) параметр
0	43	44	45	46	47
I	I. Разг., простореч., фам., детск., вульг., арго, школьн. II. Книжн., науч., тех., спец., газет., публичн., канц., офиц., поэт., нар.-поэт.	Бран., ирон., неодобрит., шутл., презрит., пренебр., укор., торж., ритор., эвф., неприл.	1) Помета неправ. или отсутствие такой пометы 2) Варианты, допустимые (на входе: <i>санаторий</i> — <i>санатория</i> и т. д.)	—	I. В пометах: иностр., истор., дореволюц., загр. II. В дефинициях: напр., <i>есаул</i>
II	Сленг и коллоквиализмы с отдельными входами: <i>Dictionary of slang: Ananias: A liar. Lady: The reverse side of a coin. The keeper of the gunners' small stores. Queen, in card games</i>	—	—	—	—

Таблица 1 (продолжение)

Родство (генети- ческая общность, этимоло- гический параметр)	Заимствования	Синонимы	Антонимы	Омонимы	Ассоциативный параметр	
0	48	49	50	51	52	
I	—	Указанием языка и исходной формы в нем: <i>dervish</i> — перс. <i>дэргиз</i> — ни- щий; <i>derby</i> [дэ . . . по имени основателя этого вида спорта английского лорда Derby]; азерб., амер., греч., мек- сикан., новолатин., тат., укр., чагат.	(Только как разно- видность дефиниции: депрессия — упадок, застой . . ., заго- лить — обнажить, оголить)	—	—	Обозначаются на входе цифровыми индексами: <i>кося<sup>1</sup></i> (прическа); <i>кося<sup>2</sup></i> (орудие); <i>кося<sup>3</sup></i> (уз- кий мыс); <i>кося<sup>4</sup></i> (по- лоса леса)
II	—	—	Вход от списка клю- чевых слов. От са- мых синонимиче- ских эквивалентов входа нет (т. е. нет индекса): <i>aban- don, abdicate, ajure, forswear, recant, vacate. . .</i>	Так же, как и си- нонимы, к тем же самым «ключам»: <i>abandon. . . (keep, cherish, adopt. . .)</i>	Омографы и омофо- ны — в рубрике Double Meaning Words: <i>ancient — aged per- son, a flag, cabin — a hut, a ship's apart- ment</i>	Ideas (Association of) — в виде ассо- циативных полей, групп к некоторому списку ключевых слов: <i>alien — foreig- ner, immigrant, new- comer, stranger</i>

Таблица 1 (продолжение)

98

	Семантическое поле	Тематическая группа	Лексический класс	Ономастический параметр	
0	54	55	56	57	
I	—	—	0	—	
II	Родо-видовые отношения заданы двумя входами (оба от рода, различие в том, что в первом есть упорядоченность по числу букв в названиях видов и минимизированные дефиниции): 1) Clues and their Answers; 2) Clues and Alternative Solutions	40 крупных тематических областей, обозначенных 72 входами по алфавиту их названий (Occupations — см. также Professional, Trade, etc.)	0	Задан тематически, т. о., что охватывает все ономастические объекты: а) фамилии знаменитостей по областям деятельности, марки вин, автомобилей, названия клубов, фильмов, школ, мужские и женские имена, топонимы, административно-политические названия, корабли, зоонимы, этноНимы...	
	Терминологический параметр	(Лингво-) исторический параметр	Библиографический параметр	Лексикографический параметр	Лексическая сочетаемость
0	58	59	60	61	62
I	Авиац., апт., биол., бухг., ест., зоол., казак., малаяр., ме-теор., охот., порт., религ., с.-х., филос., шахм., этнол... (всего более 70 помет)	Частично; только в по-метах: нов., церк.- книжн., старин., устар.	—	—	—
II	—	—	—	В разделе «Clues and their Answers» — указание на словарь, откуда взята	—

Таблица 1 (*продолжение*).

	Аббревиатуры	Рифма (совпадающие финали слов)	Паронимы	Картинный па- раметр (иллюстра- ции рисунком как эквивалент остенсивного определения)	Общее число па- раметров: а) лексико- графируемых в данном языке; б) представленных в данном словаре полностью; в) представленных частично
0	63	64	65	66	67
I	—	—	—	—	a) 63 б) 38 в) 6
II	Abbreviations (отдель- ным входом)	Rhyming Words (в алфа- вите букв, относящихся к окончанию): -ague: plague, vague..., -owl: cowl, foul, growl..., -url: burl, churl, girl...	Some Solutions of spe- cial Similarity of Spell- ing and Lettering (``Trap-words''): Confu- sion — mess, toss; Light — kindle, candle	—	a) 58 б) 25 в) —

Таблица 1 (продолжение)?

0	1	2	3	4		
III	Webster-7	Английский	1) Орфографический, прямой алфавитный; 2) по параметру аббревиации; 3) по параметру рифмы	—	Около 300 тыс. слов	
IV	Trésor	Французский	Один: орфографический, прямой алфавитный	1) Время первой письменной фиксации по отношению к каждому значению (в разделе словарной статьи, посвященной этимологии слова); 2) рамки словарника (соответственно иллюстраций) — 1789-1960 гг.; 3) дата употребления в иллюстративном примере	(Не окончен)	
0	5	6	7	8	9	10
III	Побуквенно, варианты подаются на вход в алфавитном порядке: <i>heart'i-ly</i>	—	Постановкой значка после ударного слога: <i>cɔ:g'ro· ra'tive</i>	1) Полное; транскрипционное обозначение всех звуков — [k-]spain; 2) частичное: для фиксации вариантов произношения и для сложных слов: <i>head'mas'ter</i> (-mas-ter)	Разделение слов знаком ударения или точкой в орфографическом (см. 5, 7) и транскрипционном (см. 8) написании	AJ — adjective, AV — Adverb, CJ — conjunction, DA — definite article, IA — indefinite article, IJ — intonation, N — noun, PN — pronoun, PP — preposition, VB — verb
IV	Побуквенно — с вариантами написания и кратким очерком исторической эволюции современного орфографического облика слова	—	—	Французская транскрипция с указанием вариантов произношения: <i>abandon</i> — abādō, abā-dōn	—	Subst., adv., verbe, adj., . . .

Таблица 1 (*продолжение*)

0	11	12	13	14	15	16		17
III	0	Для pluralia, singularia tantum — спец. помета с ссылкой к статье: see PLURAL, Note...	ø	0	0	—	Только для супплетивов, которые поданы на вход: best — superl. of GOOD	0
IV	masc., fém.	—	-ée, -ë	0	0	—	Только для супплетивов	0
, 0	18	19	20	21	22	23		24
III	VI — verb, intransitive VT — verb, transitive	VM — verb, impersonal	Предлоги-наречия в словарной статье глагола — по алфавиту, — как модификаторы лексического значения входной единицы: catch, catch on; catch up	Reflex.	—	—	VP — verb, imperative	
IV	Verbe trans., verbe intrans. — с тонкой градацией реального обладания объектом или кажущегося и т. п.	Impers.	Беспредложное или предложное — с указанием соответствующих предлогов: emploi trans. — acclamer..., emploi intrans., — acclamer a..., acclamer pour...	Emploi pronom — s'abandonner aux mouches	—	—	—	

Таблица 1 (*продолжение*)

0	25	26	27	28	29	30	31
III	На вход	—	—	0	На вход с пометой VA — verbal auxiliary	Для неправильных глаголов: met, written, spoken — в словарной статье основной формы и на входе в алфавите словарника	—
IV	На вход	—	—	0	—	Part. passé adj. et subst. — на вход и в разделе словарной статьи, посвященной словообразовательным характеристикам единицы	—
0	32	33	34	35	36		
III	—	0	Суффиксы и префиксы в алфавитном порядке общего словарника подаются на вход с пометами: AS — adverb suffix JS — adjective suffix NS — noun suffix PF — prefix SF — suffix VS — verb suffix	Частично фиксируется в словарной статье, частично подается на вход с пометами: CP — combining form NC — noun combining form VC — verb combining form	Numeration of values within the dictionary article		
IV	—	0	Суффиксы и префиксы даны на вход; в разделах словарной статьи «Этимология» и «История» данного слова прослеживаются предпосылки его современного морфологического членения	Дериваты включаются в общий алфавитный словарь и подаются на вход как дополнение к словарной статье, с указанием формы, послужившей исходной для каждого случая	Практически не проводится различий между оттенком и значением, поэтому число значений (единая нумерация внутри словарной статьи) для многозначных слов очень велико		

Таблица 1 (продолжение)

0	37	38	39	40
III	Филологическое определение для каждого значения обычных слов и элементы энциклопедического определения (т. е. «как сделано или из чего состоит, как и где применять») для терминов	Lit. — fig. (прямое — переносное), с подразделением внутри каждого на буквенное и цифровое различие оттенков	Dial., Anglo-Ind., Austral., Can., Brit. ...	В виде помет в словарной статье, касающихся особенностей сочетаемости и синтаксического статуса слова: <i>be·ware</i> , commonly followed by of, lest, how, that not.
IV	Предельно сжатая дефиниция каждого значения, поскольку семантическая информация обильно рассеяна по другим разделам словарной статьи (при характеристике сочетаемости, стилистических особенностей и т. п.)	Прямое (A) — переносное (B) как основное деление на два комплекса значений, внутри которых употребления классифицируются по сферам языка (терминологическое или общее), по оттенкам значения в данной сфере, по грамматическим свойствам и т. п.	В ремарках к отдельным употреблениям: <i>dial.—région</i> .	В виде подробной характеристики синтаксических партнеров данного слова: агента (по его принадлежности к тематическим классам лиц, вещей, животных), первого объекта (лицо, вещь, место, животное или абстрактное — конкретное), второго объекта (предложная связь), устойчивых определений к слову, особенностей употребления каждого из названных объектов с глаголами активного и пассивного залогов и т. п.

Таблица 1 (продолжение)

<i>0</i>	41	42	43	44	45
III	В конце статьи — перечень идиом с данным словом и комментарий	Иллюстрации (главным образом речения) — не к каждому значению, редко — к стилистическим пометам об особенностях употребления и к синонимическим отсылкам	Статус единицы по отношению к литературному языку определяется в общем виде: <i>slang</i> , <i>colloq.</i> , <i>obs.</i>	—	—
IV	В конце статьи перечень фразеологизмов с комментарием	Иллюстрации даны исключительно литературными цитатами, ко всем употреблениям (по несколько примеров, различающихся хронологией), к каждому параметру (произношение, характеристика объекта и т. д.); каждая иллюстрация документирована (произведение и год)	Каждое значение потенциально может иметь стилистическую ремарку; пометами снабжены также варианты произношения; самостоятельный параметр под знаком <i>STYL.</i> выделяется в тех случаях, когда особым образом характеризуются особенности употребления, напр.: <i>abgrisot</i> , <i>aboucher</i> ...	Тонкие оттенки эмоциональных компонентов употребления комментируются, если они имеются, для каждого значения в словарной статье: <i>acclimatation</i> . . . <i>Au fig., rare, iron</i> . . .	Варианты произношения

Таблица 1 (*продолжение*)

	46	47	48	49	50
III	—	— Не в качестве самостоятельного параметра, а только сведения в <b>деконструкциях и комментариях</b> : <i>gen·darme [F.] One of a body of policemen organized, armed, and drilled as soldiers, in France and some other European countries</i>	Максимальная глубина — OS — Old-Saxon, AS — Anglo-Saxon, OF — Old French. . . beyond. . . AS. <i>begeondan</i> , prep. and adv., fr. <i>be-</i> + <i>geond</i> yond, yonder	Дается недифференцированно с этимологическим параметром: <i>bey. . . [Turki beg]. . .</i>	Выделены отдельным большим разделом в словарной статье под знаком <i>Syn.</i> для каждого значения; от синонимов даны перекрестные отсылки к синонимам синонимов; степень близости значений (степень синонимичности) обозначена постановкой двоеточия перед словом, которое обладает свойством взаимозаменимости с <b>входным</b>
IV	Указана частота каждого слова ( <i>Fréq. abs. litt., Fréq. rel. litt.</i> ), отдельно для XIX в. и отдельно для его употребления в литературе XX в.	— Информация дана <b>имплицитно</b> — в комментариях к употреблениям, при раскрытии этимологического и исторического параметров	Глубина раскрытия — до латыни (где это возможно) или до старофранцузского; в первом случае указаны соответствия данного слова во всех романских языках, если они имеются; на основании общих фонетических закономерностей прослеживается переход к французской форме	— Не как самостоятельный параметр, а только в связи с этимологией: <i>ab ovo, loc. adv. lat. . .</i>	Специально не выделены, но при комментировании отдельных употреблений и иллюстраций синонимы (хотя и не систематически) указываются, т. е. они даны не как самостоятельный структурный параметр, а как средство обычной дефиниции: <i>se donner, se confier, se livrer a</i> — синонимы (вместо дефиниции) к переводному употреблению <i>s'abandonner</i>

Таблица 1 (продолжение)

	0	51	52	53	54	55	56
III	Tак же как синонимы, под знаком Ant., перекрестные отсылки не используются	Омографы отмечены дважды: 1) в соответствующей словарной статье перечислением после синонимов; 2) на входе как самостоятельное слово с необходимыми перекрестными отсылками	—	См. 50, 51	—	—	0
IV	— Tак же как синонимы, специально не выделены, но используются для более глубокого раскрытия семантики, как бы компенсируя недостаточность дефиниции: так, <i>se tenir</i> упоминается как антоним к одному из значений <i>s'abandonner</i>	Омонимы подаются в качестве самостоятельной входной единицы и снабжаются цифровыми индексами	—	—	—	—	0

Таблица 1 (*продолжение*)

0	57	58	59	60	61
	Значения ономастического параметра распределены по 4 отдельным входам: имена знаменитостей; общие английские (американские) имена и фамилии; топонимы мира (с указанием происхождения и местонахождения); названия английских и американских колледжей (с указанием местонахождения)	Задан сокращенным обозначением соответствующей отрасли знаний (Astr., Mus., Chem. . .), а также введением в дефиницию термина или слова, имеющего и терминологическое употребление, элементов энциклопедической характеристики (ср. 37)	— Без комментариев, только пометой obs. к отдельным значениям	—	—
IV	—	Отмечен разветвленным обозначением отраслей знаний (воен., религ., мистич., философ., право, каноническое право. . .), каждое из которых снабжено хронологически упорядоченными документированными иллюстрациями употребления	Этот параметр включает: 1) исторический комментарий (под знаком HIST.) — содержит указание даты первой письменной фиксации, характеристику этапов эволюции семантики, особенностей употребления и т. п. с документированной иллюстрацией каждого изменения; 2) историческое изменение внешнего облика слова — прослеживается также в разделе словарной статьи PRONONC. ET ORTH.; 3) новообразования снабжаются на входе индексом néo.	Подан специальным разделом в словарной статье под знаком BBG. и включает в зависимости от рассматриваемой единицы от 1 до 20 наименований исследований, объектом которых она является	—

Таблица 1 (*продолжение*)

0	<b>62</b>	63	64	<b>65</b>	66	67
<b>III</b>	—	Отдельным входом в виде приложения к словарю — с расшиф- ровкой	Отдельным входом, по алфавиту финалей слов	—	Для характеристики некоторых специаль- ных понятий, относя- щихся к различным областям профессио- нальной деятельно- сти людей, науки, тех- ники	a) 58 б) 37 в) 3
<b>IV</b>	Освещается несистематически, в виде ремарок к отдельным употреблениям и стилистических рекомендаций	—	—	—	В ремарках, раскрывающих стилистические тонкости употребления; <i>Il y a similitude partielle de contenu entre abstraction «pouvoir d'abstraction» et abstractivité; mais abstraction souligne le caractère actuel du pouvoir en question, abstractivité souligne l'aptitude générale</i>	a) 59 б) 32 в) 7

Таблица 1 (*продолжение*)

7 ю. к. Гарячев	0	1	2	3	4	5
V. Oxford Dictionary	Английский	Один: алфавитный прямой	1) Хронологические рамки словника: IX в.—XX в. (в иллюстрациях нет примеров позднее 1900 года, однако отсутствие иных хронологических по-мет — см. ниже — позволяет идентифицировать слово как употребляющееся в том же значении и ко времени выхода словаря в свет); 2) время первой письменной фиксации каждого отдельного значения (указаны год и произведение); 3) время «последней» письменной фиксации: для слов, вышедших из употребления и помеченных на конце знаком †; 4) окказионализмы, т. е. слова с единичной фиксацией — for the попсе	Более 400 тыс. слов	Побуквенно, с указанием этапов становления современной буквенной формы слова по векам: beauti- ful... forms: <i>beaute-</i> , <i>beuti-</i> , <i>beuty-</i> , <i>bewti-</i> , <i>bewtyfull</i> , <i>beuty-</i> , <i>butyful</i> , <i>beauti-</i> , <i>beautiful], beautiful</i>	
VI. Grand Larousse	Французский	Один: алфавитный прямой	1) Время первой письменной фиксации: <i>carte I tōgſeau de carton</i> (XV-e s.); <i>II carte à jouer</i> (1398); <i>III carte de géographie</i> (1636)	Около 200 тыс.	Побуквенно, с вариантами написания, причем редкие даты в конце статьи ( <i>Rem. ambroisiēn, -enne</i> — <i>ambroisiēn, -enne</i> ), а частые — в начале: <i>agalactie</i> ou <i>agalaxie</i>	

Таблица 1 (продолжение)

0	6	7	8			9	10	11
V	—	a) Указано в транскрипции постановкой точки после ударного слога: Agate (æ·gët), Hobby (hö·bi); б) для вышедших из употребления слов — в орфографической записи: Alfe·res, A·nredly	Полная транскрипция на основе специально разработанной системы обозначений, с вариантами произношения: allegiance (ali·dzans, alΓ·dʒləns) либо с подачей варианта на входе: shak-shak, variant of chac-chac; слова, вышедшие из употребления (+), сведениями о произношении не снабжаются	—	Adj., adv., conj., int., pron., prep., sb.	—	0	
VI	—	—	Полная транскрипция на основе международного фонетического алфавита с указанием вариантов произношения: albatros [albatros ou albatros]	—	Adj., adv., conj., n. m. (nom masculin), n. f.	—	Только для имен существительных, одновременно с указанием принадлежности к этой части речи: п. т., п. ф.	
0	12			13	14	15	16	
V	Обозначены особые случаи образования множественного числа, отклонения: Antenna... Pl-æ, garely -as.	0	0	0	Указываются при наличии колебаний в образовании: beautiful... Occas. compared with -er, -est, usually with more, most	0	Tr., intr. Admire, .. v. 1. intr. To feel or express surprise. 2. tr. To view with wonder or surprise...	
VI	1) Для имен singularia или pluralia tantum: branchiopodes — п. м. PL; 2) в случаях отклонений от стандартных способов образования множ. числа: bec-de-corbin Pl. des becs-de-corbins; bail Pl. des baux [bo].	На входе указаны наряду с окончаниями формы м. р. также и окончания ж. р.: assimilateur, -trice; ambrosien, -enne; avant-courier, -ere	ø	ø	На входе, только для супплетивных форм: meilleur, -e I. Comparatif de l'adjectif «BON» II. «Le meilleur», «La meilleure» — formes de superlatif de «BON». Кроме того, охарактеризовано употребление коррелирующих форм: plus bon, beaucoup meilleur, loien meilleur	0	v. tr.; v. intr.	

Таблица 1 (*продолжение*)

0	19	20	21	22	23	24
V	—	Обозначено подразделениеми словарной статьи внутри непереходного значения, т. е. как часть definicji: <i>admire</i> . . . a. simply (т. е. без предлога, — <i>IO</i> , <i>K.</i> ) obs. . . b. with <i>at</i>	Указывается как компонент лексического значения в соответствующей части definicji: <i>balance</i> . . . v. . . 5. To steady. . . b. <i>refl.</i> and <i>inir</i> . To keep oneself in equilibrium	—	—	—
VI	V. impers. (pleuvoir — t. impers.)	Visiter I. quelqu'un. . . II. quelque chose. . . croire v. tr. I. qch, et l'infinitif, que. . . II. qn. . . III. qn ou qch v. intr. I. a. . . II. en. . . IV. de. . .	V. pr. (v. pronominal) visiter: se ~ v. pr.; croire: se ~ v. pr.	—	—	Только в особых случаях, когда образование и написание императивных форм может вызывать трудности: aller — s'en ~ . . . L'imperatif s'écrit va-t'en, allons-nous-en, allez-vous-en
0	25	26	27	28	29	30
V	На вход	—	Только пометы о типе спряжения: v. <i>str.</i> , <i>or go</i> . — verb strong, or weak	0	—	Даны на входе, если употребляются как прилагательные (ppl. a., participle adjective): balanced — ppl. a., balancing — ppl. a.
VI	На вход	—	Парадигмы глагольного словоизменения сведены в таблицы в начале словаря; на входе в словарь глаголы, имеющие трудности в спряжении, снабжены индексами, отсылающими и соответствующей парадигме	9	Употреблению глаголов в этой роли отводится специальный раздел статьи: <i>avoir</i> — I, II, III. . . v. auxil. — <i>avoir se construit avec le part. passé du verbe pour exprimer l'action accomplie: J'ai dit, Il a fait</i> . . .	Дан на входе отдельной статьей, если употребляются как прилагательные: su, sue — adj. (part. passé de <i>saTOir</i> )

Таблица 1 (продолжение)

100

	0	32	33	34	35
V		Pra. t. — past tense, форма подается на вход только в случае отклонений — для неправильных глаголов: same... pra. t. of COME v. ...	—	1) Приставки, суффиксы, части сложных слов поданы на вход самостоятельно в общем алфавитном порядке словарника; 2) в разделе словарной статьи, посвященном происхождению, в случае производного слова показана его деривация: Beeriness. [f. Beery + -ness]; Coachmanship. [f. Coach + Man + -ship]	Дериваты даны на входе отдельными статьями, но без отсылочных помет (т. е. отсутствует указание словообразовательных связей); направление деривации можно установить на основе морфологического членения (см. 34); иногда — также и на основе definizioni: asceti-cally. After the manner of an ascetic
VI	—	—	0	Элементы словоизводства даны трижды: 1) в виде отдельного входа в начале словаря (до основного словарника) — список суффиксов существительных, прилагательных и т. д.; префиксов — по происхождению (лат., греч., фр.); компонентов сложных слов с указанием их происхождения, принадлежности к grammaticalской категории, способа соединения и характеристики результата соединения; 2) в этимологической части словарной статьи: antilogie (gr. antilogia, de <i>anti</i> — contre et de <i>logos</i> — discours); 3) отдельной словарной статьей на вход	Словообразовательное гнездо задано универсальным перечнем отношений внутри него, которые даны отдельным входом в предисловии к словарю: La construction des unités lexicales dans la perspective synchronique. I. La suffixation: I. A. La relation nom → verbe I. B. La relation adjetif → verbe I. C. La relation verbe → nom II. La préfixation III. La composition ....

Таблица 1 (*продолжение*)

	36	37	38
V	<p>1) Отдельные значения обозначены арабскими цифрами (1, 2, 3, сплошной нумерацией, независимо от других группировок), в основе их упорядочения лежит принцип семантической деривации: от более общего (независимо от того, вышло оно из употребления или нет) — путем импликации к более узким, переносящим (fig.) или специальным;</p> <p>2) значения подразделяются на употребления с оттенками (а, б, с. . .);</p> <p>3) значения группируются в более крупные комплексы (I, II, III. . .) по генетическому принципу</p>	<p>Филологическое определение, с широким использованием синонимов: <i>ascending</i>, . . . <i>The action of the verb Ascend; ascent, ascension</i>. . .</p>	<p>1) Отсутствие пометы означает обычное употребление;</p> <p>2) lit. — literally (букв.);</p> <p>3) fig. — figuratively (перен.)</p> <p>Agate. . . I. A precious stone. . . 2. fig. A very diminutive person, in allusion to small figures cut in agates for seals</p>
VI	<p>Отдельные значения обозначены римскими цифрами и применяются два способа их упорядочения:</p> <p>1) семантическая деривация — carte (см. № 3);</p> <p>2) если семантическая деривация не прослеживается, то значения располагаются в порядке их временной фиксации: <i>cageau</i>: I. <i>Trait d'arbalete</i>. . . (1070); II. <i>Objet à quatre côtés</i>. . . (XII); III. <i>Dessin</i>. . . (XVI)</p>	<p>1) Филологическое определение;</p> <p>2) энциклопедические сведения о лингвистических понятиях и грамматике французского языка: <i>accents</i>, <i>accord</i>, <i>actualisation</i>, <i>morphologie</i>, <i>versification</i>, <i>le champ sémantique</i>. . .</p>	<p>Дана очень тонкая разработка параметра транспозиции значения: помимо генерального противопоставления litér. — fig., внутри последнего дана дифференциация по способу переноса значения — <i>métonymie</i>, <i>par opposition</i>, <i>par analogie</i>, <i>par allusion</i>, <i>par extension</i>. . .</p>

Таблица 1 (продолжение)

0	39	40	41
V	Чисто диалектные слова не включены в словарик; учитываются только те, что ныне сохранились в диалектах, а прежде были общеупотребительными; помета — <i>dialectal</i> , (dialectal), с крупными подразделениями: north. — <i>northern</i> , e. <i>midl.</i> — east midland, w. <i>midl.</i> — west midland, s. w. — south western; из помет о территориальном распространении формы либо значения употребляются только две — <i>Sc.</i> (Scotch) и <i>U. S.</i> (United States)	Минимальная информация о роли в предложении: noun of action, noun of agent, object, attrib., subject, . . .	На ключевое слово: 1) в разделе словарной статьи, названной <i>Combinations</i> , даны штампы, крылатые выражения: <i>Jack in a box, Sea Horse. . .</i> ; 2) в разделе <i>Phrases</i> — идиоматичные употребления слова: <i>catch me! or catch me at it: a phrase-expressing emphatically that one will never be found doing a thing (colloq.)</i>
VI	Отдельные диалектные слова, встречающиеся в прессе и современной художественной литературе, снабжены пометой <i>dialectal</i>	Подробно разработана сочетаемость для многозначных и высокочастотных глаголов: <i>convenir</i> . . . v. tr. ind. (-irect). . . «convenir à»: <i>avoir de la convenience avec</i> . I. <i>Avec un sujet désignant une chose</i> . . . II. <i>Avec un sujet désignant un être animé</i> . . . III. <i>Employé impersonnellement</i> . . . «convenir de, que»: <i>tomber d'accord sur, que</i> . I. <i>Avec un sujet désignant une ou des personnes</i> . . . II. <i>Employé impersonnellement</i> . . .	Идиомы и устойчивые словосочетания даны без помет, в общем алфавите словарика по первому слову речения: <i>chemin de fer</i> , либо по ключевому слову — с толкованием, комментарием и иллюстрацией: <i>fair du charme, fair la cour à quelqu'un. . .</i>

Таблица 1 (*продолжение*)

0	42	43	44
V	Документированные цитаты с указанием "года" и источника — для каждого значения отдельно, в хронологическом порядке, от первой письменной фиксации слова	Развернутая дифференциация признаков: slang, pop. (popular), colloq. (colloquial), university colloq., commercial slang, nautical slang, public school slang, ...	Несистематические пометы: jocular, euphemistically
VI	1) Литературные цитаты с указанием только автора (без точного адреса); 2) речения и словосочетания	Стилистические характеристики распределяются по трем сферам речи — книжной (littér. — littérature, littéraire, acad. — académie, rhét. — rhétorique, scolast. — scolastique, didact. — didactique), разговорной (pop. — populaire, fam. — familier, triv. — trivial, vulg. — vulgaire) и арготической (arg. — argot, arg. mil. — argot militaire, arg. scol. — argot scolaire, arg. du sport)	Специально не выделен, все эмоциональные оттенки употреблений включены в definiciji соответствующих значений, где и комментируются; используется только одна помета: ironiq.

Таблица 1 (*продолжение*)

	45	46	47
V	<p>Нормативный параметр в строгом смысле слова не представлен, так как словарь построен на дескриптивистских, а не на прескриптивистских принципах, однако спорадически встречаются те или иные признаки нормативного параметра:</p> <p>1) в дефинициях и иллюстрациях используется специальная помета, которой обозначаются ошибочные (catachrestic), путанные, неточные употребления: <i>Balancé</i>... 17. To ballast. Obs. rare;</p> <p>2) употребляются пометы: erron. (erroneous), ellipt. (elliptical), euphem. (euphemistically), mod. (modern) occas. (occasional);</p> <p>3) обозначены допустимые варианты произношения (<i>ægarik</i>, <i>agærik</i>) и правописания (<i>Agast</i>, <i>aghast</i>)</p>	<p>В чистом виде не представлен, отмечены лишь употребления с минимальной частотностью (единичные): <i>Admōtion. Obs. rare</i>—, где индекс —1, —0 означает, что в картотеке зафиксирован лишь один или даже нет ни одного примера употребления этого слова</p>	—
VI	<p>Словник и примеры отобраны в соответствии с существующей нормой, поэтому неправильности и ошибочные употребления не отмечены; в ряде случаев даны варианты: <i>abiétacées</i>... Rem. On a dit <i>naguère abiétinées</i>; степень употребительности охарактеризована двумя пометами: <i>inusite</i>; <i>peu usité</i></p>	<p>Статистические сведения даны отдельным входом и не полностью, т. е. не для каждой входной единицы: в начале словаря (с. LXXXIII—XC) дан список 1063 наиболее употребительных слов (с частотой 20 и более на массиве 163 текстов); указаны ранг слова, число текстов, в которых данное слово встречается и абсолютная частота; напр., I. Stre (163; 14083), 609, <i>apergevoir</i> (30; 42), 1031, <i>goût</i> (18; 20)</p>	<p>Хотя лингвострановедческая информация специальным образом не выделена и не сконцентрирована, т. е. ее нельзя извлечь из словаря в виде самостоятельного параметра, однако сведения об обычаях, особенностях быта, традициях поведения и т. п. обильно включены в иллюстративный материал, представлены в цитатах из художественной литературы</p>

Таблица 1 (продолжение)

0	48	49	50
V	<p>Максимальная глубина этимологизации — до древнегерманского и готского или древнефранцузского и латыни (греческого), с указанием форм в некоторых родственных языках:</p> <p>1) Abysm. [a. OFr. <i>abîme</i>, <i>abîme</i> (<i>cognâte</i> with Provengal <i>abisme</i>, Spanish <i>abismo</i>): — late pop. L. <i>*abyssimus</i>, a superlative of <i>abyssus</i>, lit. the profoundest depth.. .</p> <p>2) Acre. OE. <i>acer</i>, <i>acer</i>, cogn. w. Goth. <i>akr-s</i>, ON. <i>ager</i>, OSax. <i>accar</i>, OFris. <i>ekker</i>, OHG. <i>achar</i>; L. <i>ager</i>, Gr. <i>ἀγρός</i>, Skr. <i>ajras</i> plain; originally open country, untenanted land, forest . . .]</p>	<p>Глубина раскрытия заимствования — до орфографии (которая в нижеследующих примерах опущена), произношения и исходного значения в языке-источнике, с указанием степени натурализации в английском, где знак II свидетельствует о неполной натурализации:</p> <p>1) Agape (æ'gæpi) Pl. agapæ, -ai, rarely agapes. Gr. <i>ἀγάπη</i> — brotherly love. . .</p> <p>2) Alcohol (æ'lkohol). [a. med. L. alcohol, ad. Arab. <i>al-koh'l</i> 'collyrium', the fine powder used to stain the eyelids, f. . . <i>kahala</i>, Heb. . . <i>kākhāl</i> to stain, paint. . .]</p>	(Только как средство дефиниции)
VI	<p>Глубина этимологизации — до латыни: <i>accablier</i> [akable] v. tr. (comp. d'une form normande <i>cable</i>, de l'anc. franc. <i>chaable</i>, du lat. pop. <i>*catabola</i>, machine à lancer des pierres; XIV-e—XV-e s., «abattre», d'où «écraser» [encore au XVII-e s.]</p>	<p>Маркированы указанием языка (Anglais, Arabe, Latin, Portugais. . .); глубина раскрытия заимствования — до орфографии в языке-источнике:</p> <p>1) 2. car [kar] n. m. (abbrév. de autocar; XX-e s.; au XIX-e s., car, mot angl. issu lui-même de car, forme normande de char, désignant une voiture sur rails [1873, Hubner]);</p> <p>2) lazzarone ou lazaron, lazaron n. m. (napolitain lazaron, mendiant, de lazaro, ladre, lépreux [v. LAZARET]; 1781, Brunot, écrit <i>lazzaron</i>; laz(z)aron, fin du XVIII-e s. [var. <i>lazares</i>, milieu du XVII-e s., et <i>lazarelli</i> — plur., — 1739, Brosses])</p>	<p>Синонимам отведен специальный раздел словарной статьи — в ее конце, причем к каждому значению слова указана своя синонимическая группа: <i>achalandage</i>. . . Syn.: 1. <i>clientèle</i>; 2. <i>approvisionnement</i>, stock</p>

Таблица 1 (продолжение)

	51	52	53	54	55	56
V	—	Не имеют специальных помет, а даны на входе отдельными статьями, без всяких индексов, лишь с указанием на различия в произношении, грамматических характеристиках, деривации или этимологии: Agate (æ:gĕt), sb. . . Agaté (Sgei·t), adv. . .	—	—	—	0
VI	Даны в одном разделе с синонимами: autre. . . Syn.: 1. dissemblable; 2. nouveau, second. — Contr.: 1. identique, même, pareil	Различение омонимии проводится очень строго и используемые обозначения помогают дифференцировать типы омонимов: а) омографы снабжаются цифровым индексом и различаются указанием их грамматической принадлежности: 1. boucher n. т., 2. boucher v. tr.; б) абсолютные омонимы также снабжаются индексами (арабские цифры) и различаются семантической характеристикой: 1. air=fluide, 2. air=manière d'être, 3. air=melodie; и) если семантические расхождения не являются полными, то омонимы даются как бы одной статьей, но разделенной римскими цифрами на самостоятельные части: aile: I. Organe de vol. II. Ce qui a la forme d'aile. III. Partie latérale (architecture)	—	—	—	0

Таблица 1 (продолжение)

0	57	58	59	60
V	<p>—</p> <p>Учтены лишь дериваты от имен собственных, имеющие статус самостоятельных слов:</p> <p>1) Ascham (<i>æskám</i>) named from Roger Ascham, author of <i>Toxophilus</i>. A sort of cupboard. . .</p> <p>2) Henrician (<i>henri-ʃan</i>), a. and sb. <i>Eccl. Hist.</i> [ad. med. L. <i>Henrīciān-us</i>, f. <i>Henricus Henry</i>]. . .</p>	<p>Развернутая маркировка слов, значений и отдельных употреблений по широкому кругу областей человеческих знаний и интересов: anatomy, architecture, astrology, crystallography, horticulture, metaphysics, mineralogy, theology, phrenology, ornithology, . . .</p>	<p>История развития форм и история развития значений рассматриваются: а) в разделе словарной статьи, посвященной этимологии, б) в семантической части дефиниции и в) поэтапно — в иллюстративной; например, по поводу эволюции произношения слова <i>Abyss</i> сделано тонкое наблюдение: [...] <i>Probably abisme was at first merely an artificial spelling, in imitation of the Fr.; we find abisme rhyming with time as late as 1616; the modern pronunciation follows the spelling</i></p>	
VI	<p>—</p> <p>Только дериваты от имен собственных: <i>api</i>. (<i>francisation du nom de Claudius Appius, qui aurait apporté cette pommes du Péloponnèse à Rome</i>. . .)</p>	<p>В словарик включено большое число научно-технических терминов, однако пометы даются недифференцированно: <i>scientifique</i> — для общенаучных, <i>spécialement</i> — для уникалий и технических терминов (значений); исключение сделано лишь для следующих областей знаний, которые отмечены особо: <i>Juridique, linguistique, mathématique, militaire, médical</i></p>	<p>История слова дана в редуцированном виде, без сведений о его семантической эволюции:</p> <p>— этапы становления орфографического облика представлены в этимологическом отделе словарной статьи: <i>accoutume</i> (<i>de coutume; 1190, le Livre des Rois, écrit acustumer</i>) — см. также пример в № 49;</p> <p>— соответствующий исторический период развития языка обозначен пометой: <i>ancien, vieux, médiéval, classique, néologisme</i></p>	

Таблица 1 (окончание)

	61	62	63	64	65	66	67
V	Как самостоятельный параметр, отражающий фиксацию данного слова другими словарями, не представлен; однако спорадически даны отсылки к словарям или энциклопедиям, из которых заимствована та или иная информация о данном слове: (Jm.) — in Jamieson, Scottish Dictionary; (R) — in Richardson's Dictionary; (D) — in Davies (Suppl. Eng. Glossary) и т. д.	—	—	—	Отмечены тремя способами: 1) либо в этимологической части: Balance . . . [a. F. balancer, f. balance] sb. (Like the sb., in certain senses confused with ballast); 2) либо в definитивной под знаком π: Balance . . . π <sup>17</sup> . To ballast. Obs. rare; Cf. BALANCE sb. π; 3) либо верbalным образом в самой definии: Agate . . . 5. Sometimes erroneously contused with gagates, jet	—	a) 58 б) 31 в) 8
VI	—	—	—	—	—	—	a) 59 б) 36 в) 6

## § 3. ПУТИ АВТОМАТИЧЕСКОЙ ОБРАБОТКИ ЛЕКСИКОГРАФИЧЕСКИХ ПАРАМЕТРОВ

Движенья нет, сказал мудрец брадатый. . .  
A. С. Пушкин

### Параметризация и автоматизация

Цель предлагаемого обзора состоит не в том, чтобы с достаточной полнотой охарактеризовать результаты и возможности автоматизации разного рода лексикографических работ. Автор ограничивается попыткой определить место осуществленного эксперимента по машинному построению русского тезауруса в общем потоке исследований в области автоматической лексикографии. Электронно-вычислительная техника создавалась не для лексикографии, однако ее применение в решении многообразных задач — и прежде всего информационного плана — показало, какая важная роль отводится в этом деле словарю. Таким образом, почти одновременно встали и проблема «словарь для машины», и проблема «машина для словаря». Это произошло примерно четверть века назад, и, пожалуй, главным итогом прошедших лет стало то, что к настоящему времени изменился самый статус словаря: он перестал считаться экземплярным достижением пауки. Если раньше осуществление крупных лексикографических предприятий (таких, как Оксфордский словарь, Уэбстерский, Большой академический словарь русского языка — БАС) занимало 30—50 лет, то теперь время подготовки словаря в значительной степени сокращено, причем многие этапы работы над ним передаются машинам. Даже если оставить в стороне всю вычислительную лексикографию и необозримое море специальных словарей, используемых в разнообразных автоматизированных информационных системах<sup>134</sup>, все равно приходится констатировать, что за последнюю четверть века резко возросло количество словарей в мире и увеличилось их разнообразие. В этих условиях в основу нашего обзора можно было бы положить указанное разграничение (т. е. «словарь для машины» и «машина для словаря»), пойдя по пути рассмотрения теоретических принципов и технических приемов, с одной стороны, вычислительной лексикографии, с другой —

---

<sup>134</sup> См., напр.: Марчук Ю. Н. Вычислительная лексикография. М., 1976.

автоматизированной обработки лексикографической информации в процессе подготовки словарей различного типа для использования их человеком. Так могла бы возникнуть — в дополнение к многочисленным существующим — еще одна типологическая классификация словарей.

Нам представляется, что поставленной цели можно достичь более экономным способом, если давать характеристику основных направлений автоматизации не по типам словарей, а по параметрам словарной структуры, т. е. по конституирующем лексикографическим характеристикам.

Рассматривая особенности нынешнего словаростроения и движение современной лексикографии, мы определили выше основную линию ее развития как тенденцию к лексикографической параметризации языка, как стремление закреплять в словарной форме результаты изучения всех уровней языковой структуры, представлять в виде лексикона все языковые единицы содержательного и формального плана и отношения между ними, экстраполировать в словарь результаты анализа самых разнообразных языковых явлений. И дело здесь не только в том, что словарь оказывается самой удобной формой фиксации наших знаний об изучаемом объекте. Скрытой, возможно, не всегда осознаваемой, но мощной пружиной отмеченной тенденции является то, что она закладывает базу для максимально формализованных описаний естественного языка. А в наши дни все более широкое распространение получает представление, что формализация языка скорее всего достижима на пути создания всесторонних (т. е. разноспектральных) и исчерпывающих инвентарей его единиц, процессов, явлений, в числе которых предусматриваются инвентарии семантические, лексико-семантические, морфо-семантические, семантико-статистические и др.<sup>133</sup>

Принцип лексикографической редукции, лексикографической интерпретации отдельных языковых явлений, преломляясь в реальные словари различных языков, позволяет обнаружить две разнонаправленные линии расширения и роста словарей. Центростремительная линия характеризуется целевой установкой на создание универсального словаря, охватывающего в идеале все сведения о данном языке, в том числе грамматические, синтаксические и экстралингвистические. Центробежная линия, наоборот, связана с построением однопараметровых словарей, определяется стремлением лексикографировать каж-

дое отдельное явление данного языка, каждое измерение данной языковой структуры. В качестве примера можно сослаться на имеющиеся почти в каждом письменном языке орфографические словари (параметр побуквенного состава слова, или орфографии), орфоэпические словари (параметр произношения), словообразовательные, синонимов, словоизменительные и др.

Таким образом, под параметром в самом общем виде понимается способ лексикографической интерпретации того или иного структурного элемента или функционального явления языка и их экстралингвистических соответствий. В предыдущем разделе мы в pragматическом смысле определили лексикографический параметр как некоторый квант информации о языке, который в экстремальном случае может представлять для пользователя и самостоятельный интерес, но, как правило, выступает в сочетании с другими квантами (параметрами) и находит специфическое выражение в словарях. Отметим здесь некоторые особенности параметров, существенные для нашего дальнейшего изложения.

Прежде всего, «экстраполяция в словарь» того или иного явления может осуществляться не в том виде, как оно традиционно фиксируется при описании данной языковой структуры, и в этом отношении параметр не обязательно соотносится однозначно с языковой единицей или категорией. Далее, в содержательном аспекте с точки зрения передаваемой ими информации все лексикографические параметры можно разделить на две группы. В одну из них войдут собственно языковые, т. е. структурогенные — ударения, орфографический, произносительный, словоизменительный, словообразовательный, категориальные, сочетаемости и др. Они дискретны по своей природе, возможность ихарьования распространяется только на способы их задания, фиксации в словарях. О таких параметрах мы говорим, что они обладают «значением»: так, параметр «род» может принимать три значения в русском (м., ж., ср.), два значения во француз-

---

<sup>135</sup> Городецкий Б. Ю. Теоретические основы прикладной семантики. Автореф. докт. дис. М., 1978, с. 10, 12, 13, 21; Он же. Семантические проблемы построения автоматизированных систем обработки текстовой информации. — В кн.: Вычислительная лингвистика. М., 1976, с. 28 и след.; Викс Ю. А. Классификаторная модель эstonской морфологии (автоматический синтез глагольных словоформ). Автореф. канд. дис. Таллин, 1978.

ском (т., ф.) и полностью отсутствует в английских существительных; параметр «*словоизменение существительных*» предполагает приписывание имени одного из падежных аффиксов, а задан он может быть в русском языке не только как набор 12 служебных морфем, но и с помощью ключевых форм, на основе которых порождается полная парадигма (для большинства имен в русском языке такими ключевыми формами являются, помимо данной на входе во всякий словарь русского языка формы имениительного падежа единственного числа, также род. п. ед. ч. и им. п. мн. ч.).

Другую группу составляют параметры, содержание которых включает по необходимости и экстралингвистический фактор — денотативный, историко-культурный, pragматический, т. е. моменты, вторичные по отношению к самому языку, связанные не только с языком, но с его бытованием и изучением, т. е. языкоznанием. Эти параметры недискретны, характеризуются изменяющейся глубиной раскрытия и отражают не собственно структурные отношения, а процессы — диахронические, психолингвистические, синтагматические, взаимодействия языков, интерференции и т. п. В противоположность собственно языковым, или структурогенным, мы называем их лингвистическими и относим к их числу параметр родства (этимологический), страноведческий, стилистический, библиографический (сведения об исследованиях по поводу данного слова, отраженные, например, в упомянутом в сносках 123 словаре французского языка), хронологический (дата первой письменной фиксации), частотный, экземплярно-иллюстративный и др.

Параметры могут находиться в отношениях взаимной эквивалентности. Так, параметр дефиниции, помимо обычного толкования значения слова, может быть адекватно представлен синонимами (а синонимия — это один из самостоятельных системных параметров: ср., например, синонимический словарь); примерами (экземплярами) или иллюстрациями употребления данного слова (ср. словарь-конкорданс, где этот параметр приобретает экстремальное выражение и значение задано перечнем всех контекстов)<sup>136</sup>; «неявным определением»<sup>137</sup> в виде набора элементов соответствующего семантического поля (ср. идеографический словарь); переводом на другой язык (ср. двуязычный словарь); дополняющим текст рисунком («картинный параметр»), который в совокупности с этим текстом представ-

ляет собой разновидность оstenсивного определения<sup>138</sup>; наконец, в словарях сочетаемости мы находим еще один параметр, эквивалентный рассматриваемому, — указание синтаксической валентности слова. Содержание, передаваемое каждым из этих параметров по отношению к одному и тому же слову, не тождественно во всех этих случаях, но в значительной степени, очевидно, является общим, почему мы и имеем основания говорить об их эквивалентности.

Вероятно, нам нет смысла подробно останавливаться на способах автоматической обработки и анализа каждого параметра. И не только потому, что разбор 67 пунктов потребовал бы много места и отнял бы много времени у читателя. Дело в том, что сами приемы анализа в ряде случаев повторяются, а известная часть параметров и обрабатывается так же, как бывает задана — синкетически. Поэтому поступим следующим образом: с определенной долей условности разобьем валовой состав параметров на три группы.

В качестве основания разбиения примем три формальных критерия:

- технологический критерий степени препарированности параметров для машинной обработки и, соответственно, — степени сложности предмашинного анализа для выделения самого параметра и установления наилучшего способа его представления в машинных процессах;
- источник получения параметра — текст и его разновидности, словарь или специальные лингвистические исследования;

---

<sup>136</sup> В логике существует, в частности, понятие «экземплярного определения», состоящего в приведении примеров. См.: Горский Д. П. Определение (логико-методологические проблемы). М., 1974, с. 90.

<sup>137</sup> Различие явных и неявных определений проведено французским математиком Ж. Д. Жергонном: «Явными называются определения, в которых смысл определяемого термина полностью передается через смысл определяющих терминов. В неявных определениях отсутствует тождество смыслов определяемого и определяющих терминов; смысл определяемого термина здесь передается лишь частично, путем установления его отношения к другим терминам языка» (цит. по: Попа К. Теория определения. М., 1978, с. 53).

<sup>138</sup> Под оstenсивным определением понимают прием установления смысла языкового выражения путем одновременного произнесения слов и указания на обозначаемый ими предмет (Горский Д. П. Указ. соч., с. 83 и след.).

— цель автоматической обработки — постр<sup>о</sup>ение словарей, получение новой информации о языковой структуре и синтезирование текстов.

Тогда в первую группу войдут параметры, требующие только широкой компиляции материала, домашинная обработка которых минимальна, а решение проблем их выделения и обобщения не превышает по своей сложности задач простой сортировки. К этому ряду следует отнести орфографический параметр, количественный (подсчет числа слов в словаре или вообще подсчет каких-то элементов словаря), хронологический, параметр длины слова, рифмы (скажем, словарь рифм поэта или вообще совпадающие финалы слов, а шире — обратный алфавитный порядок следования букв), аббревиатуры, экземплярно-иллюстративный параметр, статистический и некот. др. Источником для их извлечения служат тексты, а целью их обработки является получение определенной базы данных. Назовем их компилятивными.

Вторую группу составят параметры, автоматическая обработка которых предполагает решение определенных аналитических задач, и объединяют они такие элементы информации о языковой структуре, которые воплощаются в категориях и формах. Сюда мы включаем параметр удаления и связанные с ним параметры произношения и слогоделения, часть речи, вид глагола, переходность, управление, залоги, причастие, будущее время, прошедшее время, морфологическое членение слова, многозначность — однозначность, транспозицию значения, ареальный параметр, фразеологический, стилистический и некот. др. Параметры этой группы выделяют на основе словарей (толковых, синонимических, идеологических, словообразовательных, грамматических), а также специальных языковедческих исследований, и результаты их автоматической обработки предназначены для анализа текста и построения новых типов словарей.

Границу между второй и третьей группами параметров провести труднее, но общий принцип здесь таков, что усложнение идет по линии комплексности параметра. Так, если в предыдущей группе каждый параметр воплощается в той или иной стандартной форме (причастие, вид глагола, морфологическое членение слова) или в модели, отражающей стандартный способ получения такой формы (управление глагола, будущее время, переходность), то параметры, относящиеся к третьей группе, предполагают

развертывание уже в целый набор форм (или слов), представляют собой своего рода «парадигму», понимаемую в широком смысле этого термина, т. е. позволяют порождать любую форму или комплекс соответствующих форм. В этом смысле параметры третьей группы можно назвать «конструктивными», поскольку, будучи образованы по законам построения «новых лингвистических объектов» (см. Введение), они способны не только обобщить известное, но и генерировать новую информацию о языке (ср. фиксацию в словоизменительном словаре таких форм, которые не засвидетельствованы в текстах, но образование которых не противоречит законам системы, и потому они как бы предсказаны соответствующей парадигмой: *блудши, волгнув, гния, ехав, жалев, кажась, могши, сидев*<sup>139</sup>; ср. также выявление «пустых клеток», т. е. потенциальных слов в словообразовательной системе русского языка<sup>140</sup>).

Последующее изложение в этой главе определяется охарактеризованным разделением параметров, так что каждой из названных групп — компилятивной, аналитической и конструктивной — посвящается специальный раздел. И разбиение параметров на три группы, и дальнейшее их рассмотрение в связи с автоматизацией лексикографических работ сделано в основном с ориентацией на русский язык. Это важно подчеркнуть, поскольку параметры, будучи в какой-то части универсальными, обладают вместе с тем и национальной спецификой, а это значит, что мы в своем анализе не будем касаться, например, параметра «лексический класс», свойственного языкам с классными показателями, или параметра «породы глагола», представленного в семитских языках. Тем не менее, такое сплошное обследование лексикографических параметров одного языка, с достаточно широким привлечением материала и других индоевропейских языков, дает, как нам кажется, основания утверждать, что оно воссоздает довольно полную картину осуществляемого в настоящее время изучения метаязыка словаря с помощью электронно-вычислительной техники. При этом под метаязыком словаря в широком смысле понимается совокупность лексикографических параметров, отражающая всю информацию,

---

<sup>139</sup> Зализняк А. А. Грамматический словарь русского языка. М., 1977, с. 98—130.

<sup>140</sup> Тихонов А. Н. Школьный словообразовательный словарь русского языка. М., 1978.

передаваемую словарем о структуре данного/языка, его истории, распространении, функционировании в настоящее время и изучении. Тогда метаязыком в узком смысле этого термина мы назовем то, что традиционно под ним и понимается, — язык описания значений в толковом или переводном словаре, язык дефиниций. Непосредственно этому предмету — анализу языка дефиниций — посвящены III и IV главы книги, а метаязыку словаря в широком смысле отводится здесь лишь роль общего фона, к характеристике которого мы и переходим.

### **Компилятивные параметры и банки данных**

Как было отмечено выше, домашинная обработка параметров этой группы минимальна и ограничивается некоторыми формальными процедурами, связанными с подготовкой материала к вводу в машину. Общее движение исследования в этом случае — от текста к словарю. Вероятно, одной из первых лексикографических процедур, подвергшихся автоматизации, было составление словарика, т. е. перечня всех слов, встретившихся в тексте (текстах). Причем на заре автоматизации проблему составляло даже получение простого алфавитного списка слов данного текста с подсчетом частоты их встречаемости, не говоря уже о составлении словоуказателей к текстам, включающим информацию об адресе данного слова — томе, наименовании произведения, странице, строке<sup>141</sup>. К настоящему времени автоматическое получение словарников, которое является первым необходимым шагом в построении частотных, обратных словарей, конкордансов и в перманентном ведении, пополнении картотеки национального словаря, представляет собой простейшую задачу. Предварительная разметка текста при ее решении предполагает специальное обозначение конца строки, конца строфы (в поэзии), номера страницы, конца абзаца (в прозе) или монолога, реплики (в драме), выделение в текстах слов типа *глава, действие, эпилог, сцена, предисловие*, а также названия произведения<sup>142</sup>.

Подобный словарник или словоуказатель, с одной стороны, хотя и содержит всего лишь перечень слов, с точки зрения параметризации не является элементарным параметром, а объединяет целый комплекс их: побуквенный состав слов (т. е. орфографический параметр), частотный, хронологический (поскольку обработке могут быть под-

вергнуты тексты, охватывающие только определенный исторический период), а также имплицитно присутствующие дистрибутивный и длины слова (количество букв в слове). Каждый из перечисленных параметров в зависимости от поставленной задачи может стать доминирующим, и тогда мы имеем дело либо с орфографическим<sup>143</sup>, либо с обратным<sup>144</sup> и т. п. словарем.

С другой стороны, словоуказатель — это вспомогательный для лексикографа материал, а собственно словарь начинается с картотеки. В картотеке слову соответствует от одной до нескольких десятков карточек, на каждой из которых зафиксирован тот или иной контекст этого слова. Частным случаем такой картотеки является конкорданс к произведению или совокупности произведений одного автора. Общие филологические словари составляются, естественно, не на материале одного автора, и лексикографы прибегают к выборкам. Выборки, сделанные с помощью конкордансов, способны более адекватно отражать и обобщать особенности словоупотребления той или иной эпохи, чем выборки, осуществляемые на основе опыта и интуиции лексикографа. Как считает Д. Кубурлис, составивший на ЭВМ несколько конкордансов, «если имеется конкорданс, исследователю уже не нужно ходить пешком, он способен к полету»<sup>145</sup>.

Автоматическое составление конкордансов сейчас более распространено применительно к поэтическим текстам, что связано с относительной простотой определения в них минимального контекста: он совпадает здесь со стихотворной строкой. Обработка текста при составлении, в частности, словаря, упомянутого в сноске 145, включала следующие шаги. 1) Транслитерация с кириллического алфавита на латинский (что обусловлено особенностями машины, на которой проводилась работа по этому словарю) вместе со страницей, названием стихотворения

<sup>141</sup> Колин А. Ж. Т. Автоматическое составление словарника. — В кн.: Автоматизация в лингвистике. М.—Л., 1966.

<sup>142</sup> Вертель В. А., Вертель Е. В., Рогожникова Р. П. К вопросу об автоматизации лексикографических работ. — ВЯ, 1978, № 2, с. 109.

<sup>143</sup> Viks Ü. Õigekeelsussõnaraamatut arvutivariant. Verbid. — Keel ja kirjandus, 1974, N 10.

<sup>144</sup> Штандлова И. Обратные словари. — В кн.: Автоматизация в лингвистике. М.—Л., 1966.

<sup>145</sup> A concordance to the poems of Osip Mandelstam. Ed. by D. J. Kubourlis. Ithaca—London, Cornell University press, 1974, p. XV.

Таблица 2

Зва	Страница	Стихотворение	Стих
Звать 1 <i>Но ложечкой</i> звать, но умильно смотреть. . .	39	«Мороженно!» Солнце. Воздуш. . .	B5
Зовет 2 <i>И</i> зовет к себе конюха мальчика Илью: День и ночь тебя зовет, сквозь слезы тебя ищет. . .	311 313	Пришли четыре брата, несходки Как прочел письмо родитель	30 8
Зовется 1 Так и зовется по имени	213	Эта, какая улица?	1
Зову 3 <i>«Цо»</i> — это я тебя зову! <i>«Во»</i> — это я тебя зову Но я тебя зову	68 68 90	Твое чудесное произношенье Твое чудесное произношенье Я наравне с другими	16 B16 8
Зовут 1 Которую Арменией зовут	169	В год тридцать первый от ро. . .	3
Зовущая 1 Хриплые горы к оружью зовущая —	153	Орующих камней государство	9

(условно все стихи названы здесь по первой строке) и номером строки. 2) Шифровка на перфокартах с последующей автоматической и визуальной корректировкой перфорации. 3) Автоматическое извлечение титулов (названий стихотворений) и их сокращение. Затем номера, приписанные каждому стихотворению, были автоматически заменены на сокращения соответствующих титулов, слова одной группы (т. е. словоформы одной лексемы) сведены вместе, проведена дифференциация омоформ и осуществлен обратный перевод с латинского на кириллический алфавит.

Полученный результат автор назвал «клластерным конкордансом» (*cluster concordance*), поскольку слова в нем собраны как бы «гроздьями» под «главным словом» (*head-word*), т. е. словоформы — под своей лексемой или слова — под гиперлексемой. В качестве примера приведем здесь одну группу слов лексемы «звать»<sup>146</sup> (цифры возле словоформы указывают ее частоту, или количество строк, в которых она встретилась) — см. табл. 2.

Понятно, что такой «клластерный конкорданс» уже с некоторой долей условности может быть назван словарем, составленным из компилятивных параметров, поскольку по крайней мере две процедуры здесь выходят за рамки чистой компиляции и требуют проведения определенного анализа. Это, во-первых, разрешение трудности, которую Д. Кубурлис в другой своей работе назвал «диффузией словоформ» (*word-form diffusion*)<sup>147</sup> и суть которой заключается в том, что формы одного слова в обычном конкорданссе, будучи упорядоченными по алфавиту, могут оказаться в разных местах словаря (ср. англ. *see* — *saw*, *is* — *are* — *were*, рус. *идти* — *шел*, *звать* — *зову* и т. п.). Во-вторых, аналитической работы требует «нечувствительность» машины к омоформам (*homoform insensitivity*), из-за которой в одну единицу соединяются разные слова на основе их буквенного совпадения.

Компилятивный параметр представлен и в словаре рифм. Дж. Т. Шоу — составитель такого словаря на материале пушкинской поэзии — следующим образом ха-

---

<sup>146</sup> Ibid., p. 140.

<sup>147</sup> Koubourlis D. J. From a word-form concordance to a dictionary-form concordance. — In: Computers in the humanities. Minneapolis, 1975, p. 225.

рактеризует этапы работы над словарем <sup>148</sup>. Были сделаны ксерокопии всех произведений поэта по академическому изданию. Каждая строка в этом тексте была помечена ударением для позиции конечного иктуса и добавлен код места строки, который позволяет легко связывать любое рифмующееся слово с его партнером. Затем каждая строка была перфорирована на отдельной карте. Одновременно всем конечным словам в строках была приписана информация, отражающая грамматический статус этого слова, его синтаксические особенности, форму рифмующегося сегмента, т. е. мужская ли это рифма с ее разновидностями или женская или дактилическая и т. п. Эта информация была перфорирована на второй для каждой строки карте. Обе карты записывались на магнитную пленку для получения единой комбинированной информации о строке. Для обработки всего массива были составлены многочисленные программы, и результат был размножен в машинной печати.

Словарь пушкинских рифм состоит из трех частей.

I — лексикон всех конечных слов в строках (как с рифмой, так и без нее), например:

игр'ал ... 1 ем 2  
игр'ал ... 2 ем 1  
игр'ал ... 3 ем 3

Цифры в последней колонке означают здесь количество строк с данной рифмой (т. е. ее частоту), другие обозначения содержат грамматическую информацию — 1-е (или 2-е, 3-е) лицо, единственное число, мужской род.

II — конкорданс рифм — в обратном алфавитном порядке, сгруппированных под рифмующимися сегментами, например:

/'ЕЧ/  
ПЛ'ЕЧ  
р'ечь ежИ-1  
до пл'еч мсР12  
2\*C2.71.2,4.a\*  
м'еч емИ-1  
с пл'еч мсР41  
2 РЛ.2.141,42.6\*

Это означает, что под указанным сегментом слово *плеч* встретилось два раза; справа от слова дана грамматическая информация о нем, которая для последней, например, формы — *с плеч* — расшифровывается так: мно-

жественное число, средний род, родительный падеж, а цифры (41) означают тип предложного управления. Буквенно-цифровая запись под каждой рифмующейся парой сообщает ее адрес: в последнем случае это «Руслан и Людмила», глава 2, страница 141, строка 42; буква со звездочкой символизирует положение данной рифмы — в одном и том же предложении или в разных выступают партнеры, один и тот же ли субъект они характеризуют, к прямому или косвенному объекту относятся и т. п.

III — индекс стихотворений, с помощью которого можно расшифровать сокращенную запись в адресе.

Словари, составленные на материале поэзии Баратынского и Батюшкова<sup>149</sup>, представляют дальнейшее усложнение, увеличение числа параметров и объединяют конкорданс со словарем рифм. Принципы их построения не добавляют ничего нового к сказанному выше о каждом из типов словарей в отдельности. Здесь надо только подчеркнуть, что хотя рифма, согласно нашим представлениям, — это один отдельный параметр, в словаре рифм, как мы видели из приведенных примеров, выступают и многие добавочные параметры, в частности фонетическая, грамматическая и синтаксическая информация о конечном слове строки. Точно так же и в конкордансе — к элементарному иллюстративному параметру присоединяются орфографический, частотный, дистрибутивный. Более того, практика автоматического построения конкордансов приводит лингвистов к выводу, что просто выигрыш в скорости по сравнению с ручным их составлением уже не может считаться преимуществом: современные автоматические конкордансы должны включать максимально возможное число структурных параметров, оснащая исследователя исчерпывающей, быстро обобщаемой и легко доступной информацией о языке<sup>150</sup>.

Создание узкооднопараметровых словарей преследует, как правило, какую-то вполне определенную, часто вспо-

---

<sup>148</sup> Pushkin's Rhymes. A dictionary. By J. Th. Shaw. Madison, 1975, p. IX.

<sup>149</sup> Baratynskii. A dictionary of the rhymes & A concordance to the poetry. By J. Th. Shaw. The University of Wisconsin Press. Ann Arbor, 1975; Batjushkov. A dictionary of the rhymes & A concordance to the poetry. By J. Th. Shaw. The University of Wisconsin Press. Madison, 1975.

<sup>150</sup> Ingram W. Concordances in the Seventies. — Computers and the Humanities, 1974, v. 8, N 5—6.

могательную цель, как, например, в случае хронологического словаря английского языка, вход в который осуществляется именно от хронологического параметра — года первой письменной фиксации данного слова<sup>151</sup>.

Этот словарь был необходим для установления объема и состава английского лексикона времен Шекспира, чтобы изучение языка этого автора можно было осуществлять на фоне языка соответствующей эпохи. С этой целью «Shorter Oxford English dictionary» (SOED), содержащий 80 096

словарных статей, был целиком введен в машину и проанализирован по ряду критерииев. В частности, были исследованы омографы, число значений каждого слова, этимологии, упорядочены данные о письменной регистрации каждого слова, выверены характеристики по частям речи и т. п., что позволило выявить значительное число неточностей и ошибок в словарных статьях. Затем входные слова

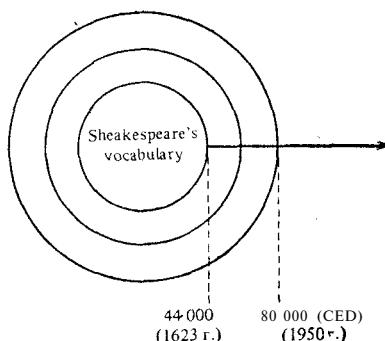


Рис. 1

были переранжированы по времени их появления в письменном языке и получен «Chronological English dictionary» (CED). Объем лексикона, ограниченный 1623 г., — датой появления первого собрания сочинений Шекспира — и определил тот фон, на котором исследуется язык классика английской литературы<sup>152</sup>(см. рис. 1). Результаты анализа представлены в полном и комплексном словаре шекспировских текстов (SHAD), который характеризует каждое слово по 32 (по нашему подсчету) параметрам, объединяющим информацию шекспировского конкорданса Дж. Бартлетта (1894 г.), автоматически составленного современного конкорданса (SHAC), а также упомянутых выше словарей SOED и CED. Естественно, что далеко не все из этих параметров являются чисто компилитивными, и прежде всего это касается информации, связанной с лемматизацией, т. е. сведением текстовых форм к исходным, словарным формам. Сюда относится также установление грамматической и синтаксической двусмыслинности во фразе, определение принадлежности

слова к той или иной специальной сфере, установление общей и более узкой этимологии и т. п.

Комплекс параметров по SHAD, охватывающий информацию и о структуре английского языка, и о его функционировании в Англии на рубеже XVI—XVII вв. в сопоставлении с современным английским (здесь, в частности, используется параметр, который я называю «словарным» и который указывает другие словари, в данном случае английского языка, содержащие сведения о соответствующем слове и указывающие на различия — если они имеются — в его характеристике), и о специфике его употребления в шекспировских текстах, приближается по своей сложности, объему и задачам к так называемому банку данных по подъязыку определенной области, в данном случае — но языку писателя.

Банки данных, как, впрочем, ясно из самого названия, призваны систематизировать и аккумулировать информацию об особой сфере знаний, техники, человеческого опыта с тем, чтобы по запросу исследователя компоновать эту информацию заданным образом и выдавать в необходимых объемах для решения различных задач. Банки данных, по самой структуре своей предполагающие широкую компиляцию материала, в нашем случае — сортирование языкового материала для его лексикографического представления, строятся в основном из компилятивных параметров и являются для этой группы, очевидно, «вершинным достижением», представляя собой максимальную возможность структурирования на основе простой сортировки.

Рассмотрим в качестве примера устройство машинного архива финских диалектов, созданного в университете г. Турку<sup>153</sup>. Архив подготавливался в течение почти двух десятилетий. Записи диалектного материала, собранного

---

<sup>151</sup> Finkenstaedt Th., Leisi E., Wolff D. A chronological English dictionary listing 80 000 words in order of their earliest known occurrence. Heidelberg, 1970.

<sup>152</sup> Spevack M., Neuhaus H. / ., Finkenstaedt Th. SHAD: a Shakespeare dictionary. — In: Computers in the humanities. Edinburgh, 1974, p. 114; Neuhaus H. / ., Spevack M. A Shakespeare dictionary (SHAD): some preliminaries for a semantic description. — Computer and the Humanities, N. Y., 1975, v. 9, N 6.

<sup>153</sup> Описание дано на основе личного знакомства автора с указанной системой и бесед с ее создателями. Из публикаций о ней см., напр.: Ikola O., Karjalainen Y. Syntax archives of Finnish dialects for computer work. — In: Computational and mathematical linguistics, I. Firenze, 1977.

финскими лингвистами и энтузиастами изучения родного языка с конца прошлого века, транскрибировались по единой системе (при необходимости переводились с магнитной пленки в обычный текст на бумагу), и текстовые формы размечались по определенной программе, предусматривающей как структурогенные — в нашей терминологии, т. е. морфологические, синтаксические, фонетические и прочие параметры, так и лингвистические — название диалекта и ареал его распространения, имя, возраст и другие данные об информанте, время записи и проч. Отличительной особенностью этого архива (можно сказать даже «словаря», имея в виду, что последний тоже есть систематизированный комплекс параметров, вход в который осуществляется от одного или нескольких из них) является исключительно широкая представленность в нем синтаксических параметров — их 33 из общего числа около 100. Это и понятно: хотя так же, как в обычном словаре, перед вводом в машину каждое слово здесь было занесено на отдельную карточку, и ему была приписана вся связанная с ним морфологическая, синтаксическая и другая информация, а сами слова упорядочены как в прямом, так и в обратном алфавитном порядке, но не надо забывать, что данный архив имеет дело с текстом — текстом финских диалектных записей, отсюда и обилие, и особое значение в нем синтаксических параметров. Среди последних тоже можно выделить как структурогенные параметры (например, члены предложения; главное и подчиненное предложение; порядок следования главного и подчиненных; повествовательное, вопросительное или восклицательное предложение; прямая или косвенная речь), так и лингвистические (например, число предложений в тексте; порядковый номер предложения в данном тексте; число слов в предложении). При этом почти все названные параметры, за исключением, может быть, первого — «члены предложения», получают определенные значения только в конкретном тексте, т. е. являются иными словами «текстовыми», и потому не могут быть включены в общий перечень лексикографических.

Аналогичный банк данных по другому подъязыку (не диалектному) создается в Университете дружбы народов им. П. Лумумбы и межфакультетской лаборатории вычислительной лингвистики МГУ для обработки записей русской разговорной речи. Система, названная АЛЕКС, представляет собой лексикографическую информационно-

поисковую систему, основной структурной единицей которой является Лексикографический Банк Данных.

«Лексикографическим Банком Данных (ЛБД) называется система, состоящая из:

1) массива Модулей Лексикографических Данных (МЛД);

2) справочника Обращения к МЛД;

3) программы Управляющего Блока, осуществляющего функции накопления, коррекции, хранения, поиска и выдачи модулей по запросам.

Модулями лексикографических данных могут являться тексты (минимальные выборки), словарные статьи, иллюстрирующие контексты и другие лексикографические единства. Модули состоят из Записей Единиц Данных. Единицами данных являются словоформы, словарные формы, ключевые слова, указывающие место данной словоформы в структуре его словарной статьи (в парадигме) или в структуре предложения, синтаксические паспорта словоформ, указывающие их синтаксические функции, частоты употребления словаформ и словарных слов»<sup>154</sup>.

С позиций лексикографической параметризации языка, с, которых мы ведем изложение, рассматриваемая лексикографическая ИПС включает главным образом компилитивные параметры, которые распределяются между четырьмя типами вспомогательных банков данных: аккумулятором, или банком входных данных, содержащим тексты стандартной длины с информацией об источнике, типе записи, типе речи и т. п.; словарем, или банком выходных данных, содержащим лексему с относящейся к ней информацией — часть речи, парадигма, указатель синтаксических функций, набор частот и др.; конкордансом, или банком резервных данных (предложений, контекстов), содержащим адреса хранения управляющих и управляемых слов; индексатором — специальным блоком для реализации диалогового режима.

Банк лексикографических данных, помимо того, что он может использоваться для получения самой разнообразной информации о языке (как современном, так и лю-

---

<sup>154</sup> Анищенко В. М. Некоторые проблемы автоматизации лексикостатистических работ. — В кн.: Вопросы лингвостатистического анализа русской разговорной речи. М., 1976, с. 69; см. также в том же сборнике статью: Ваников Ю. В. Лингвостатистический справочник русской разговорной речи и получение информации высших уровней.

бого исторического периода), представляет собой необходимую базу автоматизации работ в словаростроении. Так, корпус текстов на 5 млн. словоупотреблений, презентирующий функционирование английского языка в Америке наших дней<sup>155</sup>, послужил базой для автоматического построения частотного, ряда школьных словарей<sup>156</sup>, для уточнения при переиздании некоторых других известных словарей компаний Мерриам—Уэбстер.

Аналогичное применение находят банки данных английского<sup>157</sup>, шведского («The Swedish logothèque: a computer-based text and word bank»)<sup>158</sup>, итальянского<sup>159</sup>, древнегреческого<sup>160</sup> и других языков.

### **Особенности аналитической группы параметров**

Примерно полтора десятилетия назад, когда вопросы автоматизации лексикографических работ обсуждались в основном в их перспективе, в модальности будущего времени, и результаты были еще незначительны, лингвистам представлялось, что роль ЭВМ в процессах подготовки словарей будет чисто вспомогательной. Она ограничивалась обработкой компилятивных параметров, т. е. сортированием и обобщением данных, на которых базируется словарная работа: подбором иллюстративного материала, выбором канонической письменной и произносительной форм слова, упорядочением помет об употребительности и т. п.<sup>161</sup> Так возникло понятие и направление computer-aided lexicography, которое ознаменовалось значительными достижениями, особенно при подготовке больших, полных словарей, тезаурусов — в первом, исконном смысле этого слова — как сокровищниц языка, предполагающих обработку больших массивов текстов и объединение максимума словарной информации о языке за определенный исторический период. Таким способом, например, создается в Нанси (Centre de recherche pour un Trésor de la langue française) самый полный словарь французского языка, охватывающий период почти в два века (условно от 1789 г.) и базирующийся на автоматически обрабатываемой картотеке в 150 млн. словоупотреблений. Он соединяет в себе черты словаря исторического и современного и является одним из немногих, в каждой словарной статье которого помимо прочей информации широко представлены библиографический параметр — полный список исследований, посвященных данному слову во француз-

ской и мировой лингвистической литературе, и лингвоистический, отражающий эволюцию означающего (письменной и произносительной форм слова) и означаемого, т. е. его семантики<sup>162</sup>.

Если первый из названных параметров входит в число компилятивных, то второй уже относится к аналитическим. Для их выделения, представления в удобной для машинных операций форме и автоматической обработки недостаточно простого сравнения, на котором основывается опознание, идентификация компилятивных параметров, а требуются более сложные действия по разложению языковых единиц на составные части, обобщению элементов по новым признакам и выбору альтернативных решений. Направление исследования здесь противоположно тому, что мы фиксировали для компилятивных параметров, потому что, как правило, источником для аналитической обработки соответствующих лексикографических параметров служат толковые словари: обычный словарь перестраивается в автоматический, содержащий аналитические параметры. Последний позволяет решать и такие задачи, в которых, на первый взгляд, как бы генерируется новая информация о языке. Однако при ближайшем рассмотрении эти задачи нельзя считать эвристическими, и системы с аналитическими параметрами — так же, как и

---

<sup>155</sup> Paikeday T. M. The American heritage intermediate corpus. — In: Computational and mathematical linguistics, I. Firenze, 1977.

<sup>156</sup> См. рецензии на эти словари: Smith R. N. The American heritage school dictionary. N. Y., 1972; The American heritage word frequency book. N. Y., 1971. — In: Computers and the Humanities, 1974, v. 8, N 5—6, p. 335—336.

<sup>157</sup> Leonard R. The computer archive of modern English texts. — In: Computational and mathematical linguistics, II. Firenze, 1977.

<sup>158</sup> Zettersten A. Current Scandinavian computer-assisted language and literature research. — Computers and the Humanities, 1976, v. 10, N 5, p. 275.

<sup>159</sup> Spogli elettronici dell'italiano delle origini e del duecento. Utrecht—Bologna, 1974; Spogli elettronici dell'italiano letterario contemporaneo. Il Mulino, 1975.

<sup>160</sup> Bruner Th. F. El proyecto «Thesaurus linguae graecae». — In: Utilización de ordenadores en problemas de lingüística. Madrid, 1976 (Revista de la Universidad Complutense, v. XXV, N 102).

<sup>161</sup> Bailey R. W., Robinson J. L. The computer in lexicography. — In: Lexicography and dialect geography. Wiesbaden, 1973, p. 38.

<sup>162</sup> Trésor de la langue française. Dictionnaire alphabétique de la langue des XIX-e et XX-e siècles (1789—1960), t. 1—5. Paris, 1971—1976.

системы, базирующиеся на компилятивных параметрах, — обобщают и выдают («порождают») только то, что было заложено в них с самого начала. Впечатление оригинальности получаемой информации возникает в этих случаях за счет некоторой переклассификации материала, что само по себе, безусловно, важно и необходимо. Вместе с тем автоматический, а тем более машинный словарь всегда проигрывает своему источнику — толковому словарю: при большей аналитичности и точности задания отдельных параметров число их всегда меньше, чем в толковом словаре, что обуславливается более строгой адресностью, более узкой целевой предназначенностю автоматического словаря. Отсюда можно сделать вывод, что автоматический анализ метаязыка словаря практически никогда не осуществляется в полном объеме, если понимать метаязык в широком смысле. Рассмотрим несколько систем такого рода.

В любой системе автоматического анализа текста словарь занимает центральное место. Для венгерского языка такой словарь создан на базе семитомного толкового, параметры которого были частично переосмыслены и переориентированы. В итоге в автоматическом варианте он включает 10 комплексных параметров, из которых только два последних являются компилятивными<sup>163</sup>:

1) сложное или простое слово (из скольких корней состоит лексема), 2) омонимия, 3) часть речи (в случае конверсии каждая отдельная форма квалифицируется как самостоятельная), 4) число значений, 5) стиль, 6) морфологические данные (для единиц, способных к словоизменению), 7) управление (только для глаголов), 8) этимология, 9) суффикс, 10) длина слова. Переразложение параметров позволило выделить новые аспекты их сопоставления и обобщения и в результате предложить решение ряда неясных вопросов в грамматике венгерского языка. В частности, за одним из способов образования притяжательной формы 3-го лица ед. ч. — с *j* или без него, — оценившемся как нерегулярный (*lab* — *laba* 'его нога', но *comb* — *combja* 'его бедро'), обнаружена закономерность, согласно которой способ с *j* используется как маркированный, для обозначения необычного, редко встречающегося или недавно появившегося в венгерском языке сочетания согласных в конце основы. Новые закономерности выявлены также в распределении правил сингармонизма между разными группами слов,

Переклассификация синонимических и антонимических отношений слов в английском языке с целью выявления сети взаимосвязанных семантических классов и получения точной и максимально полной картины структурирования английской лексики осуществлена с помощью ЭВМ на базе «Webster's New Dictionary of Synonyms»<sup>164</sup>. Эта работа выполняется на материале дефиниций, и одной из аналитических процедур, к которым прибегают авторы, является различение словарной единицы (dictionary entry) и лексической единицы (lexical item). Так, если в Уэбстерском словаре синонимов и антонимов многозначное слово simple трактуется как одна отдельная словарная единица с пятью значениями, то в машинном варианте словаря предлагается каждое значение рассматривать как самостоятельную лексическую единицу, т. е. ввести simple 1, simple 2 и т. д. Использование лексической единицы дает возможность представить синонимические и антонимические отношения как бинарные и переходить к лексико-семантическим классам как группировкам лексических единиц.

Традиционный лексикографический параметр дефиниции может подвергаться другим разновидностям аналитических процедур с целью получения новой информации о семантико-синтаксических свойствах слов. При создании автоматического словаря итальянского языка дефиниция эксплуатируется как бы дважды — с одной стороны, как носитель характеристики означающего, как семантический эквивалент слова, с другой — как модель синтаксической сочетаемости характеризуемого слова<sup>165</sup>.

<sup>163</sup> Papp F. Automatic analysis of Hungarian texts and linguistic data. — In: Computational and mathematical linguistics, I. Firenze, 1977, p. 287. — Как раз на этом примере удобно продемонстрировать сокращение, сжатие общего числа параметров в автоматическом варианте словаря по сравнению с исходным толковым. Словарная статья в «A Magyar nyelv ertelmező szótára» (v. I—VII, Budapest, 1959—1962) включает помимо названных параметры частичного произношения, ареальный, арго, подробно разработанный словообразовательный параметр, историко-лингвистический, а также элементы страноведческой информации.

<sup>164</sup> Edmundson H. P., Epstein M. N. Research on synonymy and antonymy: a model and its representation. — In: Papers in computational linguistics. Budapest, 1976.

<sup>165</sup> Ferrar G. Dictionnaire automatique et dictionnaire — machine: une hypothèse. — In: Computational and mathematical linguistics, I. Firenze, 1977.

При этом автор словаря - Дж. Феррари - исходит из следующих предпосылок: во-первых, основой для конструирования словаря, предназначенного к анализу текста, остается лексическая и семантическая информация, тогда как синтаксическая используется в редуцированном виде и минимальном объеме; во-вторых, каждая словарная статья — прямо или косвенно, т. е. в снятом, нейтрализованном виде, — отражает один или несколько типов контекстов данной единицы; и в-третьих, дефиниция каждой словарной единицы принадлежит одновременно и языку, поскольку передает сведения о некоторой реальности — лингвистической или впевзыковой, и метаязыку, поскольку каждый ее элемент может рассматриваться как по определенному дифференциального признака. Так, из определения КРАСОТЫ как «качества того, что красиво» мы узнаем не только, что такое «красота», но и то, что это слово принадлежит к классу качеств, свойств и обладает набором признаков (семантических, логических и синтаксических — как заместитель определенных позиций во фразе), характеризующих этот класс. Точно так же определение СОБАКИ как «домашнего млекопитающего из рода плотоядных» дает нам сведения, что в классе животных собака имеет свойства млекопитающих, а значит, в лингвистическом плане это слово связывается с группой слов, называемых действиями: ACCOUCHER, POULINER, VELE, METTRE BAS etc. В итоге общая структура автоматического словаря предстает в виде трехуровневого дерева, первое ветвление которого отражает набор генерализованных категорий, второе устанавливает иерархические отношения между словами в вокабуляре и третье фиксирует отношения эквивалентности между ними.

В зависимости от задач, на которые ориентирован словарь, его масштабы и соответственно объем аналитических параметров в нем могут существенно колебаться. В проекте универсального макрословаря русского языка, строящегося как объединение словарников основных русских филологических, энциклопедических и специальных словарей и рассчитанного на 1 млн. лексем (соответственно 10 млн. словоформ), главная аналитическая проблема связана с 'экономным' представлением парадигм. Ее решение достигается за счет отказа от явного задания всех словоформ и изменения структуры словаря. «Структура словаря должна представлять собой сочетание словарей двух типов: словаря машинных основ и словаря словоформ,

сжатых по методу Купера. В словаре основ все словоформы данной парадигмы порождаются от общей машинной основы, представляющей собой цепочку букв от начала слова, общую для всех словоформ одной парадигмы. Эта машинная основа не обязательно должна быть связана с традиционной основой. Порождение нужной словоформы осуществляется стандартной операцией апплицирования машинной флексии»<sup>166</sup>.

Г добавлением к категориальным, формально-грамматическим параметрам семантических усложняются задачи их аналитической обработки. В английском словаре, предназначенному для систем, понимающих естественный язык, наряду с набором постоянных, ядерных, признаков, каждая часть речи характеризуется еще специфической серией параметров, и все они являются результатом специального анализа. В число постоянных, относящихся ко всем частям речи, включены такие:

- 1) орфографический, т. е. сама словарная единица,
- 2) часть речи,
- 3) семантическое поле,
- 4) словарная дефиниция,
- 5) нерегулярная словоизменительная форма,
- 6) словообразовательная морфология,
- 7) синонимы, включая синонимические ссылки,
- 8) антонимы,
- 9) примеры употребления для каждого случая, имеющего дефиницию,
- 10) время реакции информанта на предложение и усредненный тип ответа по всем информантам,
- 11) данные об информанте.

Имя существительное, например, характеризуется сверх того следующими параметрами, которые извлечены из дефиниции с помощью особых приемов анализа:

- 1) синтаксико-семантические признаки: ± человек, ± одушевленное, ± исчисляемое, ± конкретное, ± мужское, ± женское;
- 2) падежные маркеры (в смысле Филлмора);
- 3) метафорические переносы значения;

---

<sup>166</sup> Беляева Л. Н., Крисевич В. С., Липницкий С. Ф., Пиотровский Р. Г. О многоцелевом автоматическом словаре русского языка (МАРС). — В кн.: Вопросы общей и прикладной лингвистики. Минск, 1975, с. 156—157.

4) социолингвистические ограничения употребительности<sup>167</sup>.

Аналогичной структурой обладает дрезденский машинный словарь, в котором в силу его специфики отсутствуют психолингвистические и социолингвистические параметры<sup>168</sup>.

Двуязычные и многоязычные словари, опирающиеся на такие же аналитики, принципиально не отличаются по своему устройству от рассмотренных выше и, не имея возможности останавливаться подробно на их характеристике, ограничимся здесь упоминанием лишь некоторых отечественных и зарубежных работ такого рода<sup>169</sup>.

Завершая рассмотрение автоматических словарей и систем с использованием большого числа аналитических параметров, укажем еще две работы, которые представляются наиболее полными по охвату параметров и наиболее сложными — по их структурированию — из всех, отнесенных нами к данной группе. Первая из них — снимающий неоднозначность словарь (*The disambiguation dictionary*) для автоматического распознавания смысла высокочастотных английских слов<sup>170</sup>. Как известно, наиболее употребительные слова являются и самыми многозначными. Вместе с тем из 117 значений и употреблений глагола *take* или 91 значения существительного и конвертированного глагола *hand* (по словарю «*The Random House dictionary of the English language*») далеко не все оказываются употребительными и даже реально встречающимися в современных английских текстах и устном общении. Авторы поставили перед собой задачу проанализировать, в каких значениях выступают 1815 самых частотных слов в текстах на 500 тыс. словоупотреблений, относящихся к различным сферам общественной жизни, и разработать основы словаря, который позволял бы в ходе контент-анализа однозначно идентифицировать смысл, в котором употреблено данное многозначное слово в данном контексте. Результатом работы системы является приписывание каждой единице анализируемого текста номера значения этой единицы, зафиксированного в словарной статье автоматического снимающего неоднозначность словаря. Последний, таким образом, имеет два источника: корпус текстов и толковый словарь. Текст был обработан на уровне компилиативных параметров и представлен в виде KWIC-конкорданса (*keyword-in-context*) для 1815 исходных слов — со всевозможными статистическими

характеристиками. Параметры же толкового словаря подверглись аналитической обработке, с неизбежной в подобных случаях редукцией некоторых из них (в частности, этимологического, исторического, словообразовательного, произносительного и некот. др.).

В числе аналитических процедур были следующие: создание анализатора словоизменительных форм, который давал возможность представить каждую словарную единицу как основу, к которой прибавляется — в случае регулярной парадигмы — определенный набор суффиксов (например, -s, -ed, -ing для глагола); установление части речи, в функции которой выступает данная форма, что в значительной степени определяет синтаксические свойства ее ближайшего окружения; идентификация — с помощью толкового словаря — значения слова, реализованного в данном контексте, с выявлением неопределенных двусмысленных или, наоборот, генерализованных его употреблений и т. п.

Отличительная особенность полученного словаря заключается в том, что в нем исключительно детальную характеристику приобретают синтаксические и семантические параметры, названные здесь «категориальными маркерами». Среди десяти групп синтаксических параметров:

— детерминативы, включающие, помимо artikelей (a, an, the) и демонстративов (this, these; that, those), также

<sup>167</sup> Maxwell E. Я., Smith R. N. A computerized lexicon of English. — In: Computers in the humanities, Minneapolis, 1974; Smith R. N.-Maxwell E. R. An English dictionary for computerized syntactic and semantic processing systems. — In: Computational and mathematical linguistics, I. Firenze, 1977, p. 310—311.

<sup>168</sup> Нейберт Г. О структуре и применении машинного словаря научно-технических подъязыков. — В кн.: Использование математических моделей и электронных вычислительных машин в лингвистике. София, 1976.

<sup>169</sup> Шаляпина З. М. Англо-русский многоаспектный автоматический словарь (АРМАС). — В кн.: Машинный перевод и прикладная лингвистика, вып. 17. М., 1974; Deweze A. The trilingual computational dictionary «THESEE». Thesaurus, compatible with the French, English and German indexing systems. — В кн.: Использование математических моделей и электронных вычислительных машин в лингвистике. София, 1976; Stachowitz A. Beyond the feasibility study: lexicographic progress. — In: Computational and mathematical linguistics, I. Firenze, 1977; Wilton M. T. Bilingual lexicography: computer-aided editing. — Ibid.; Зиман Ю. Л., Егорова Е. В. Использование ЦВМ для автоматизации составления и редактирования словарей. М., 1977.

<sup>170</sup> Kelly E. F., Stone Ph. J. Computer recognition of English word senses. Amsterdam—Oxford, 1975.

показатели принадлежности («генитивы» в терминологии авторов: *my, our, your*. . .), числительные (в двух субкатегориях: количественные и порядковые) и предартикли (исчисляемые: *both, each, few, every*. . ., и смешанные: *all, any, some, half*. . .);

— указатели позиции слова во фразе (первое слово, последнее слово);

— специальные глаголы и глагольные элементы, содержащие девять субкатегорий, в том числе — глаголы-связки (*seem, look, feel, remain*. . .), глаголы превращения (*become, get, turn into, grow*. . .), модальные глаголы (*can, may, shall*. . .) и др.

Семантические параметры классифицируются в соответствии со стандартной практикой контент-анализа и представлены в данном случае шестнадцатью семантическими полями, часть которых предполагает еще и дополнительное внутреннее подразделение: I одушевленное, II человек, III коллективное, IV абстрактное имя, V общественное место, VI часть тела, VII политическое понятие, VIII экономическое понятие, IX цвет, X коммуникация (в том числе — газета, книга, телефон, печать, карандаш. . .), XI эмоции, XII частота (редко, случайно, часто. . .), XIII оценочные прилагательные (плохой, хороший, прекрасный, трудный. . .), XIV прилагательные величины, XV прилагательные позиции (низкий, верхний, средний. . .), XVI наречия степени<sup>171</sup>.

В результате применения специальных правил анализа текста и его параметризации получен автоматический словарь (он воспроизведен в машинной печати в виде приложения к книге), содержащий очень большой объем самой разнообразной информации и находящий применение не только при контент-анализе, но и для целей информационного поиска и при некоторых других видах работ с текстами. В качестве образца приведем без комментариев одну из коротких словарных статей — для слова *гнев*<sup>172</sup>.

Наконец, последняя из рассматриваемых систем относится к машинному представлению всех лексикографических параметров Уэбстерского словаря — и компилитивных, и аналитических — с целью автоматического редактирования, переклассификации используемых данных, добавлений и изменений при переиздании, а также для возможного применения его к задачам автоматического поиска информации. Система разработана на базе стандартного способа структурирования библиографических

	Suffixes:	Root	s	ed	ing	Total
ANGER	Overall frequencies:	20	0	2	0	22
1) NOUN WRATH, IRE		19	0	0	0	19
2) VERB TO MAKE ANGRY		1	0	2	0	3
rule test from to for logic	pass fail					
1) TAG -1 -1 PRON —	SENSE 2 NEXT	1	0	0	0	1
2) TAG +0 +0 ROOT —	SENSE 1	19	0	0	0	19
	SENSE 2	0	0	2	0	2

данных MARC (machine readable catalog), имеющего распространение в Великобритании и США, и в применении к указанному словарю получила название WEBMARC<sup>173</sup>.

Приведение одной из словарных статей данного словаря в качестве примера потребовало бы слишком развернутого комментария из-за сложности используемой записи, поэтому мы ограничимся здесь самой общей характеристической системы. Если исходить из понятия параметра, развивающегося в наших работах, то число таких параметров в системе WEBMARC не превышает 19. В классификации автора этой системы опорными лексикографическими категориями являются «поля», которых выделяется шесть: поле побуквенной записи (орфографическое), поле дефиниции, поле полного произношения, синонимическое поле, этимологическое поле, поле частичного произношения, — и которые подразделяются на субполя. Между полями и субполями — с одной стороны, и параметрами — с другой стороны, нет однозначного соответствия: параметр может совпадать и с полем (например, этимологический), и с субполем (например, омографический), некоторым субполям и даже полям (например, «частичного произношения») может не соответствовать ни один параметр, и наоборот, одно субполе может включать два параметра. Например, субполе, названное «точки слогоделения» — • hyphenation points — и кодированное знаком \$b в орфографическом

<sup>171</sup> Kelly E. F., Stone Ph. J. Op. cit., p. 17—21.

<sup>172</sup> Ibid., p. 127.

<sup>173</sup> Sherman D. A common structure for lexicographic data. — In: Computers in the humanities. Minneapolis, 1974.

поле, содержит информацию о числе слогов (один параметр) и количестве букв в каждом слоге (другой параметр), считая с конца, от точки, отделяющей последний слог слова; так, для слова *dic.tio.nary* этот код записывается \$b33.

Рассмотрим подробнее субполевое структурирование орфографического поля в соотношении с нашими параметрами. Прежде всего, до побуквенной передачи слова оно (поле) содержит указание на класс слов, к которому принадлежит данное, причем понятие «класс слов» оказывается несколько шире категориального лексикографического параметра «часть речи», поскольку включает частично морфологическую (PF — prefix, SF — suffix, VS — verb suffix . . .), частично словообразовательную информацию (CF — combining form, NC — noun combining form . . .). Таким образом, мы имеем основание записать:

класс слов = параметр «часть речи», параметр морфологического членения, параметр словообразования.

Далее следуют субполя:

\$a — побуквенная запись = собственно орфографический параметр,

\$b — точки слогоделения = параметр числа букв, параметр слогоделения,

\$c — показатель ударения = параметр ударения,

\$d — статус альтернативного класса слов (в случае конверсии)=0,

\$e — код альтернативного класса слов = 0,

\$1 — статус принадлежности данной единицы к тому или иному социальному срезу языка (dial., slang. . .) — ареальный параметр (указание региона), параметр арго (указание типа арго),

\$i — варьирование, типы вариантов = 0,

\$s — статус вариантов (равноценный, предпочтительный, второстепенный) = параметр нормативности.

Итак, пример рассмотренной системы со всей наглядностью показывает, что широкая разработка и использование аналитических параметров всегда опирается на толковый словарь, т. е. источник получения их вторичен по отношению к тексту, тогда как результат их обработки и систематизации как бы возвращается к тексту на новом уровне, поскольку предназначен для его автоматического анализа.

## **Конструктивные параметры и генерирование новой информации о языке**

Под конструктивными понимаются параметры, требующие для своей обработки построения специальных эвристик. Как правило, это параметры, включающие семантический компонент — словообразовательный, синтагматический (свободная и связанные сочетаемость), ассоциативный, лингвострановедческий, терминологический, параметры словоизменения (склонения, спряжения, степени сравнения, наклонения) и другие, и потому как бы подразумевающие присутствие человеческого фактора. Оперирование конструктивными параметрами с необходимостью предполагает и предварительную *компилацию* соответствующих данных, и *апализ* определенных характеристик, по главной особенности является связанное с ними решение эвристических задач, которое ведет к созданию новых лингвистических объектов. Новым лингвистическим объектом мы называем такой объект (это может быть словарь, лингвистический атлас, функционирующая модель фрагмента языковой структуры, целеориентированная грамматика и т. п.), который возникает не в результате описания некоторого языкового материала, а в результате эксперимента. Этот объект, будучи построенным на определенных теоретических предпосылках, сам дает новый материал для дальнейших наблюдений, исследований, выводов о данном языке. Таким образом, в случае работы с конструктивными параметрами одно их свойство как бы обуславливает другое: решение эвристических задач опирается на эксперимент, а возникающий в ходе эксперимента новый лингвистический объект позволяет генерировать новую информацию о языке, выявлять нетривиальные, иногда непредсказуемые его свойства.

В числе исследований подобного рода — работы разной степени сложности: в одних ставится задача построения, так сказать, однопараметровых объектов, другие направлены на объединение параметров, их синтез и универсализацию полученного объекта. Рассмотрим некоторые параметры.

Автоматический словарь эстонских глаголов<sup>174</sup> ре-

---

<sup>174</sup> Викс Ю. А. Классификаторная модель эстонской морфологии (автоматический синтез глагольных словоформ). Канд. дис. Таллин, 1978.

шает задачи опознания грамматического значения любой правильной формы (т. е. позволяет приписывать некоторой словоформе соответствующее время, лицо, залог, наклонение, число и т. п.) и задачи конструирования необходимой словоформы по заданному набору грамматических значений. Построение такого словаря для сильно флексивного языка, каким является эстонский, потребовало представления каждого члена парадигмы в виде особой программы, которая предусматривает эвристическую задачу установления по характеру основы — типа спряжения, соответствующих ему правил дистрибуции основы и ее морфонологической перестройки в процессе глагольного словоизменения. Сокращение программных процедур и их типизация в машинном эксперименте достигается с помощью оптимальной классификации самих основ, флексий и типов морфонологических чередований. Важной особенностью данного словаря является то, что он значительно увеличивает число «входов» в морфологию. Если в обычном описании вход осуществляется либо от словоформы (анализ), либо от слова-образца (т. е. через парадигму), то в автоматическом словаре, при сохранении указанных возможностей, вход осуществляется также и от перечня грамматических значений (как от каждого значения в отдельности, так и от различных их комбинаций), и от перечня форм (классический анализ), и от классификации. Последняя, являясь организующим стержнем этого нового лингвистического объекта, представляет собой предельно сжатое выражение структуры эстонской глагольной морфологии, т. е. сетку запретов и разрешений, которая как бы сдерживает «комбинаторный взрыв» логически возможных комбинаций грамматических значений и их фонологических репрезентаций. Этот словарь на ограниченном материале глагольного словоизменения моделирует в известном смысле знание языка, поскольку с достаточной надежностью по типу основы и ее месту в классификации предсказывает ее морфологическое и морфонологическое поведение, т. е. относит ее к классу определенных парадигм.

Машинный толковый терминологический словарь, относящийся к области ядерной физики и по своему типу приближающийся к традиционному энциклопедическому, разработан как составная часть системы, понимающей естественный язык<sup>175</sup>. Ограниченнность предметной области дает здесь возможность рассматривать наименова-

**ния ядерных реакций как единый тематический ряд.** План содержания каждой из подгрупп терминов, входящих в тематический ряд, описывается методом компонентного анализа, и все описания соединяются в одном классификаторе. Особенность построенного таким образом аналитико-смыслового словаря в том, что принцип порождения распространяется и на план содержания. Разработанный классификатор содержит три разбиения и дает возможность конструировать наборы семантических множителей (позиционные смысловые коды), которые однозначно соответствуют содержанию определенного научного понятия. Связи и отношения, свойственные описываемой предметной области, находят отражение в самой структуре классификатора.

При составлении позиционного кода из каждого разбиения берется по одному семантическому множителю, и смысл любого термина предстает в виде трехместного цифрового кода, внутренним свойством которого является значимый порядок следования позиций. Так, термину «упругое рассеяние нейтронов» соответствует смысловой код 2.2.1, где первая цифра характеризует тип налетающих частиц (нейтроны), вторая цифра — тип вылетающих частиц (нейтроны), третья цифра характеризует состояние остаточного ядра (остаточное ядро равно начальному). Для порождения плана выражения соответствующего термина строятся матрицы перехода, отражающие связь между двумя планами. Использование матриц позволяет выявить такие закономерности в отношениях между планом содержания и планом выражения, которые в вербальной формулировке остаются скрытыми. Эксплицитное обозначение этих закономерностей служит основой для представления их в виде алгоритма. «Комбинаторный взрыв» и порождение бессмысленных кодов нейтрализуются здесь тоже специальными семантическими алгоритмами.

Таким образом, на основе алгоритмов, программ и исходных данных ЭВМ строит словарь, состоящий из эксплицитной записи смысла термина в соответствии с классификатором и формы выражения этого смысла в виде терминологического словосочетания или символа. Словарь

---

<sup>175</sup> Марютина А. И. Лингвистические основы и структура аналитико-смыслового словаря для систем, понимающих естественный язык, Канд. дис. М., 1978.

хранится в ЭВМ в потенциальной форме, допускает расширение за счет введения новых разбиений в классификатор, новых дифференциальных признаков в каждое разбиение и т. д. и не требует при этом пересмотра всей структуры. Кроме того, данный словарь предусматривает использование его для лексикографического перевода терминов той же предметной области на английский и французский языки.

Рассмотренный лингвистический объект использует несколько эвристик, имитирующих знание языка: во-первых, умение отсеивать бессмысленные наборы семантических множителей, что требует «знания» о мире, знания данной предметной области; во-вторых, умение переходить от смысла к его выражению с учетом структурных особенностей соответствующего языка — возможностей комбинирования основ с флексиями и типов модификаций основ при образовании словоформ — для русского, или выбора формы артикля — для французского и т. д.

Присутствие человеческого фактора наиболее отчетливо проявляется при разработке такого параметра, как ассоциативный. Другой характерной чертой опытов автоматического построения ассоциативных словарей можно считать то, что в них наглядно подтверждается наличие системных связей между самими параметрами, обнаруживаются тесные взаимозависимости внутри отдельных их подгрупп. Установление же таких зависимостей, в свою очередь, несет новую информацию и о языковой структуре в целом, что дает основания отнести этого рода работы к рассматриваемой, третьей группе конструктивных параметров.

Что касается первой из названных черт — человеческого фактора, то его присутствие совершенно очевидно, поскольку машинной обработке в этом случае подвергаются результаты тестовых ассоциативных экспериментов. Так, авторы ассоциативного тезауруса английского языка, отобрав по психолингвистическим (т. е. по принципу наиболее принятого использования данных слов в качестве стимулов в ассоциативных экспериментах), частотным и смысловым критериям 8400 слов-стимулов, проанализировали затем анкеты, полученные от 100 информантов, каждому из которых было предложено ответить на 100 стимулов<sup>176</sup>. Тезаурус представлен в виде сетей словесных ассоциаций и содержит 55 732 узла, которые характеризуются входящими связями. Исходящие

связи тоже учитываются. В другой форме представления данный тезаурус выступает как перечень слов-стимулов с указанием к каждому из них всех входящих и исходящих реакций в порядке убывания их частотности. Наконец, для удобства пользования представлен и обратный перечень — всех входящих связей, какие только появились в ответах испытуемых. Ценной особенностью этого словаря является то, что в нем зафиксированы, а затем проанализированы не просто слова-реакции в их лемматизированной форме, а учтены словоизменительные разновидности, в которых эти слова выступали в ответах информантов, зафиксированы ошибки, варианты произношения и т. п. Статистическая обработка этих данных и их семантический анализ позволили выявить взаимозависимости между собственно ассоциативным параметром, частотным и рядом морфологических характеристик. Например, определение силы ассоциативной связи — как для входящих, так и для исходящих отношений — опирается на частотный критерий; обобщение данных по частоте словоизменительных форм в ответах приводит авторов к некоторым психологическим оценкам отдельных аспектов английской грамматики.

Указанные особенности выгодно отличают рассмотренный ассоциативный тезаурус от другого подобного ему словаря английского языка, в котором ассоциативный параметр выступает как бы в изолированном виде<sup>177</sup>. Здесь автор аксиоматически вводит 118 понятийных категорий, которые выступают как семантические примитивы и являются, по его мнению, достаточными для передачи значения любого слова. Они получают условные мемориальные обозначения и могут образовывать различные комбинации (AIM, BODY, EVER, GLAD, HEAR, LADY, MOTV, TIME, WRIT, NO. . .). Как справедливо отмечалось в рецензии на этот словарь, даже для выявления только сети ассоциативных связей «простые корреляции между словами оказываются неадекватными вне их соотношении с другими аспектами языковой структуры»<sup>178</sup>.

---

<sup>176</sup> Kiss G. et al. An associative thesaurus of English and its computer analysis. — In: The computer in linguistic and literary studies. Edinburgh, 1973.

<sup>177</sup> Laffal J. A concept dictionary of English. Essex—N. Y., 1973.

<sup>178</sup> Huntsman J. F. A concept dictionary of English, by Julius Laffal. — Computers and the Humanities, 1975, v. 9, N 1, p. 47.

Необходимость использования — при построении новых лексикографических объектов — не только собственно конструктивных, но и параметров других рангов, а именно компилятивных и аналитических, можно продемонстрировать на примере организации семантического архива английского языка в Объединении системных исследований Санта Моника (США)<sup>179</sup>. Само название этой системы («Архив») ориентирует как будто на восприятие ее как банка лексико-семантических данных, однако уже в проекте ее создания предусматривалось, что она будет функционировать одновременно и в роли информационного центра, и в качестве исследовательской базы — источника получения новых сведений о лексике английского языка. Исходные данные этой системы скомпилированы из четырех типов источников — Уэбстерского словаря (7-е изд.), тезауруса Роже, «Брауновского корпуса» (*«Standart corpus present-day edited American English»*)<sup>180</sup> и большого круга лингвистических, философских, психологических, антропологических и другого рода гуманитарных исследований, в которых так и иначе анализируется значение и употребление конкретных слов, вошедших в общий лексикон по трем первым источникам. Таким образом, в отличие от упоминавшегося выше французского «Trésor» библиографический параметр в системе SOLAR выполняет не просто отсылочную функцию, а получает глубинную, содержательную интерпретацию, снабжая лексикографа и лексиколога сведениями о детальной разработке семантики отдельных слов. Суммарный объем используемого лексикона охватывает свыше 10 млн. словоупотреблений. Весь материал систематизирован в девяти разрядах (группах) параметров — семи основных и двух вспомогательных, для каждой из которых разработан свой машинный формат. Первые три группы: 1) семантический анализ, 2) объяснение дескриптивных констант, используемых для семантического анализа, 3) концептуальный анализ понятий, связанных с дескриптивными константами — обрабатываются и подготавливаются к вводу в машину ручным способом. Остальные группы основных параметров: 4) дополнительные признаки, коррелирующие с отдельным значением слова, 5) семантические поля, 6) развертывание дефиниции, 7) контексты употребления, а также вспомогательные — индекс слов и библиография — обрабатываются автоматически.

Соотношение друг с другом трех первых групп параметров, в которых каждая последующая выступает как разъяснение существа предыдущей, позволяет оценить их как найденный авторами удачный, на наш взгляд, прием, останавливающий регресс в бесконечность при анализе лексической информации, обусловленный свойством ее «циркулярности» в словаре, где для объяснения значения слов употребляются те же самые слова. Так, семантический анализ строится на описании значений слов, содержащихся в специальных работах — главным образом лингвистов и философов, и включает следующие параметры:

- установление предикатно-аргументных отношений, причем под предикатом понимается анализируемое значение, а дескриптивные константы (т. е. семантические маркеры=семантические компоненты=семантические дифференциальные признаки=семантические примитивы), соответствующие этому значению, указывают на роль выражений, функционирующих в качестве аргументов;
- компонентный анализ значений;
- установление пресуппозиций отдельного значения;
- коннотации данного значения;
- условия его распространения, т. е. возможности его развития в предложении;
- допустимые дополнения;
- субкатегориальные признаки.

Объяснительные примечания к дескриптивным константам, используемым в семантическом анализе, содержат толкование нестандартных символов, определение данной константы, извлеченное из оригинального текста автора, употребляющего и разрабатывающего эту константу, отсылки к другим словам, в анализе которых участвует данная константа. Наконец, концептуальный анализ, интегрируя итоги предыдущего, базируется на результатах аналитической философии и включает рассмотрение центральных понятий, связанных с дескриптивными константами, таких как «факт», «событие», «причина», «личность», «действие» и т. п.

---

<sup>179</sup> Diller T., Olney J. SOLAR (A Semantically-Oriented Lexical Archive): current status and plans. — Computers and the Humanities, 1974, v. 8, N 5—6.

<sup>180</sup> Kuiera H., Francis W. N. Computational analysis of present-day American English. Brown University Press, Providence, 1967.

Для иллюстрации процедуры обработки параметров первых трех групп приведем один пример. Семантический анализ слова *excuse* 'извинять, прощать' выявил в качестве одной из дескриптивных констант *defendant* 'ответчик'. Согласно Ч. Филлмору, использовавшему эту константу, понятием более высокого ранга, участвующим в ее определении, является *responsible* 'ответственный'. В соответствующем формате на это понятие критически рассматриваются различные точки зрения нескольких авторов, занимавшихся его анализом (в том числе и определение его в Уэбстерском словаре), и делается попытка сформулировать интегрированное представление о его содержании.

Что касается других групп параметров, то они в значительной степени традиционны. Так, к дополнительным признакам значения авторы системы относят обычные стилистические пометы о сферах использования и ограничениях в употреблении, пояснения, даваемые к словарной дефиниции в скобках, иллюстрации употребления (которые одновременно представляют собой и часть общего конкорданса и потому дублируются 7-й группой параметров), некоторые сведения о сочетаемости, извлеченные из работ по «формальному семантическому анализу» рассматриваемой единицы. Семантические поля (5-я группа) компонуются главным образом на основе Уэбстера и Роже и строятся на двух типах отношений — дефиниционных и синонимических: некоторое слово *x* включается в поле данного *y*, если, во-первых, оно участвует в его дефиниции или, наоборот, данное слово *y* входит в дефиницию *x*; во-вторых, если находится с данным в одной из широко понимаемых синонимических связей (например, *end* считается синонимом *last*). Тип взаимоотношений каждого слова в поле с заглавным указан, так же как указана и дополнительная информация — часть речи, число значений у слова, помимо того, в котором оно включено в данное поле (так называемые «омографы»), переходность—непереходность у глагола и др. Распространение дефиниции (6-я группа параметров) осуществляется на двух уровнях — по общезыковому, т. е. филологическому словарю, а затем по специальному, в зависимости от той области, к которой относится данное слово (металлургия, горное дело, химия и т. д.): из дефиниции Уэбстерского словаря выбираются слова, которые имеются в словарнике терминологического словаря соответствующей

области (первый уровень), и из последнего выписываются их дефиниции (второй уровень). Наконец, последний — 7-й формат представляет собой обычный KWIC-конкорданс, составленный по всем используемым в системе источникам.

Уже простое перечисление использованных в системе SOLAR параметров показывает, что их состав выходит за рамки собственно лексикографических (хотя, естественно, и не все из квалифицированных нами как лексикографические нашли отражение в этой системе: здесь отсутствуют, например, этимологический, словообразовательный, словоизменительный, производительный и многие другие параметры). Поэтому система способна не только служить базой при составлении различных словарей, но и давать ответы на широкий круг запросов, требующих и аналитической работы с исходными данными, и такой их переклассификации, которая выявляет неизвестные ранее свойства лексической семантики английского языка. Подобные исследования находятся уже на грани между лексикографическим объектом и моделью языка в целом, поскольку вплотную подводят к процессам «понимания» сообщения, к уровню принятия решений. Охарактеризованная система SOLAR, например, в принципе обладает теми же «знаниями» (только в иных пропорциях), что и системы, «понимающие естественный язык», а именно — семантикой, синтаксисом, логикой и представлением о мире (т. е. знанием определенной предметной области), отличаясь от таких систем меньшим динамизмом, но зато превосходя их масштабами<sup>181</sup>. Тем самым ее рассмотрением мы должны завершить наш обзор, если хотим остаться в пределах лексикографии.

Однако нам кажется необходимым остановиться на характеристике еще одной системы, которая по существу не относится к лексикографии прямо, но по своему устройству и функционированию аналогична словарю. Речь идет о фонде (банке данных) эвристических приемов решения конструкторско-изобретательских задач, и понятно, что единицей в этом своеобразном словаре уже становится не слово, а эвристический прием, т. е. краткое предписание (правило), как преобразовать прототип, чтобы получить новое техническое решение, или как синтезировать

---

<sup>181</sup> См., напр.: Виноград Т. Программа, понимающая естественный язык. М., 1976.

новое решение<sup>182</sup>. Источниками выявления и систематизации эвристических приемов при составлении фонда послужила обширная мировая литература, посвященная методам поиска новых технических решений. В республиканской научно-исследовательской лаборатории математических методов оптимального проектирования при Марийском политехническом институте им. М. Горького проанализированы методы: морфологического ящика (США), матриц открытия (Франция), организующих понятий (ГДР), алгоритма решения изобретательских задач (СССР), систематической эвристики (ГДР), ступенчатого подхода к решению задачи (США), функционального изобретательства (Англия), интегральный метод открытия (Франция), десятичных матриц поиска (СССР), мозгового штурма (США) и мн. др. Всего собрано около 800 не повторяющих друг друга приемов, которые классифицированы в несколько групп и распределены по подфондам, представляющим собой массивы информации, используемые на различных этапах работы обобщенного эвристического алгоритма поиска новых технических решений. Помимо того, что созданная машинная библиотека («словарь» в нашем смысле) эвристических приемов и поисковых процедур может использоваться как обычный банк данных, т. е. как вспомогательное информационное средство при решении научных, технических и организационных проблем, этот «изобретательский словарь» служит центральным узлом реализованной в Марийском политехническом институте автоматической системы синтеза новых технических решений. Система прошла апробирование, и с ее помощью найдены принципиально новые пути конструирования запоминающих устройств, элементов аналоговых вычислительных машин, сооружений из мягких материалов и др.

Эта работа является пионерным исследованием в области изучения и формализации приемов научного творчества: она вплотную подводит к созданию автоматической модели эвристических способностей человека, т. е. фактически создает новый научный объект, обладающий в том числе и лингвистическими свойствами, что в условиях отсутствия общей теории творчества<sup>183</sup> является важным шагом на пути ее построения. Понятно, что в этих своих качествах данная система далеко выходит за рамки лексикографии, вторгаясь в сферу исследования проблем искусственного интеллекта. Вместе с тем она демонстри-

рует основополагающую роль собственно словаря в решении целого ряда эвристических задач.

Подводя итоги сделанному обзору, следует констатировать, что в ходе автоматического построения различных лексикографических объектов с необходимостью осуществляется анализ метаязыка словаря (в широком смысле), причем в зависимости от характера этого объекта акцент делается на той или иной группе параметров, каждая из которых для своей обработки требует специфического подхода, использования особых приемов и процедур. Естественно, что при обработке параметров высшего уровня нельзя миновать предыдущие, поэтому работа, например, с аналитическими параметрами предполагает предварительную компиляцию, а оперирование конструктивными не может обойтись без предшествующего ему анализа. Как показывает опыт составления тезауруса русского литературного языка (см. главу III), все эти этапы являются неотъемлемой частью работы над ним, однако в силу особенностей самого этого объекта, для которого определяющими являются семантические параметры (т. е. параметры конструктивного уровня), центральные проблемы на пути достижения поставленной цели связаны с анализом прежде всего семантической информации. Характеристике этих проблем и способов их решения посвящаются следующие главы, но прежде следует остановиться на специфике семантической информации, отраженной в тезаурусе, и способах ее фиксации в нем.

---

<sup>182</sup> Методы поиска новых технических решений. Под ред. А. И. Половинкина. Йошкар-Ола, 1976.

<sup>183</sup> Как констатируют многие исследователи и в нашей стране, и за рубежом, вопрос о создании подобной теории по-настоящему еще не поставлен на научную основу. См., напр.: Буш Г. Я. Методологические основы научного управления изобретательством. Рига, 1974; Bender A. D. Creativity and productivity programs — how widely they used? — Research Management, N. Y., 1975, v. 18, N 5; Mc. Pherson / ., Andrews W. Collected insights for enlightened managers/administrators. — Journal of Creative Behavior, Buffalo (N. Y.), 1971, v. 5, N 3.

## Глава II

### КАК УСТРОЕН ТЕЗАУРУС, КАК ОН РАБОТАЕТ И ЗАЧЕМ НУЖЕН

---

#### § 4. ТЕЗАУРУС КАК СПОСОБ ПРЕДСТАВЛЕНИЯ СЕМАНТИЧЕСКОГО ПРОСТРАНСТВА

An idea can never make the best use of available information  
(De Bono's 1-st Law).

*E. de Bono. Practical thinking\**

В названии главы сформулированы три разных вопроса, но ответить на каждый из них по отдельности невозможно: рассматривая любой из них, мы так или иначе по необходимости затрагиваем и другие. Что касается структуры тезауруса, тот этот вопрос достаточно широко освещался в литературе, и наш читатель, вероятно, определенным образом представляет себе организацию материала в таком словаре. Будем называть тезаурусом всякий словарь, который в явном виде фиксирует семантические отношения между составляющими его единицами. При таком понимании под это определение подходят и все типы идеографических словарей — тематический, аналогический и собственно идеографический, согласно классификации В. В. Морковкина<sup>1</sup>, и специальные тезаурусы — общетеchnические и информационно-поисковые. Если в общепринятом определении последних снять терминологические и отраслевые ограничения, станет очевидно, что принципиальной разницы между общеязыковым и информационно-поисковым тезаурусом нет: «Тезаурус является лексическим инструментом информационно-поисковых систем. Он состоит из контролируемого, но изменяемого словаря терминов, между которыми указаны смысловые связи. Такой словарь, исчерпывающим образом покрываю некоторую специфическую область знаний, представляет собой перечень дескрипторов и недескрипторов (вспомогательных терминов), который упорядочен по систематическому и алфавитному принципам и содержит указания на смысловые отношения между ними — как иерархического (родо-видового), так и неиерархического типа»<sup>2</sup>. Упорядоченность двух родов — систематическая и алфа-

витная — отличает этот словарь от толкового и определяет главное структурное его свойство — наличие более чем одного входа. В самом деле, чтобы отыскать слово в обычном толковом словаре, мы можем использовать только его внешний облик, его написание, поскольку в таком словаре лишь один вход — алфавитный. Так, слово *диссонирующий* в словаре Ушакова находится между *диссонировать* и *диссоциация* и относится, согласно помете, к области музыки. Для выделения всех терминов, относящихся к музыкальной сфере в этом словаре, надо прочитать весь словарь. В тезаурусе же Касареса, например, это слово находится в ряду прилагательных в статье «Музыка», где с ним соседствуют — *мелодичный, гомофонический, консонирующий, созвучный, аккордный, гармонический, динамический, полухроматический, энгармонический, полифонический, характерный, фугированный* и т. д. Всего статья включает 508 слов, каждое из которых в своей дефиниции (во второй части словаря Касареса) содержит — либо в виде ключевого слова, либо в виде пометы — ссылку на понятие «музыка»<sup>3</sup>. Таким образом, тезаурус расширяет число возможных «точек доступа» к словарному составу языка, увеличивая тем самым размерность семантического пространства лексики: при обязательности алфавитного входа, без которого немыслим, очевидно, никакой словарь и который обеспечивает переход от «слова» (через «значение», как в словаре Касареса, или минуя его) — к «понятию» (или от «знака» — к «концепту»), для него конституирующими является вход через понятие или тему, т. е. он даст и возможность перехода от «понятия» к «слову» (от «концепта» — к «знаку»).

Именно эта черта тезауруса была всегда главным аргументом при обосновании необходимости построения подобных словарей, причем их авторы подчеркивали pragматический аспект отношения концепт—знак, а именно, возможность использования такого словаря для поисков выражения, слова, адекватно передающих соответствующее

---

\* С помощью одной какой-либо идеи никогда не удается наилучшим образом использовать всю наличную информацию (1-й закон де Бено) (*E. de Бено. Практическое мышление*).

<sup>1</sup> Морковкин В. В. Идеографические словари. М., 1970.

<sup>2</sup> Proceedings of the International conference on general principles of the thesauri building. Warsaw, 1970, p. 160.

<sup>3</sup> Casares J. Diccionario ideológico de la lengua española. Barcelona, 1951.

понятие или какой-то отдельный его признак. В предисловии к словарю Роже говорится: «Писатель или любой другой человек, который при составлении связного текста испытывает муки слова, понимая, что он хочет сказать, но не находя нужного в данный момент выражения, с удовлетворением воспользуется этим трудом, где найдет то, в чем он нуждается — богатый и легко доступный словарь, в котором слова точно квалифицированы, разбиты на группы и расположены по рангам. Иными словами, этот человек должен обратиться за консультацией к тезаурусу»<sup>4</sup>. Автор другого идеографического словаря — Х. Касарес — рассказывает про академика, который хотел в обычном толковом словаре найти слово, обозначающее того, кто родился на корабле во время плавания. Эта задача оказывается практически непосильной, поскольку для ее решения в алфавитном словаре надо не просто перебрать 200 тыс. входных слов, но просмотреть, хотя бы бегло, и их толкования, т. е. прочитать весь словарь<sup>5</sup>. Если же в нашем распоряжении имеется словарь другого типа, предусматривающий вход от понятия или темы (в данном случае «корабль» — *embarcación*, *navío*, *nave*, *nao* . . .), то мы без особого труда в числе слов, связанных с исходным понятием, обнаружим искомое. Это слово — *naonato* — в словарной статье *embarcación* входит в группу существительных, обозначающих лиц, связанных с кораблем и навигацией, — *armador*, *naviero*, *bajelero* . . .; в статье же *navegación* ‘судоходство’ оно встречается в той же группе существительных, но выступающей здесь в значительно расширенном виде. Одновременно ‘рожденный на корабле во время плавания’ включается и в перечень слов, относящихся к понятию «рождение» (*nacimiento*), но отсутствует в группе слов, характеризующих понятие «плавание» (*natación*), что, вероятно, справедливо, поскольку *naonato* не принадлежит к числу обязательных свойств этого понятия.

Итак, центральное место во всяком тезаурусе составляет его систематическая часть, которая представляет собой алфавитно упорядоченный перечень дескрипторов I (т. е. основных понятий с относящимися к ним словами) и ключевых слов (тоже являющихся разновидностью основных понятий и занимающих в систематической части свое место по алфавиту, но лишенных самостоятельной «дескрипторной» статьи, поскольку относящиеся к ним слова целиком включаются в подчиняющий их дескриптор).

Различие между дескрипторами и ключевыми словами, достаточно четко устанавливаемое в специальных тезаурусах, определяется тем, что последние, широко встречааясь в текстах определенной отрасли знаний, выступают в роли синонимов (квазисинонимов) к дескрипторам и снабжаются в словаре особыми пометами: ИСП (т. е. «использовать») — при отсылке к соответствующему дескриптору и ИВ (т. е. «использовать вместо») — при указании на квазисинонимические отношения с данным дескриптором. Например<sup>6</sup>:

Дыхание «рот в рот»	Дымовые боевые части
ИСП Искусственное	ИСП Боевые части ракет
дыхание	и Дымовые боеприпасы
Искусственное дыхание	Дымовые боеприпасы
ИВ Аппаратное	ИВ Дымовые боевые части
искусственное дыхание	Дымовые бомбы
Дыхание «рот в рот»	Дымовые гранаты
Управляемое дыхание	Дымовые мины
Ботаника	Дымовые снаряды
ИСП Растения	Дымообразующие
	боеприпасы

В тезаурусах же неспециальных, т. е. общеязыковых, введение ключевых слов в систематическую часть обусловлено стремлением увеличить количество «точек доступа» к лексическому составу и с максимальной полнотой вовлечь в систематическом указателе классификационную схему или указатель иерархических отношений дескрипторов (там, где подобные схемы вообще имеются, о чем подробнее см. ниже). Так, в словаре Х. Касареса к упомянутым выше дескрипторам в систематической части отсылают ключевые слова: Navegante (V. Navegacióñ), Navegar (V. Navegacióñ), Natal (V. Nacimiento). Вообще появление ключевого слова в систематической части является здесь сигналом того, что в соответствующей дескрипторной статье выделяется компактная и достаточно целостная подгруппа слов, которая не образует еще самостоятельной статьи, но объективно должна быть отмечена как некоторое единство, озаглавленное этим ключевым словом. Например, в упоминавшейся уже статье «Музыка»

<sup>4</sup> Roget P. M. Roget's thesaurus of English words and phrases. Penguin Books. Harmondsworth, 1978, p. VIII.

<sup>5</sup> Касарес Х. Введение в современную лексикографию. М., 1958, с. 7.

<sup>6</sup> Тезаурус научно-технических терминов. Под ред. Ю. И. Шематкина. М., 1972.

в качестве отсылочных выделены понятия: *canto* 'пение', *afinar* 'настраивать', *cantar* 'петь', *canción* ' песнь', *cantante* 'певец', *instrumentalista* 'инструменталист', *instrumento* 'инструмент', *danza* 'танец', *cante* 'песня'. Из них дескрипторами, т. е. заголовочными словами самостоятельных статей, оказываются менее половины, тогда как остальные имеют статус ключевых слов и служат промежуточными звенями для связи с другими дескрипторными статьями. Соответственно при обращении к систематической части мы находим: *afinar* (V. *Afinación*), *cantar* (V. *Canto*), *canción* (V. *Canto*), *cantante* (V. *Canto*), *cante* (V. *Canto*). Из этого перечня *cantante*, например, возглавляет в своей статье (*Canto*) целую серию слов, обозначающих разного рода певцов:

<i>cantor, ta</i>	певец, кантор;	<i>rapsoda</i>	рапсол
<i>cantista</i>	певица	<i>bayadera</i>	баядера
<i>cupletista</i>	певец	<i>zarabandista</i>	поющий и танцующий
<i>corista</i>	куплетист		сарабанду
<i>cantador</i>	хорист	<i>madrigalista</i>	поющий мадригалы
<i>juglar</i>	певец		
<i>payador</i>	трубадур	<i>romancero</i>	исполнитель
<i>pallador</i>	гаучо, поющий		романсов
	под гитару	· · · · ·	· · · · ·
	куплетист,		
	бродячий певец		

Таким образом, ключевые слова являются теми нитями, которые связывают внутреннюю структуру тезауруса, т. е. организацию словарной статьи, с его внешней структурой, образуемой парадигматическими связями дескрипторов.

Завершая рассмотрение и характеристику двух входов в тезаурус — систематического (воплощающего отношение концепт—знак) и алфавитного (отношение знак—концепт), отметим еще одну их особенность. Как правило, они выполняются в тезаурусах конструктивно раздельными: в словаре Касареса это «аналогическая» и алфавитная части, у Роже — Reference Book и Index, в минимальном идеографическом словаре (МИС) — имена полей с относящимися к ним словами и ключ к МИС<sup>7</sup> и т. п. Причем алфавитный вход может быть представлен либо в виде простого индекса слов, либо в виде толкового словаря, как, например, у Касареса, выполняя в последнем случае не только роль указателя к систематической части, но и давая каждому слову определение. Однако принципиальная похожесть представления в тезаурусе элемен-

тов систематической и алфавитной части делает возможной их конструктивную совместимость. В случае, если обе части соединяются в одном входе и для различения между дескриптором, ключевым словом и словом как элементарной единицей вокабуляра используются разные шрифты, особые пометы и разный объем информации, мы имеем дело с аналогическим словарем. Последний является как бы вырожденным тезаурусом: проигрывая тезаурусу в компактности, достаточной строгости и обозримости систематической части, аналогический словарь обладает тем преимуществом, что в нем вся информация о слове сосредоточена в одном месте — в словарной статье. Здесь и семантическая его характеристика, и стилистические ограничения, и иллюстрации употребления, и указанные с помощью целой сети перекрестных ссылок системные связи (синонимические, антонимические, омонимические, паронимические, ассоциативные), и грамматическая его квалификация, и словообразовательные отношения, и этимологический параметр, и некот. др. Таков словарь М. Молинер<sup>8</sup>, в котором «рядовые», так сказать, слова снабжены минимумом информации и определяются либо ссылками, либо через соответствующий дескриптор:

aposto. V. bajo «apostar»  
bravamente (liter.) Con bravura. ○ Muy \*bien Q \*mucho:  
‘Bravamente homos comido’...  
bravura. Cualidad do bravo (valiente). ○ Bravata

Ясно, что в этой цепочке определений решающим является слово *bravo*, которое выступает в роли дескриптора и которому посвящена полная словарная статья, содержащая весь набор перечисленных выше параметров. Подобным образом организован материал и в известном аналогичном словаре Робера<sup>9</sup>.

Логически возможные комбинации из двух элементов — концепта и знака — не исчерпываются рассмотренными нами соотношениями, а допускают еще два сочетания: концепт—концепт и знак—знак. Как мы покажем ниже, обе эти возможности тоже реализуются в тезаурусе.

Иллюстрацией того, как осуществляется переход от понятия к другому понятию (т. е. концепт—концепт),

<sup>7</sup> Караплов Ю. Н. Общая и русская идеография. М., 1976, с. 276—315.

<sup>8</sup> Moliner M. Diccionario de uso del español, t. 1—2. Madrid, 1966.

<sup>9</sup> Robert P. Dictionnaire alphabétique et analogique de la langue française. Paris, t. I—VI, 1976.

могли бы уже служить примеры из словаря Х. Касареса, приведенные в связи с обсуждением вопроса о ключевых словах и дескрипторах. Однако, будучи растворенными в систематической части, такие отношения, так же как и перекрестные отсылки в дескрипторных статьях, еще не образуют самостоятельного входа в тезаурус и не дают основания говорить об особом «измерении»<sup>10</sup> словарного состава языка. Это измерение возникает только в том случае, если тезаурус располагает классификационной схемой (блок-схемой) понятий, которая явным образом, последовательно и полно фиксирует парадигматические отношения между дескрипторами. Схема объединяет дескрипторы в блоки, дескрипторные области, фасеты и содержит указания подчиненности дескрипторов (выше-ниже), в том числе полного включения одного в другой их противопоставленности, отношений части и целого и т. п. (так называемые «сильные» парадигматические отношения), а также сходства, смежности, квазисинонимии («слабые» парадигматические отношения). Классификационная схема, наряду с алфавитной и систематической частями, представляет собой третий вход в тезаурус, а реализована она может быть разными способами.

В словаре Х. Касареса классификационная схема, составляющая здесь синоптическую часть, так сказать синоптический вход в тезаурус, строится на двух уровнях: общий план классификации<sup>11</sup>, результатом которого стали 38 дескрипторных областей (38 квадратов, по терминологии автора), и собственно дескрипторное (попутное) наполнение каждой области. Например, квадрат № 5, посвященный растительному миру и названный на схеме Х. Касареса «Ботаника», включает 85 дескрипторов, которые объективным образом делятся на несколько дескрипторных блоков. Блоки не имеют своих названий но относящиеся к каждому из них группы дескрипторов отделены при заполнении квадрата друг от друга пробелами. Первый блок охватывает самые общие понятия («выше»), которые характеризуют практически каждый вид растительности: *растительность, семя, корень, стебель, ветка, почка, лист, цветок, плод...*. Затем дается разработка видов растений («ниже»), основанная не на строго научной их классификации, а на повседневно-бытовом восприятии различий между ними, в таком порядке (напомним, что названия групп даны нами условно у Касареса их нет, но они достаточно точно отражают

Содержание его дескрипторных блоков): мхи (споровые) — *мох, плесень, гриб, водоросль, папоротник*; травы — *трава, луг, шафран, овощи, лук, спаржа, капуста, фасоль*. . .; злаки — *ишеница, рожь, овес, рис.* . . .; далее следуют растения, которые нельзя отнести ни к травам, ни к кустарникам — *кактус, агава, тростник, пальма*; кустарники — *куст, виноградная лоза, малина, роза*. . .; деревья — *лес, сосна, дуб, смоковница, яблоня, каштан*. . . Если оценивать структурирование этой дескрипторной области с точки зрения того, какой тип анализа соответствующей части объективной действительности, окружающего нас мира лежит в его основе — логический, прагматический или ассоциативный <sup>12</sup>, то следует констатировать, что решающим в данном случае является, очевидно, прагматический подход, опирающийся в своих выводах на критерии здравого смысла. Под большим влиянием логики сделана аналогичная схема в словаре Дорпзайфа:

7.1. Растение. 7.2. Виды растений. 7.3. Части растений. 7.4. Болезни растений. 7.5. Разведение растений. 7.6. Плодородие. 7.7. Бесплодие. Однако, будучи слишком общей, она затрудняет переход от понятия к понятию, проигрывает схеме Касареса в конкретности. Почти полным переносом принципов ботанической таксономии на эту часть лексики, т. е. преобладанием логического аспекта, характеризуется соответствующая схема в словаре Бабова и Воргулева, где для получения классификации используются приемы построения иерархических деревьев. В схематической записи это отражается постановкой черточек перед дескриптором, причем количество черточек обозначает число узлов дерева, предшествующих его появлению, или число оснований деления «старшего» дескриптора:

---

<sup>10</sup> Каждое из четырех отношений (знак—концепт, концепт—знак, концепт—концепт и знак—знак) можно расценивать как отдельное измерение лексической системы, и тогда полный тезаурус представляет словарный состав языка в четырехмерном семантическом пространстве, заданном этими отношениями.

<sup>11</sup> Эта схема в переводе на русский язык приводится, например, в кн.: *Морковкин В. В. Указ. соч., вклейка между 32 и 33 стр.*; или в кн.: *Караулов Ю. И. Указ. соч., с. 248—249.*

<sup>12</sup> В комбинации этих трех аспектов создаются все классификационные схемы тезаурусов. См., напр.: *Соколов А. В. Методические материалы по разработке информационно-поисковых тезаурусов. Л., 1975, с. 22-30.*

## РАСТЕНИЯ

### Декоративные растения

- огородные и парковые растений
- == травянистые
- = кустарники
- = деревья
  - == лиственные
  - == хвойные
- комнатные растения<sup>13</sup>

Как естественное продолжение этого ряда — классификационных схем в тезаурусах — можно было бы рассматривать и структуру соответствующей части уже цитированного «Тезауруса научно-технических терминов», только при этом необходимо иметь в виду принципиально различающиеся установки общеязыковых и терминологических тезаурусов: первые исходят из задачи возможно полно и притом с достаточной степенью детальности описать, представить лексически некоторую область объективной действительности, вторые же имеют целью анализ и логически непротиворечивое представление существующей и общепринятой терминологии данной отрасли. Отсюда становятся понятными, во-первых, невозможность такой детализации дескрипторов в терминологических словарях, какая есть, например, в словаре Касареса, а во-вторых, появление в классификационных схемах общеязыковых тезаурусов наряду с полноправными, так сказать, дескрипторами, и ключевых слов, которое оправдывается здесь стремлением равномерно, без лакун заполнить лексикой данную дескрипторную область и которое на основе подобной предпосылки было бы недопустимым в терминологическом тезаурусе. Напомним, что к ключевому слову привязана замкнутая, как правило, группа лексических единиц внутри соответствующей дескрипторной статьи. Эта группа ключевого слова покрывает фрагмент семантического пространства той дескрипторной области, в рамках которой это ключевое слово фигурирует, тогда как вся дескрипторная статья, к которой адресуется отсылка при ключевом слове, может относиться и к другой дескрипторной области (другому квадрату). Таким образом, ключевое слово, служа в систематической части структурированию дескрипторной статьи, облегчающему поиск нужной лексической единицы, одновременно в классификационной схеме выполняет роль

звена, обеспечивающего непрерывность членения (описания) семантического пространства. Например, в квадрате № 33 у Касареса, названном «Собственность» (имущество, достояние) — Propiedad, используются следующие ключевые слова:

— в дескрипторном блоке Adquisición 'приобретение', Toma, tomar 'взятие, получение', Admisión 'принятие', Herencia 'наследование, наследство' таким словом является Admisión, отсылающее к более общему дескриптору той же области — Recepción 'прием';

— в другом дескрипторном блоке: Provisión 'снабжение', 'запасы', Acopio 'накопление', Monopolio 'монополия', Transmisión 'передача', Trasiego (V. Permuta) 'мена' (см. обмен), Subasta 'аукцион', Compraventa 'купля—продажа', Préstamo 'ссуда, заем' отсылка от ключевого слова дается к дескриптору, относящемуся уже к 35-му квадрату (Comercio 'торговля'), в котором повторяются и некоторые другие дескрипторы этого блока;

— наконец, в блоке Gratuidad 'даровой, бесплатный', Depreciación, barato 'обесценивание, дешевый', Ganga (V. Ventaja) 'находка, легкая добыча' (см. выгода) отсылка дана к дескриптору того же квадрата.

Задаче полного, непрерывного и без остатка членения дескрипторной области в классификационных схемах служит также прием антонимического противопоставления дескрипторов и дескрипторных блоков. Так, в первом из рассмотренных блоков только одному дескриптору Adquisición противопоставлен целый «антиблок»: Privación 'лишение, отнятие', Robo 'грабеж, кража', Fraude, estafa 'обман, мошенничество', Comiso 'конфискация', Embargo 'наложение ареста'; и наоборот, второму из рассмотренных блоков (Provision) противопоставлен лишь один дескриптор — Carencia 'отсутствие, нехватка' и т. д.

Думается, что приведенные здесь примеры дают известное представление об организации и характере этого третьего входа в тезаурус, третьего измерения лексической системы, которое воплощается в классификационной схеме, дает членение (или воспроизводит структуру) определенной области (действительности, знаний), устанавливает системность отношений на уровне «концепт—концепт» и обеспечивает при поиске оптимальный переход от од-

---

<sup>13</sup> Бабов К., Въргулев А. Тематичен руско-български речник. София, 1961.

ного понятия к другому, от дескриптора — к дескрипторному блоку, а от него к смежной дескрипторной области, и наоборот. Вместе с тем мы можем уже говорить и о некоторых особенностях классификационных схем — как определенного уровня организации словарного материала и как самостоятельного входа в тезаурус. Прежде всего, классификационная схема, будучи, с одной стороны, подчинена «логике мира», а с другой — логике поиска, представляет собой такой уровень обобщения, который не имеет непосредственного выхода на тексты. Прямой переход к текстам и даже развертывание в связный текст<sup>14</sup> возможны лишь от словесного наполнения дескрипторной статьи, т. е. на уровне систематической организации тезауруса, в том измерении лексической системы, которое обозначается переходом «концепт—знак». Однако, благодаря конструктивному свойству тезаурусов, которое позволяет совмещать входы в него, классификационная схема может выполнятся заодно с систематической частью и брать тем самым на себя ее функции. В этом случае мы имеем дело с тематическим словарем.

Тематический словарь представляет собой монтаж из блоков-тем, каждая из которых допускает лексическое наполнение любого объема и развертывается в связный текст, раскрывающий соответствующую тему с большей или меньшей глубиной. Эта разновидность тезауруса характеризуется двумя входами: от классификационной схемы (например, «Жизнь и деятельность человека» → IX. Человек. → 7) Головные уборы → шапка → новая шапка, красивая шапка, теплая шапка, зимняя шапка, дорогая шапка, примерить шапку, носить шапку, ходить в шапке, надеть шапку, снять шапку; шляпа → . . . )<sup>15</sup> и от лексического наполнения темы, т. е. обычным алфавитным входом. Из трех принципов, определяющих установление парадигматических отношений между дескрипторами (в данном случае — темами), а именно — логического, прагматического и ассоциативного, — решающим здесь является прагматический, что полностью отвечает и учебным задачам такого тезауруса: «Пользуясь словарем, можно выполнять разнообразные виды устных и письменных работ. Так, на уроке, при активном участии всего класса можно составлять рассказы типа „Наша семья“, „Зима“, „Лето“, „В библиотеке“ и т. п., используя словосочетания и предложения, имеющиеся в словаре. По материалам словаря легко организовать занятия-беседы на все

важнейшие темы, актуальные для учащихся, можно выполнять различные письменные работы — сочинение, со-ставление рассказов»<sup>16</sup>.

Последнее из рассматриваемых отношений — «знак — знак» также может реализоваться в качестве самостоятель-ного (четвертого) входа в тезаурус. Он выполняется в виде пермутационного указателя терминов (дескрипторов) и предназначен для облегчения и ускорения поиска тер-минов и терминосочетаний, а также может использоваться для «доводки», совершенствования структуры тезауруса<sup>17</sup>. «С точки зрения построения раздел Пермутационный указатель терминов представляет собой расположение по ал-фавиту собрание всех значащих слов, входящих в состав терминов тезауруса. После каждого значащего слова приводятся в алфавитном порядке все термины тезауруса, содержащие это значащее слово в качестве терминоэле-мента. Таким образом, каждый термин встречается в Пер-мутационном указателе столько раз, сколько значащих слов он содержит»<sup>18</sup>. Например, если обратиться к бота-нической области, которую мы уже затрагивали выше по разным поводам, то в «Тезаурусе научно-технических терминов» мы обнаружим следующие группы:

#### Растений

Бактериальные возбудители болезней растений  
Вирусы растений  
Вредители растений

#### РАСТЕНИЯ

Лекарственные растения  
Ядовитые растения

#### Растительная

Растительная маскировка

#### Растительного

Снятие растительного слоя **грунта**

#### Растительные

Растительные масла

Растительные пигменты<sup>19</sup>

<sup>14</sup> См., напр.: *Караулов Ю. Н.* Указ. соч., с. 218—219.

<sup>15</sup> *Тихонов А. Н.* и др. Русско-узбекский тематический словарь. Ташкент, 1975.

<sup>16</sup> Там же, с. 10.

<sup>17</sup> *Веревченко А. П., Петров Ю. Г.* Применение пермутационного указателя для анализа и совершенствования тезаурусов. — НТИ, сер. 2, 1968, № 3.

<sup>18</sup> Тезаурус научно-технических терминов. Под ред. Ю. И. Шема-кина, с. 8.

<sup>19</sup> Там же, с. 596,

А соответственно первый из приведенных терминов появится в пермутационном указателе еще три раза: в группе «бактериальные», в группе «возбудители» и в группе «болезней». Таким образом, этот вход оказывается сочетаемостным, или синтагматическим.

Понятно, что значимость рассматриваемого входа в тезаурус, или измерения семантического пространства, ослабевает по мере приближения единиц в тезаурусе к унитермным. Так, практически унитермным является наполнение тезауруса Касареса: хотя в дескрипторную статью здесь включаются иногда так называемые модизмы (т. е. фразеологизмы, крылатые выражения, поговорки и другие речения), их использование не является последовательным и не охватывает всего фразеологического состава испанского языка, а самое-то главное, — и функционально и семантически эти устойчивые сочетания эквивалентны слову, т. е., по сути дела, тоже унитермны. Поэтому для данного тезауруса отпадает необходимость в пермутационном указателе, тем более, что совмещение алфавитного входа с толковым словарем оставляет место для включения в дефиницию соответствующего слова устойчивых речений с ним, которые адресуются к определенным дескрипторам («концептам») в систематической части, связанным уже не с их оформлением, а с семантикой. Например, в дескрипторной статье *Juventud* 'молодость' приводятся следующие речения: tener pocas barbas букв. 'иметь маленькую бороду' и tener la leche en los labios 'иметь молоко на губах'. В дефиниции же слова *barba* (в «заромбовой» ее части) мы находим: Tener uno pocas barbas. — Tener pocos afios o ser inexperto 'иметь мало лет или быть неопытным', где *pocos afios* есть перифраза понятия «молодой» — *joven*, «молодость» — *juventud*, а *inexperto* является ключевым словом, от которого в систематической части словаря дана ссылка к дескрипторной статье *Tnexperejencia* 'неопытность, незрелость'. Последняя в свою очередь содержит рассматриваемое речение в числе составляющих ее единиц. Аналогичным образом обстоит дело со вторым сочетанием, и при толковании этого выражения, в дефиниции слова *leche* 'молоко', мы находим те же ключевые слова: Estar uno con la leche en los labios — Ser joven e inexperto.

Необходимость оформления пермутационного входа в (нетерминологический) тезаурус как самостоятельного возникает в том случае, когда единицами тезауруса ста-

новятся синтагмы. Таковы словари поговорок и пословиц, афоризмов, которые строятся, как правило, по тематическому принципу, таким может стать и мыслимый фразеологический словарь, в котором пермутационный указатель играл бы роль дополнительного по отношению к понятийному (т. е. ономасиологическому, «от концепта») входу и который пока не создан ни для одного языка. Интересен опыт организации пермутационного входа к изданию таджикских народных рубои: «Научный аппарат завершается регистром мотивов, взятых из текста вариантов по опорным словам. . . Этот указатель служит не только ключом к содержанию всех публикуемых 8793 записей 394 произведений, но является также самостоятельным, обобщенным на уровне репертуара, совершенно новым исходным материалом и важным инструментом для углубленного изучения содержания, образов, символов и языка рубои — этого богатейшего, многовекового вида поэзии на таджикско-персидской почве. . .

Регистр опорных слов благодаря ссылкам на произведения указывает частоту, а следовательно, в известной мере и значимость определенных слов и понятий, вся синонимия которых легко вскрывается в контексте публикуемых произведений. Регистр обладает многими преимуществами частотного словаря, тезауруса или конкордансии, что, безусловно, представит особый интерес для филологов, лингвистов и историков культуры»<sup>20</sup>. Таким образом, для четверостишия, например

Если луной станешь, на твоё лицо не взгляну,  
Если водой станешь, на твой ручей не взгляну,  
Если букетом станешь в день сто раз,  
Ей богу, возьму и не буду шохать<sup>21</sup>,

в пермутационном указателе мы найдем слова: *луна*, *лицо*, *взглянуть*, *вода*, *ручей*, *букет*, *день*, *взять*, *нюхать* с их точными адресами, т. е. зашифрованным указанием номера предметной области, рубрики, номера рубои и строчки.

Процесс перехода «знак—знак» не связан рамками понятийной общности двух знаков, хотя таковая может и иметь место, но чаще осуществляется на основании

---

<sup>20</sup> Свод таджикского фольклора. Рубои (четверостишия). В 3-х т. Проспект. Душанбе, 1970, с. 4 (ротапринт).

<sup>21</sup> Там же, с. 31.

каких-то формальных признаков. Тем не менее мы готовы в подобных случаях констатировать определенную связь между двумя знаками. Так, в последнем примере *луна* и *лицо* связаны как элементы постоянной метафоры в восточной поэзии, *вода* и *ручей* принадлежат к одному семантическому полю, между *букет* и *нюхать* можно установить связь по смежности. В рассмотренных выше испанских примерах тоже ощущается некая глубинная антонимия в паре *barba* — *joven* и интуитивная близость между *leche* и *juventud*. Когда мы не можем указать точных предпосылок семантической связи между двумя единицами, т. е. установить общие компоненты в их значении или выделить логические либо прагматические основания их связи, мы апеллируем к ассоциативным закономерностям.

Индивидуальные ассоциации непредсказуемы, более того, именно ассоциативная память при дивергентном мышлении считается одной из предпосылок творческого потенциала личности. Фантастичность, неожиданность персональных ассоциаций можно продемонстрировать на примере одного эпизода из пьесы В. Розова «Ситуация». Герой пьесы — изобретатель, ломающий голову над проблемой создания втулок для скоростных станков, после бурной сцены объяснения со своими друзьями, осудившими конформность его поведения, вынужден согласиться с их доводами и направляется на завод для решающего разговора с директором. И для действующих лиц, и для зрителей этот предстоящий (но так и не состоявшийся) разговор является доминантой эпизода, да, пожалуй, и всей пьесы. Поэтому дальнейшее поведение героя, о котором мы узнаем по репликам персонажей, оставшихся в комнате (на сцене) и наблюдающих за ним из окна, воспринимается и ими, и зрителями как колебание нерешительного человека перед трудным, серьезным объяснением: «Остановился . . . подошел к скамейке. . . посидел. . . пошел дальше . . . повернулся к доске объявлений, читает . . . возвращается назад?!» Тем неожиданнее оказывается следующая мизансцена, когда герой входит и сообщает, что решил проблему втулок: «Понимаешь, продаются попугайчики! Какие, думаю, попугайчики? — Заинтересовался, читал объявление дальше. — В Подольске. . . Так ведь мы же туда ездили в командировку . . . а там есть „Завод пластмасс“! Нашел! — Втулки из пласти массы!» Напряженная мысль исследователя, даже занятая

предстоящим разговором, стремится — в соответствии с преобладающей у него установкой — все, с чем он сталкивается, поставить в связь с доминантой актуального (потому и «заинтересовался») семантического поля: «втулка» — «попугайчики» — связи нет; «втулка» — «Подольск» — связь начинает нащупываться, пока отдаленная, слабая; «Подольск» — «Завод пластмасс» — «пластмассовая втулка» — это именно то, что он искал, ведь втулки из пластмассы смогут выдерживать высокие скорости, не будут «гореть».

Тем не менее, при всей причудливости, нестандартности индивидуальных ассоциаций некоторая их часть оказывается общей для всех носителей языка и может быть выделена в качестве статистически обоснованной ассоциативной нормы. Объективность ее существования можно показать, сравнив словарные статьи двух ассоциативных словарей, при подготовке которых и условия эксперимента, и контингент испытуемых, и их число, так же как способы обработки результатов, были разными. Однако при всем этом многие словарные статьи того ассоциативного словаря, который из двух сравниваемых может быть назван ««малым»<sup>22</sup>, целиком входят (находятся в отношении полного включения) в статьи большего по объему словаря<sup>23</sup>. В «малом» словаре статья включает 14 ассоциатов, или слов-реакций, на стимул ДОМ, в «большом» их 88. Общими из 14 являются 11 ( *большой, высокий, красивый, квартира, кирпичный, крыша, здание, семья, белый, наш, мой*), два находятся друг с другом в ассоциативной связи, так сказать, ассоциации второго шага (*дым — очаг, забор — двор или забор — сад*), и лишь одно слово-реакция, ассоциированное, вероятно, на основе рифмы, не находит себе соответствия в статье «большого» словаря — *том*.

В этих условиях у нас есть как будто основания трактовать ассоциативный словарь как вырожденный тезаурус, в котором пермутационный вход совмещен с систематическим и который, таким образом, обладает всего двумя входами: наряду с указанным совмещенным, еще и алфавитным, т. е. индексом слов-ассоциатов с адресацией к соответствующим словам-стимулам. Совмещение пермута-

<sup>22</sup> Титова Л. Н. Киргизско-русский ассоциативный словарь. Фрунзе, 1975, с. 49.

<sup>23</sup> Словарь ассоциативных норм русского языка. Под ред. А. А. Леонтьева. М., 1977, с. 92.

ционного входа, реализующего отношение «знак—знак», с систематическим, воплощающим переход «концепт—знак», объясняет появление в словарной статье ассоциативного словаря, с одной стороны, слов-реакций, связанных со стимулом сочетаемостными признаками, т. е. совместной встречаемостью в минимальных контекстах, с другой стороны, — слов-реакций, связанных с ним отношениями род—вид, часть—целое, элемент—система, причина—следствие, действие—орудие и т. п., т. е. отношениями, характерными для дескриптора и слов, составляющих дескрипторную статью. Следовательно, формулу ассоциативного словаря как разновидности тезауруса можно представить в таком виде:

знак (+ концепт) ↔ знак.

Наличие в ассоциативном словаре двух входов, объединяющих три измерения семантического пространства, не исключает возможности использования и четвертого измерения (в данном случае «концепт—концепт») в качестве самостоятельного входа. Естественно, что установление парадигматических отношений между словами-стимулами должно опираться не на логический и даже не на pragматический принцип, а главным образом на ассоциативный. Опыт построения такой классификационной схемы из слов-стимулов осуществлен на материале английского языка; она выполнена в виде иерархических деревьев (прямолинейных графов), объединенных в сети словесных ассоциаций<sup>24</sup>. В этом варианте ассоциативного тезауруса три традиционных входа (через классификационную схему, т. е. «концепт → концепт»; от стимула — через систематическую часть, совмещенную с пермутационным указателем, т. е. «знак (+ концепт) → знак»; от реакции, т. е. через алфавитный вход, резюмируемый формулой, обратной предыдущей, — «знак → знак (+ концепт)») дополнены еще частотным, где слова-реакции расположены по убывающей частоте их появления в ответах испытуемых.

Итак, подводя итог анализу устройства и принципов функционирования тезаурусов, мы можем резюмировать их рассмотрение следующим образом. Структура тезауруса представляет собой своеобразную семантическую матрицу с четырьмя потенциальными входами. Эта матрица задает план содержания лексики, лексическую систему языка (или подъязыка — в случае отраслевых тезаурусов) в четырех измерениях: «концепт—знак», «концепт—

концепт», «знак—концепт», «знак—знак», первое из которых является типологически определяющим для данного типа словаря. Соответственно в полном объеме тезаурус состоит из четырех частей — систематической, или лексико-семантического указателя терминов; классификационной, или указателя иерархических отношений дескрипторов; алфавитного указателя терминов; пермутационного указателя терминов (дескрипторов). В известных разновидностях тезаурусов эта структура может быть представлена в редуцированном, свернутом виде. Так, устройство аналогического словаря, в котором пермутационный указатель всегда отсутствует, систематическая часть (вход) совмещена с алфавитной, а наличие классификационной схемы факультативно (например, в словаре М. Молинер она есть, а у П. Робера отсутствует), отражается формулой:  
знак (+ концепт) ↔ концепт.

Наоборот, структура тематического словаря, предполагающая объединение классификационной схемы с систематической частью и факультативность использования пермутационного входа, характеризуется формулой:  
концепт (+ концепт) ↔ знак.

Что касается ассоциативного словаря, в котором классификационная схема обычно не используется, хотя принципиально и допустима, а пермутационный указатель исключается по самому своему статусу, то его формула была обоснована и приведена выше:  
знак (+ концепт) ↔ знак.

Если вход в тезаурус (или, что то же — измерение семантического пространства лексической системы, заданного с помощью тезауруса) совпадает с одной из возможных структурных его частей и определяется тем или иным отношением концепта и знака друг с другом, то непосредственным инструментом входа является единица соответствующего раздела, т. е. имя. В систематической части это имя дескриптора или ключевого слова, в алфавитной — имя термина либо отдельное слово как самостоятельный элемент вокабуляра, в классификационной схеме это имя дескрипторной области и в пермутационном ука-

---

<sup>24</sup> Kiss G. et al. An associative thesaurus of English and its computer analysis. — In: The computer in linguistic and literary studies. Edinburgh, 1973.

зателе — слово как часть составного именования, или терминоэлемент.

Единицы основного, т. е. систематического, входа могут быть разбиты на классы, и число «входов» (а вернее сказать — «подвходов») может тем самым расти. Классы эти выделяются на разных основаниях: имена собственные и нарицательные; внутри имен собственных — топонимы, гидронимы, зоонимы, антропонимы, номенклатурные знаки, марки и т. п.; внутри нарицательных — тематические объединения, синонимы, омонимы, антонимы, термины, идиомы и т. д.; в свою очередь, антропонимы, например, могут подразделяться опять на тематические группы — имена известных деятелей театра, науки, кино, военного дела, а зоонимы — на клички (известных, т. е. породистых) лошадей, собак и т. п. Потребность в увеличении числа входов, точек доступа к лексическому составу языка (подъязыка) возникает при необходимости оперативного извлечения информации из словаря.

### § 5. СФЕРЫ ИСПОЛЬЗОВАНИЯ ТЕЗАУРУСА, ИЛИ ЗАДАЧИ, КОТОРЫЕ ОН РЕШАЕТ

Просвещение внедрять с умеренностью, по возможности избегая кровопролития.

*М. Е. Салтыков-Щедрин. История одного города*

Вопрос «зачем нужен» тезаурус, в частности тезаурус литературного языка, мог бы показаться неуместным и даже кощунственным с точки зрения свободы научного творчества: в самом деле, к идею создания такого словаря ученых приводит сама логика развития общего языкоznания и лексикографии, а в рамках этой логики вовсе не обязателен ответ на вопрос «зачем». Для Х. Касареса, например, этот вопрос был второстепенным, но рассмотренный с позиции лингвоконструирования, он приобретает решающее значение и предполагает освещение и характеристику проблем — как прикладных, так и теоретических, — которые можно решать с помощью тезауруса.

#### Системные связи в лексике и лексическая система

Если попытаться собрать все высказываемые в современной литературе суждения о системности в лексике, то мы получим довольно неоднородную картину представ-

лений, которая заставляет вспомнить древнеиндийскую притчу о слепцах, пытавшихся описать слона, воспринимаемого ими изо всех сторон. Здесь и стремление подойти к трактовке системных связей от отдельного слова — через анализ его смысловой структуры и типизацию семантических отношений между его значениями; и попытка прочертить некоторые сквозные линии через весь словарный состав, расположив на них лексику определенными рядами, группами, гнездами (синонимия, антонимия, омонимия, паронимия, аллосемия); опыты построения отдельных ономасиологических групп с анализом взаимоотношений между составляющими их элементами и приемы типизации синтаксических связей и валентностей лексической единицы с последующей реконструкцией ее семантики по моделям сочетаемости; здесь, наконец, и опыты лексикографической интерпретации системных отношений — в словарях идеографического типа. Беда только в том, что все эти частные разработки обладают ограниченной силой, не имеют универсальной значимости.

Показательна в этом отношении составленная Г. Вольфом обобщающая схема выделяемых и изучаемых в лексике системных отношений, которая должна, по мысли автора, иллюстрировать «систему». На дело схема, включающая 16 отдельных пунктов, — по сути дела, самостоятельных, не соотносящихся друг с другом измерений (лексическое гнездо, омонимы, значения полисемантического слова, синонимы, территориально-социальные дублеты, антонимы, разговорные варианты или коллоквиализмы, соотношения части и целого, семантическая валентность, степени интенсивности, ключевые слова, элемент и множество, тематический ряд, понятийная группа, поле, логические отношения) <sup>25</sup> — подкрепляет скорее противоположный тезис — об отсутствии единой системы в лексике.

Ситуация напоминает состояние биологии в преддарвиновский период, когда теория эволюции еще не была создана, или положение в химии в доменделеевское время, когда химические элементы группировались в пучки, триады, из них строились тетраграммы и другие привлекательные геометрические фигуры, но за всем этим не стояло единой и целостной системы. Нужна была «лест-

---

<sup>25</sup> Wotjak G. Zu den Systembeziehungen in der Lexik. — Beiträge zur Romanischen Philologie, 1975, XIV Jg., H. 1, S. 125—126.

ница естественного отбора» или периодическая система элементов, чтобы внести последовательную упорядоченность в хаос организованных каждая по отдельности групп. Так и в лексикологии ныне: из детально проработанных фрагментов не возникает отчетливого образа системы, который мог бы вылиться в краткую и членораздельную формулу, сконцентрировавшую, вобравшую в себя всю семантическую энергию, способную при обратном движении «расщепляться» и доводить исследователя от универсальной схемы до любого отдельного лексического элемента. Думается, что с мыслью (явно выраженной или подразумеваемой) о такой сокровенной формуле пишутся все лексикологические работы. Может статья, что подобной формулы вообще не существует, однако эта идеализация идет явно на пользу языкознанию, поскольку служит основанием хотя бы для методических приемов по сжатию и компактному представлению лексико-семантической информации, что само по себе немаловажно в науке.

Можно полагать, что тезаурус дает известное приближение к целостному представлению лексической системы в ее идеальном понимании: компактно обобщая ассоциативно-дивергентные связи слов (мы имеем в виду, что полисемия дивергентна, тогда как смысловая связь ассоциативна по своей природе, а статья тезауруса оптимальным способом соединяет то и другое), тезаурус увеличивает размерность семантического пространства лексики, задавая его четырьмя измерениями. Каждое из этих измерений (концепт—концепт, концепт—знак, знак—концепт, знак—знак) пронизывает лексику насквозь, а в своей совокупности они позволяют охватить весь словарный состав и найти определенное место каждому слову языка, опять-таки с учетом его синонимических, антонимических и прочих отношений. Напротив, явления синонимии, антонимики и т. п., взятые по отдельности, представляют собой системно между собой не связанные организационные стержни, параллельные линии в евклидовом пространстве, которые, никогда не пересекаясь, позволяют группировать лексику в разных направлениях, но не исчерпывают всего лексического состава и не образуют единого каркаса, организованного и замкнутого семантического пространства. Так, не все доминанты синонимических (антонимических) рядов могут быть дескрипторами, тогда как обратное — справедливо: любой дескриптор может выступать

в качестве доминанты в синонимическом словаре, а в идеографическом словаре (тезаурусе) он обязательно сопровождается некоторым множеством синонимов. Например, доминанта *похолодеть* ряда *похолодеть, помертветь, омертветь, оледенеть* или доминанта *твердить* ряда *твердить, долбить, гвоздить, талдычить, зудеть*<sup>26</sup> не имеют соответствия среди дескрипторов ни словаря Роже, ни словаря Касареса, тогда как выбранные наугад из первого словаря дескрипторы *obedience* 'послушание' с синонимами *observance, compliance, submission*. . . и *store* 'запас' с синонимами *stock, fund, mine*. . ., или из второго — *debilidad* 'слабость' с синонимами *debilitation, debilitamiento, aflojamiento* и *posesion* 'собственность' с синонимами *posesion, tenencia, tencion*. . .<sup>27</sup> совпадают с доминантами синонимических рядов в цитированном выше русском словаре (с учетом возможности перехода в другие части речи): *послушный, запасной, слабый, собственность*.

Дескрипторы, в противоположность доминантам, поддаются упорядочению в дескрипторные блоки и дескрипторные области (фасеты), которые, в свою очередь, способны образовывать «универсальную фабулу». Фабула (первоначально — акроним словосочетания «фасетно-анкетно-блочная унифицированная лексика») «представляет собой схему фасетов и входящих в них дескрипторных блоков. Фабула выполняет функцию путеводителя по систематической части тезауруса в процессе поиска. Она включает также анкетные и фасетные формулы, необходимые для индексирования. Поскольку в ней представлены в обозримой форме результаты упорядочения терминологии данной тематической области при создании тезауруса, эта схема получила название фабулы (сюжет, замысел)»<sup>28</sup>. От такого понимания структуры тезауруса уже недалеко до упоминавшейся выше идеальной универсальной формулы, тем более, что сама фабула предполагает возможность единого — для разных отраслей и общенационального языка — исполнения на самых общих ступенях членения лексики, т. е. до определенного

---

<sup>26</sup> Словарь синонимов русского языка, т. П. Л., 1971, с. 213, 536.

<sup>27</sup> Roget's thesaurus of English words and phrases. N. Y., 1979, p. 277, 221; / . Casares. Diccionario ideológico de la lengua española. Barcelona, 1951, p. 151, 445.

<sup>28</sup> Соколов А. В. Указ. соч., с. 38.

уровня детализации (универсальные фасеты — хронологический, пространственный, межотраслевой, формальный; предметные фасеты, процессуальные фасеты, инструментальные фасеты<sup>29</sup> и т. д.).

Тем не менее было бы неоправданным сверхупрощением представлять дело так, будто тезаурус воплощает саму систему в лексике. Отражая парадигматические связи слов, тезаурус в любой информационно-поисковой системе, любом банке данных функционирует как часть информационно-поискового языка наряду с некоторым комплексом принятых здесь синтагматических правил. Иными словами, тезаурус составляет статичную часть информационно-поискового языка, тогда как динамизм последнего достигается за счет правил преобразований, за счет правил линейных связей и переходов между единицами. Опыт эмпирического конструирования отраслевых тезаурусов показывает, что тенденция к упрощению составляющих тезаурус единиц влечет большую гибкость и глубину поиска, по одновременно и создает значительный информационный шум. Дело в том, что сумма смыслов 'составляющих терминосочетание слов не адекватна смыслу соответствующего многословного термина; при большом числе последних в той или иной тематической области синтаксис поискового языка оказывается неэффективным, и для успешного поиска требуется введение в тезаурус терминосочетаний, т. е. синтаксиса, так сказать, в застывшем виде. Аналогичным образом недостаточность учета только парадигматических отношений между словами при попытке моделировать лексическую систему литературного языка подтверждают и излагаемые в последующих главах результаты нашего опыта машинного построения русского тезауруса. Отказ от учета синтаксических свойств отдельного слова привел к значительным трудностям как на всем словаре (см. ниже «Феноменологическая модель тезауруса»), так и в трактовке семантики отдельной единицы. В масштабе всего словаря эти трудности были разрешены чисто конструктивным образом, т. е. на основе анализа их языковой природы, а с опорой на формальный статистический критерий, который не поддается (пока?) достаточно надежной содержательной интерпретации (см. ниже «Асимптотическая модель тезауруса»). При этом фундаментальная проблема соотношения семантики и синтаксиса в трактовке значения отдельной единицы словаря, проблема отражения, проекции синтаксических

свойств слова в его семантике осталась на этом этапе нерешенной (она подробно рассматривается в IV главе). Суть ее, на примере кодирования семантической информации при составлении тезауруса русского литературного языка, состоит в следующем. На основе дефиниции слова *поклон*, которая в словаре С. И. Ожегова выглядит так:

1. Наклонение головы или верхней части туловища в знак приветствия, благодарности, покорности и т. п. 2. Приветствие наклонением головы. 3. *перен.* Пожелание благополучия, привет, осуществляется семантическая запись (код) смысла этого слова с помощью компонентов (алгоритм кодирования см. в следующем параграфе) в виде: покло накло голов тул привет благод покор пожела благоп. При этом оказывается, что на уровне компонентов различие между первым и вторым значениями отсутствует, поскольку все компоненты второго представлены уже в первом, а дифференцирующим признаком третьего значения является смысл словосочетания «пожелание благополучия», который при записи его в виде суммы компонентов «размыается» и в ходе поиска и распределения слов по дескрипторам создает информационный шум, относя данное слово к дескрипторам, с которыми оно реально не имеет семантической связи.

Итак, тезаурус является приближенной интерпретацией лексической системы, ее усредненной и агрегированной моделью. Его использование именно в этом качестве, несмотря на неполноту модели, дает исследователю ряд преимуществ. Одно из них, касающееся возможности компактного и обозримого представления системных отношений, мы рассмотрели. Такое представление, кроме того, позволяет наблюдать уровневую организацию лексики и выявлять типы отношений, которые вне тезауруспой схемы остались бы не выражеными явно: имеются в виду ключевые слова и их роль в парцеляции семантических полей (дескрипторных статей) и ликвидации разрывности семантического пространства между дескрипторными блоками. В этом смысле говорят даже о «тезаурусном методе» в семантике, употребляя это понятие как синоним системного изучения лексики и семантики<sup>30</sup>.

---

<sup>29</sup> Там же, с. 45.

<sup>30</sup> Степанов Ю. С. Семантика. — В кн.: Русский язык. Энциклопедия. М., 1979, с. 281.

Вместе с тем как метод тезаурусное изучение семантики допускает «оборачивание», когда уже не тезаурус выступает как отражение системности в лексике, а наоборот, сама системность наводится, предписывается лексике с помощью тезауруса. Так, проводя анализ трех больших групп прилагательных в русском языке — цвета, оценки и пространства, — Ж. П. Соколовская в качестве аксиоматической логико-семантической основы для анализа их смысла использует структурную схему словаря Халлига и Вартбурга: «В соответствии с иерархией „рубрик“ словаря устанавливаем иерархию элементов смысла (или сем): классемы — подклассемы — родосемы — видосемы»<sup>31</sup>. Причем в роли классем используются рубрики первого членения универсума у Халлига и Вартбурга (Вселенная, Человек, Человек и вселенная), в роли подклассем — рубрики второго членения (Небо и небесные тела, Земля, ... и т. д.), в роли родосем — рубрики следующего членения (Небо и небесные тела, Погода и ветры. . . и т. д.), а видосемы вычленяются уже в ходе анализа конкретного словарного материала и тоже образуют иерархию (видосемы первого порядка, второго, третьего и т. д.).

Такой подход подводит нас к пониманию другого преимущества тезаурусной интерпретации лексической системы, связанного с устранением встречающегося еще в ряде исследований предрассудка, будто систему можно изучать несистемно, когда в качестве объекта исследования выбирается некоторый изолированный фрагмент словарного состава (слова, выраждающие понятия «конец», «постоянный», или «помнить», «думать», «дрожать», или глаголы движения, говорения, знания, зрительного восприятия, прилагательные температурного восприятия и мн. под.), выбирается по принципам «замкнутости» и «естественной (легкой) выделимости», и изучаются взаимоотношения составляющих его единиц. При этом результаты подобного анализа не поддаются точной формулировке, их, как правило, трудно бывает использовать другим исследователям, и ценность они представляют только для автора, способствуя расширению его представлений об изучаемом материале. В тех случаях, когда отбор материала и комплектование подобных групп осуществляется с помощью толковых словарей и тезаурусов, авторы вносят в словарные классификации поправки, направленные на дополнительные ограничения состава групп,

и, таким образом, замкнутость выбранных фрагментов практически отождествляется с их закрытостью, никаких выходов за рамки 10—30 отобранных единиц не происходит, что недопустимо при системном подходе. Между тем тезаурусная интерпретация позволяет отчетливо показать, что замкнутость дескрипторных областей, блоков и статей, т. е. системных фрагментов словаря, определяется наличием в них некоторой стабильной структуры я коррелирует не с закрытостью, а наоборот, с открытостью их состава, допустимостью пополнения их списка при сохранении до некоторого момента неизменности структурных отношений внутри блока. С другой стороны, замкнутость означает интенсивность внутренних связей между составляющими данного фрагмента, по сравнению с его внешними связями, без изучения которых опять-таки немыслим системный анализ. Таким образом, тезаурус создает базу для подлинно системного изучения лексики, делая ненужными произвольные группировки слов, вычленяемые на разных основаниях, зачастую не соотносящиеся друг с другом по своей внутренней структуре и, как правило, не имеющие подобия в своем составе.

Однако эти изучения нельзя считать вовсе бесполезными: если они несопоставимы по своим результатам, то известный интерес может представлять их исходный материал. В случае комплектования полного русского тезауруса с помощью ЭВМ было бы полезно собрать, скомпилировать автоматическим способом все подобные группы, о которых сообщалось в литературе, выделенные для разных языков, и, переведя этот материал на русский язык, использовать его для пополнения и проверки состава дескрипторных статей русского тезауруса.

### Тезаурус и лексикография

О том, какое место занимают тезаурусы в словарных типологиях, написано теперь довольно много. Л. В. Щерба, который употреблял термин «тезаурус» в его первом, исконном смысле, — как сокровищница, собрание всех слов языка, а по отношению к рассматриваемым нами словарям пользовался термином «идеологический», —

---

<sup>31</sup> Соколовская Ж. П. Система в лексической семантике. (Анализ семантической структуры слова). Киев, 1979, с. 21.

характеризовал словари этого типа по их свойствам, отличающим их прежде всего от «обычных», т. е. толковых словарей<sup>32</sup>. Классификация Ж. Малкиля строится на трех основаниях (широта охвата материала, перспектива его рассмотрения, способ презентации), и решающим для характеристики идеологического словаря в ней являются значения дифференциальных признаков по второму и третьему основаниям: исторический или синхронический словарь, алфавитный или семантический, дескриптивный или прескриптивный<sup>33</sup>. В одной из недавних типологических классификаций, где термин «тезаурус» трактуется в смысле, близком к его пониманию Щербай, в качестве определяющего признака словаря этого типа выдвигается принцип полной регистрации словарного материала некоторой подсистемы (отрасли, подъязыка, языка в целом) «без всякого отбора»<sup>34</sup>. Тематические и аналогичные словари при таком подходе попадают в другую группу и характеризуются соответственно признаками: «неалфавитный, тематический» и «алфавитный, гнездовой, аналогический».

Некоторые исследователи идут дальше и выдвигают задачу «выделения словарей идеографического типа (тезаурусов) в особую область лексикографии, которую можно было бы назвать тезаурусной лексикографией. В этой области знания должны быть разработаны общие принципы составления как обычных идеографических словарей, так и информационных тезаурусов. В области ИПТ в последнее время сделано значительно больше, нежели в сфере обычных идеографических словарей».

Тезаурусная лексикография должна систематизировать все сведения о созданных к настоящему времени ИПТ, должна предложить их типологию. Таким образом, тезаурусная лексикография имеет широкий круг общих и специальных задач, которые могут оказаться далекими для специалистов по общей лексикографии или по той ее области, в которой разрабатываются принципы составления толковых словарей»<sup>35</sup>.

Безусловно, типология тезаурусов возможна, и один из ее вариантов<sup>36</sup> — на основе внутренней структуры словаря — намечен в первой части данного параграфа, где речь шла об устройстве тезауруса. Но вряд ли будет эффективным при этом отрыв тезаурусов от словарей иных типов и даже какое-то противопоставление их толковым словарям. Как свидетельствует практика составле-

ия и использования тезаурусов (и информационно-поисковых, и общенациональных, т. е. идеографических, словарей), они тесно связаны со всей системой словарей языка, а по отношению к толковым выступают и в качестве объекта и в качестве субъекта лексикографии.

Тезаурусы могут рассматриваться в роли объекта по отношению к толковым словарям в связи с тем, что последние уже на заре системных исследований в области лексики воспринимались как надежный источник выявления постоянных семантических связей и системной обусловленности единиц словарного состава языка: «Указанный метод разложения содержания на элементы необходимо обобщить. Однако для того, чтобы упорядочить все лексические факты в соответствии с описанным принципом, предстоит выполнить чрезвычайно трудоемкую работу. Правда, большая подготовительная работа уже выполнена лексикографией: лексикографические определения одноязычных словарей являются по сути дела первым важным приближением к решению поставленной задачи»<sup>37</sup>. В дальнейшем наряду со скептической оценкой информации, даваемой толковым словарем, и тенденцией к критическому пересмотру принципов словарных дефиниций росла и противоположная тенденция: у лингвистов крепла убежденность в системной ценности целостного восприятия и анализа семантики толкового словаря, в том числе и с отдельными его ошибками, частными промахами, вызванными как бы субъективизмом составителей, но на деле отражающими объективные противоречия

---

<sup>32</sup> Щерба Л. В. Опыт общей теории лексикографии. — В кн.: Щерба Л. В. Языковая система и речевая деятельность. Л., 1974, с. 290—297.

<sup>33</sup> Malkiel J. Distinctive features in lexicography. A typological approach to dictionaries exemplified with Spanish, pt 1. — Romance Philology, 1959, v. XII, N 4.

<sup>34</sup> Цывин А. М. К вопросу о классификации русских словарей. — ВЯ, 1978, № 1, с. 104.

<sup>35</sup> Герд А. С. и др. Информационно-поисковый тезаурус как объект лексикографии. — В кн.: Структурная и прикладная лингвистика, вып. 1. Л., 1978, с 171.

<sup>36</sup> Из других подходов к типологии тезаурусов можно назвать классификацию В. В. Морковкина в его книге «Идеографические словари» или параметрический анализ, предложенный нами в гл. 1 настоящей книги.

<sup>37</sup> Ельмслев Л. Можно ли считать, что значения слов образуют структуру? — В кн.: Новое в лингвистике, вып. 2. М., 1962, с. 135.

между системой, нормой и узусом. Поэтому следующий конструктивный шаг, сделанный в той или иной форме разными исследователями, состоял в приравнивании (с известными оговорками) или в допустимости замены тех «содержательных элементов» (семантических компонентов), о которых говорил Ельмслев и которые задают смысл отдельного слова, — к сумме слов, составляющих его definiciu в толковом словаре<sup>38</sup>. Итогом такого «оборачивания метода», т. е. усмотрения реального смысла в бывших вначале абстрактными конструктами компонентах, стало то положение, что в значительной части современных лексикологических и лексикографических работ, посвященных вычленению, системному или тезаурусному анализу и представлению фрагментов лексического состава языка, толковый словарь бесспорно признается основой для получения объективных критериев семантических классификаций<sup>39</sup>.

По этому пути — от толковых словарей к тезаурусам — смелее и решительнее других двинулись специалисты той области, где тезаурус сразу же стал жизненно настольной потребностью, а именно — терминологии. От эмпирических поисков формальных критериев так называемой условной эквивалентности к осознанию онтологической (системной) обусловленности классов условной эквивалентности и лежащего в их основе понятия «семантическая связь», а затем и к попытке извлечь «строительный материал» для тезаурусов из толковых словарей<sup>40</sup> — таков один из путей развития тезаурусастроения. И поэтому уже вполне обоснованной выглядит такая точка зрения исследователя: «Наш подход к решению вопроса выявления семантической структуры значения термина заключается в следующем: допускается, что семантические связи между терминами можно не задавать, а получать путем семантического анализа текстов определений значений слов, приведенных в словарях толкового типа — русского языка и терминологических»<sup>41</sup>.

Такова, собственно, и наша позиция, которая получила теоретическое обоснование и прошла эмпириическую проверку на материале, изложенном в книге «Общая и русская идеография». Здесь при построении семантических полей (которые в принципе соответствуют дескрипторным статьям тезауруса) решающим критерием для включения слова из исходного списка в поле было наличие **одинаковых слов-определителей в дефинициях этого слова**

й имени поля (дескриптора). Так, в толкованиях всех слов, включенных в поле Развитие, например, мы обязательно найдем один из компонентов, составляющих дефиницию этого дескриптора:

#### РАЗВИТИЕ

подъем	вырасти	расти
прогресс	повышать	расширяться
развитие	повышаться	увеличивать
расширение	поднимать	увеличиваться
распространение	подниматься	укреплять
рост	прибавлять	усиливать
взрастать	развертывать	усиливаться
всходить	развиваться	и т. д. <sup>42</sup>

Однако в свете развивающейся нами концепции между тезаурусом и толковым словарем существует не только односторонняя зависимость, когда первый получается из второго, является его «объектом», но и равнозначная взаимообусловленность. Она предполагает, естественно, что и тезаурус по отношению к толковому словарю может выступать как источник информации о семантической

<sup>38</sup> См., напр.: Караулов Ю. Н. Структура лексико-семантического поля. — НДВШ. Филол. науки, 1972, № 1; Он же. Общая и русская идеография, с. 62—67.

<sup>39</sup> Вот несколько случайно выбранных для иллюстрации этого положения названий из обширнейшей библиографии, которую можно было бы привести в подтверждение его правильности; во всех этих статьях источником выяснения самого состава исследуемых групп служит толковый словарь соответствующего языка: Богданова И. Г. Из опыта исследования глаголов одной лексико-семантической группы в английском языке. — Учен. зап. Азербайджанского пед. ин-та русского языка и литературы им. М. Ф. Ахундова, сер. XII, № 1. Баку, 1975; Грачева П. А. Опыт семантического сопоставления глаголов со значением «дрожать» в современном английском языке. — В кн.: Системное описание лексики германских языков, вып. I. Л., 1976; Дементьева П. В. Из опыта исследования лексико-семантического поля существительных французского языка, обозначающих «конец». — В кн.: Сборник научных трудов МГПИИ им. М. Горького, вып. 100. М., 1976; Седлова Л. И. Исследование лексико-семантической группы прилагательных температурного восприятия (на материале современного французского языка). Автореф. канд.

дис. М., 1976; Щекина И. А. Словарная статья.— объективный индикатор семантики языковых единиц. — В кн.: Вопросы романо-германской и славянской лексикографии. Минск, 1976.

<sup>40</sup> Плотников Б. А. Об использовании лексикографических данных при построении тезауруса. — НТИ, сер. 2, 1975, № 9.

<sup>41</sup> Пищеничная Л. Э. Тезаурус в документальной ИПС. Киев, 1977, с. 5.

<sup>42</sup> Караулов Ю. Н. Общая и русская идеография, с. 335.

структуре слова и системных связях между отдельными структурами, т. е. как своеобразный «субъект созидания». Эта двусторонняя зависимость двух типов словарей друг от друга помогает более определенно ответить и на вопрос: «зачем нужен тезаурус лексикографии?».

В формализованных языках полная эквивалентность тезаурусного и дефинитивного способов задания семантики отдельного элемента была с самого начала введена аксиоматически: «... для того чтобы, опираясь на принятые принципы, мы могли создать язык-посредник, нужно: 1) выбрать необходимый минимальный синтаксис... и 2) выбрать подходящий набор термов. Однако сделать такой выбор не очень просто. Удачный набор термов — это в действительности то же самое, что и набор тем в тезаурусе. В тезаурусе каждое слово определяется последовательностью названий тем, в которые оно входит, в нашем же языке каждая формула определяется множеством термов, которые ее составляют. С математической точки зрения аналогия с тезаурусом является полной»<sup>43</sup>.

Расширению этой идеи в применении к естественным языкам мешали, во-первых, малая распространенность идеографических словарей (т. е. общеязыковых тезаурусов), во-вторых, отсутствие их теории. Однако о том, что между дефиницией слова и классификацией всей лексики существует определенная связь, говорил еще III. Балли: «Рациональная классификация должна основываться на каких-то реальных связях между понятиями. Но что это за связи и каким образом их установить? Как нам кажется, в этом деле лучше всего исходить из определения; мы будем рассматривать его как естественную среду, в которой должно фигурировать определяемое слово или выражение; полное и толковое определение как бы указывает, в какой раздел следует поместить интересующее нас слово. Следовательно, определение уже заключает в себе идею классификации, и наоборот, при наличии готовой классификации определение является незаменимым средством для того, чтобы установить (или исправить) принцип расположения материала»<sup>44</sup>. Правда, и он говорит главным образом об одностороннем движении — от определения к классификации, от дефиниции к тезаурусу, но уже здесь содержится намек на возможность и противоположного движения. В книге «Общая и русская идеография» мы показали, что если в идеографическом

словаре, например Халлига и Вартбурга<sup>45</sup>, рассмотреть последовательно по восходящей иерархии все уровни понятий, в которые включается словосочетание *bateau à vapeur*, то мы получим следующую картину:

Les navires	— суда
Par mer	— (для плавания) по морю
La voie d'eau	— водные пути
Le transport, la circulation	— транспортные средства
L'homme au travail	— человек в труде
L'homme, être social	— человек как общественное существо
L'HOMME	— ЧЕЛОВЕК

Эта ступенчатая структура при ее вертикальном прочтении дает дефиницию соответствующего слова или словосочетания: *bateau à vapeur* — это разновидность судов, предназначенных для передвижения по водным путям, в частности по морю, и представляющих собой транспортное средство для перевозки людей и грузов<sup>46</sup>.

Практическое использование в лексикографии этих свойств тезаурусной классификации для систематизации и упорядочения в толковых словарях дефиниций слов, относящихся к одной лексико-семантической группе, одному полю и шире — к одной теме, только начинается. Н. Ю. Шведова отмечает, что при сопоставлении семантически близких определений выявляется разнобой в толковании словарем С. И. Ожегова (9-е изд.) таких, например, слов, как *завтрак*, *обед*, *полдник*, *ужин*: *завтрак* — утренняя еда; *обед* — 1. прием пищи, обычно в середине дня, 2.. пища, приготовленная для этой еды, 3. время этой еды, обычно полдень, 4. перерыв в работе в середине дня, в обеденное время; *полдник* — легкая еда между обедом и ужином; *ужин* — 1. вечерняя еда, 2. пища, приготовленная для вечерней еды. «Чтобы избежать разнобоя в словаре, — говорит автор, — лексикограф должен работать не с алфавитным списком слов, а с определенными их лексическими группировками внутри отдельных частей

<sup>43</sup> Паркер-Роуд А. Ф. О последних работах в области машинного перевода методом тезауруса и языка-посредника. — В кн.: Математическая лингвистика. М., 1964, с. 234 (разрядка наша. — Ю. К.).

<sup>44</sup> Баллы Ш. Французская стилистика. М., 1961, с. 155.

<sup>45</sup> Hallig Ft. und Wartburg W. von. Begriffssystem als Grundlage für die Lexikographie. 2. Aufl. Berlin, 1963.

<sup>46</sup> Карапулов Ю. Н. Общая и русская идеография, с. 20—21.

речи. Только на этом пути может быть достигнуто единство в разграничении значений, однотипность толкований, непротиворечивые приемы подачи фразеологизмов и примеров, последовательность и единообразие стилистических характеристик»<sup>47</sup>. Конструктивным для толкового словаря, по мнению Н. Ю. Шведовой, было бы разбиение слов внутри одной части речи сначала на тематические классы и подклассы (каждый подкласс по отношению к следующим, нисходящим группировкам сам является классом), а затем — выделение внутри этих подклассов лексико-семантических групп слов (во всех случаях классифицируются не слова, а лексические единицы, т. е. значения). В данной статье Н. Ю. Шведова предлагает собственную классификацию конкретных имен существительных, которая охватывает всю соответствующую лексику однотомного словаря и служит основой для унификации словарных определений. В этой связи можно вообразить и идеальную ситуацию, при которой существовала бы некоторая их универсальная тезаурусная классификация для всей русской лексики, способная в зависимости от глубины словарного наполнения и глубины семантизации каждой единицы порождать любые словари в соответствии с запросом и целями, которым такой словарь должен удовлетворять.

Аналогичным путем идут авторы Исторического словаря английского языка в Канаде (*Historical Dictionary of Canadian English*), подготавливаемого в университете Виктории: для достижения упорядочения и взаимосвязи дефиниций они используют 19 тематических рубрик, по которым распределяется вся обрабатываемая лексика, причем все абстрактные существительные составляют здесь один общий класс.

Приемы использования фрагментов тезаурусной классификации находят применение и в работе над переизданием Большого академического словаря русского языка. Результаты анализа 40 глаголов со значением местонахождения (*находиться, помещаться, пребывать, лежать, стоять, сидеть, висеть, жить* и т. д.) показывают, что некоторые общие свойства их сочетаемости, определяемые их принадлежностью к данной семантической группе, получают отражение в словарных толкованиях<sup>48</sup>.

Рассмотрение положения тезауруса в лексикографии и двунаправленной зависимости между тезаурусом и толковым словарем позволяет теперь вернуться к понятию

замкнутости системы, которого мы коснулись в предыдущем разделе, и более точно его сформулировать. Параметрический анализ словарей показал, что в лексикографии существует несколько способов задания семантемы, семантического содержания слова, причем была установлена не полная, а частичная эквивалентность этих способов друг другу (см. § 1, гл. I), в том числе частичная эквивалентность между дефиницией в толковом словаре и дескрипторной статьей в тезаурусе. Но дескрипторная статья (семантическое поле) в тезаурусе или идеографическом словаре служит «неявным определением» только по отношению к дескрипторам, а не к любому элементу, входящему в это поле и вообще в тезаурус. Что касается каждой единицы словаря, то, как мы видели выше, ее определение в хорошо сконструированном тезаурусе задастся иерархией дескрипторов, и последние, таким образом, выступают в этом случае как семантические множители. Учитывая сказанное, мы можем считать, что строгим показателем замкнутости системы является возможность определить любой ее элемент на основе его положения по отношению к другим, на основе его системных связей, т. е. в случае лексики — на основе тезаурусной формулы. Поэтому словарь, скажем, в 10 тыс. слов, единицы которого распределяются по полутора тысячам дескрипторов так, что ни один из дескрипторов «не забыт», каждый включает в свою статью некоторое число единиц из исходного словаря (см. след. параграф), мы трактуем как замкнутый, но на фоне лексики всего языка, естественно, открытый, допускающий свободное пополнение, которое, тем не менее, не влияет на системные отношения дескрипторов, остающиеся в синхронии стабильными. И наоборот, группу слов со значением «думать» либо, например, выражавших понятие «прекрасного» в определенном языке, мы считаем закрытой (в синхронии, конечно, поскольку в диахронии вообще не существует закрытых классов), по не замкнутой: для независимого описания семантики составляющих ее слов

---

<sup>47</sup> Шведова И. Ю. Однотомный толковый словарь: специфика жанра и некоторые перспективы дальнейшей работы. — В кн.: Русский язык. Проблемы художественной речи. Лексикология и лексикография. М., 1981.

<sup>48</sup> Александрова Л. П. Разграничение значений полисемантических глаголов на основе анализа их сочетаемости. — В кн.: Современная русская лексикография. 1977. Л., 1979.

недостаточно отношений внутри группы, требуется выход за ее пределы, поскольку само базовое понятие («думать» или «прекрасное») определяется вне самой группы, в отношениях с другими понятиями того же уровня. С другой стороны, семантическую систему толкового словаря мы считаем замкнутой, если дефиниции слов, составляющих его словарик, ориентированы на некоторый тезаурус (возможно, идеальный, о котором говорит Н. Ю. Шведова, либо реально существующий, как в идеологическом словаре Х. Касареса или в инструкциях составителям<sup>49</sup>).

Основное содержание этого раздела составило рассмотрение взаимоотношений тезауруса с толковым словарем, и это естественно, поскольку определяющим параметром того и другого является фиксация семантической информации о слове — семантемы, а также информации о семантических связях слов. Но, занимая определенное место в совокупности словарей, осуществляющих лексикографическую параметризацию того или иного языка, тезаурус стимулирующим образом воздействует и на расширение разновидностей словарей других типов. Так возникает идея ономасиологического словообразовательного словаря, в котором вход осуществляется от значения, а на выходе представлены все деривационные элементы и соответствующие процессы данного языка (т. е. концепт — словообразовательный знак). Аналогичным образом может быть предложена идея так называемого «обратного словаря фразеологизмов», в котором пишущий или вообще готовящий некоторый текст человек сможет найти все разнообразие идиомов данного языка, передающих то или иное понятие (концепт — фразеологический знак).

### **Тезаурус как способ систематизированного представления знания**

Первые словари такого рода создавались в истории цивилизации не с целью изучения лексики, естественно, а служили отражением представлений о мире в соответствующую эпоху. Среди таких лексикографических произведений называют древнекитайский (13 канонов), словарь древнегреческого языка Ю. Поллукса и древнеиндийский<sup>50</sup>. Исследовав историю и структуру арабоязычных идеографических словарей, А. Е. Бертельс показал обусловленность принимаемой в том или ином словаре

классификации мировоззренческой (религиозной) приверженностью составителя. «В арабоязычных „Трактатах чистых братьев“ (Х в.), где сделана попытка суммировать все знания того времени о мире и человеке (этую книгу нередко, с нашей точки зрения — ошибочно, называют первой энциклопедией), принято такое общее деление. Часть первая „Трактатов чистых братьев“ „макрокосм“, движение вселенной от единства ее творца к множественности, от бога через сотворенные им степени: первичный интеллект, первичную душу (движущую силу), первичную субстанцию, первичную природу, четыре стихии (огонь, воздух, земля, вода), планеты, звезды и их сферы, вплоть до минералов, растений и животных — уже высокой степени совершенства. Часть вторая книги посвящена „микрокосму“, человеку, венцу творения, „совершенному человеку“, поднявшемуся через самопознание, познание мира и бога до состояния совершенства „малого мира“, человеку, в духовном мире которого происходит через познание и утверждение идеи единого бога возврат от множественности вещей к единственности творца»<sup>51</sup>. Вместе с тем сами ячейки, рубрики (мы бы теперь сказали «дескрипторные статьи») идеографических словарей обнаруживают известную степень общности для разных культур и эпох: «Рассматривая рубрики средневековых арабско-персидских тематических словарей, в частности „Мукаддимат ал-адаб“ Замахшари (ум. 1143) . . . мы встретились с исследованием, указывающим на непосредственную связь одной рубрики словаря Замахшари с рубрикой словаря среднеперсидского языка „Фарханг и пахлавик“ и зороастрской „Книгой бытия“ — „Бундахиши“. Речь идет, условно говоря, о „фруктах“, или „орехах и фруктах“. Они делятся в указанных источниках на те, у кото-

---

<sup>49</sup> «При формировании дескрипторных блоков важно соблюдать правило сочетаемости блокообразующих понятий. Правило сочетаемости требует, чтобы блокообразующие понятия не были взаимоисключающими, а те признаки, которые входят в блок, напротив, были бы взаимоисключающими. При соблюдении этого правила для всестороннего описания предмета достаточно взять по одному дескриптору из каждого блока» (Соколов А. В. Указ. соч., с. 50).

<sup>50</sup> Классификационные схемы двух последних приведены в кн.: Морковкин В. В. Указ. соч., с. 13—15.

<sup>51</sup> Бертельс А. Е. О происхождении рубрик тематических (идеографических) словарей. — В кн.: Переводная и учебная лексикография. Русский язык. М., 1979, с. 23.

рых едят только наружное (вишни, персики), и те, у которых едят то, что внутри (арбузы, грецкие орехи и т. п.). В словаре Замахшари слова арабского и персидского языков расположены по тематическим главам, причем внешне, по методике это расположение не отличается от, скажем, словаря Ф. Д. Дорнзейфа — только что словарь Замахшари двуязычный<sup>52</sup>. Истоки этой общности автор склонен видеть в исторической преемственности и зависимости средневековых европейских классификаций от арабских и персидских, которые в свою очередь восходят к аккадской системе письменности, где перед словом ставились детерминативы, обозначавшие его категорию — древнее общее понятие, абстракцию.

Мысль о связи и д е о г р а ф и и в нашем ее понимании, — как теоретической основы систематических классификаций в лексике и построения словарей соответствующего типа, — с и д е о г р а ф и ч е с к и м письмом (Китай, Вавилон, Египет) представляется правдоподобной и, во всяком случае, заслуживающей дальнейшего изучения. Тем более, что на новом витке спирали, по которой развивается человеческое знание, мы опять приходим к своеобразной «иероглифической» фиксации общих и частных понятий, дающей возможность сжатия информации и экономного выражения любой мысли: речь идет о разного рода классификационных языках пауки. Например, в известной классификации Ш. Р. Рапганатана (colon classification), где знак двоеточия, показывающий «деление по проблемам», вошел в само название системы («классификация посредством двоеточия»), выражение (тема) «астрономия в трагедиях Шекспира» передается записью — O2J64 : 90B9, где «О (большая латинская буква) — литература (основной класс), 02 — трагедия (из основной таблицы), J64 — 1564 год, т. е. год рождения Шекспира (по таблице хронологических делений), 9 — литературная критика (одно из «проблемных» делений в разделе О — литература), 0 (нуль) — знак склонности, B9 — астрономия (по основной таблице)<sup>53</sup>. Классификация Рапганатана в своих истоках преследовала библиографические цели: путем фасетного анализа, т. е. установления принадлежности понятий, входящих в название книги, к определенным категориям, областям знания, формировать шифровую запись этого названия, которая состоит из указания фасетов, подфасетов, блоков, подблоков и т. д., полностью задавая объем и содержание соответствующего

предмета исследования и как бы получая тем самым своего рода «дефиницию»<sup>54</sup>. Как видим, аналогия с тезаурусом довольно сильная.

Тезаурусный способ систематизации знания наиболее полное воплощение нашел в отраслевых тезаурусах, предназначенных для использования в специфических информационно-поисковых системах. В этом случае систематизация знания равнозначна систематизации терминологии в данной отрасли. Некоторые примеры таких тезаурусов, общее число которых по разным отраслям и для разных языков к настоящему моменту составляет около тысячи, приводятся в начале следующего параграфа, в связи с обсуждением вопроса о месте и роли общеязыкового тезауруса по отношению к отраслевым. В своем типологическом делении последние повторяют охарактеризованные выше по четырем измерениям (входам) разновидности общеязыковых. Так, из двух опубликованных почти одновременно русских тезаурусов по языкознанию один организован по собственно идеографическому принципу, но имеет единственный вход, поскольку дескрипторная его часть (концепт—знак) совмещена с алфавитной (знак—концепт)<sup>55</sup>, тогда как другой представляет собой аналогическую разновидность тезауруса<sup>56</sup>, где строго разработана структура словарной статьи и система перекрестных отсылок, позволяющие фиксировать тонкие взаимоотношения между терминами, но зато невозможно в целостном виде воспроизвести структуру того или иного раздела науки о языке. Например, дескрипторная статья «Семантика»— в первом тезаурусе выглядит следующим образом (с — синонимы, в — вышележащие понятия, н — нижележащие понятия, а — ассоциаты):

с Семасиология	семантический анализ
Учение о значении	Автоматический
в Семиотика	семантический синтез
Языковые уровни	Антонимия
Языкознание	Гипонимия
п Автоматический	Звуковой символизм

<sup>52</sup> Там же, с. 25—26.

<sup>53</sup> Денисов П. Н. Принципы моделирования языка. М., 1965, с. 58.

<sup>54</sup> Ранганаатан Ш. Р. Классификация двоеточием. Основная классификация. М., 1970.

<sup>55</sup> Тезаурус информационно-поисковый по языкознанию. М., 1977.

<sup>56</sup> Никитина С. Е. Тезаурус по теоретической и прикладной лингвистике. М., 1978.

Знаковая функция языка	категории
Значение	Семантического
Концептуальное поле	дифференциала метод
Лексико-семантическое поле	Семантическое
Модель «смысл—текст»	словообразование
Однозначность	Семема
Омоимия	Синонимия
Ономасиология	Смысл
Пароним	Лексикология
Полисемия	Логическая семантика
Понятийные категории	Философия языка
Порождающая семантика	Фразеология
Семантико-лексические	Язык и мышление

Одновременно любой из терминов, составляющих эту дескрипторную статью, мы найдем на своем месте по алфавиту в общем списке входных единиц тезауруса<sup>57</sup>. Глубина такого способа систематизации знания, чтобы быть эффективной, не должна быть слишком большой, а в идеале должна коррелировать с «правилом шести шагов». В рассмотренном тезаурусе в соответствии с использованным при составлении предметным рубрикатором по языкоznанию максимальная глубина равна шести, т. е. классификация содержит 7 уровней (7 терминов, 6 точек): 89. Языкоznание, 89.1 Общее языкоznание, 89.1.5 Стой языка, 89.1.5.41 Уровни языка. Язык как система, 89.1.5.41.6 Лексикология, 89.1.5.41.6.4 Ономастика, 89.1.5.41.6.4.3 Топонимика.

В принципе глубина иерархического соотношения понятий в любой систематизирующей классификации не должна зависеть от объема интерпретируемых фактов, знаний, а должна определяться возможностями воспринимающего и оперирующего этими данными субъекта. Так, в тезаурусе, систематизирующем терминологию не всего языкоznания, а только его разделов, связанных с автоматической обработкой текста, максимальная глубина классификации тоже равна шести (7 терминов, 6 черточек): «формальный класс — часть речи — глагол — время — прошедшее время — аорист — сигматический аорист»<sup>58</sup>. В то же время в «Новой Британской энциклопедии», представляющей собой попытку систематизировать все человеческое знание (15-е изд.)<sup>59</sup> и по объему анализируемых и классифицируемых сведений, казалось бы, несравнимую с тезаурусом по автоматической обработке текста, мы найдем ту же самую глубину. Опыт «Новой Британской энциклопедии»<sup>60</sup>, составленной по тезаурусному принципу, демонстрирует все преиму-

шества такой систематизации материала. Поскольку использование четырех тезаурусных входов облегчает доступ к материалу, а оптимальная глубина классификации, соизмеримая с возможностями человеческой памяти, позволяет пользователю сравнительно легко ориентироваться во всем многообразии знаний, накопленных человечеством.

Структура «Новой Британской энциклопедии» такова. В соответствии с четырьмя охарактеризованными нами размерностями семантического пространства она разделена на серии, но поскольку алфавитный вход здесь совмещается с пермутационным, то таких серий только три, а не четыре. Первая, состоящая из одного тома и названная *Propaedia*, воплощает соотношение концепт—концепт, т. с. содержит классификационную схему универсума и представляет собой «путеводитель по Британике». Вторая серия (т. I—X), названная *Micropaedia*, выполняет справочную функцию и организована по принципу знак—концепт (и знак—знак), т. е. является обычным энциклопедическим словарем с алфавитно-пермутационным входом и дефиницией на выходе. Наконец, третья, систематическая серия, предназначенная для образовательных целей (т. I—XIX), носит название *Macropaedia* и в 4207 больших статьях (*longer articles*) содержит развернутую характеристику основных понятий («дескрипторов»), охватывающих все сферы знания и человеческих интересов. Вход в эту часть осуществляется, таким образом, от понятия (концепт—знак), но взамен набора терминов и ключевых слов обычного тезауруса читатель находит в дескрипторной статье текст, описание соответ-

---

<sup>57</sup> Из состава приведенной статьи, очевидно, по случайному недосмотру, в число входных единиц данного тезауруса не попал термин «логическая семантика».

<sup>58</sup> Никитина С. Е. Указ. соч., с. 56. — Приводя этот пример, автор тезауруса почему-то интерпретирует глубину такой классификации как восемь шагов. Может быть, при этом имеется в виду, что исходный термин в цепочке («формальный класс») сам является результатом некоторого (первого) шага деления? Но тогда все равно семь шагов, а не восемь.

<sup>59</sup> The new encyclopaedia Britannica. In 30 volumes. 15-th ed. Chicago, London, etc., 1975.

<sup>60</sup> Она не является, конечно, ни первым, ни тем более единственным опытом такого рода: начиная с середины нашего века в Европе было осуществлено несколько аналогичных изданий, напр.: *Encyclopédie universelle*, t. 1—8, Verviers, Gérard, 1961—1963; *Der neue Herder*. (In 6 Bd.) Freiburg u. a., 1966—1968; и др.

ствующей области знания, включающее все необходимые термины и перекрестные отсылки. Что касается устройства классификационной схемы, то она конструировалась следующим образом. В самой этимологии слова en-cyclopaedia авторы<sup>61</sup> усматривают знак того, что знание представляет собой круг. Этот круг человеческих знаний делится на 10 частей таким образом, что в его центр помещается многоаспектный перечень направлений человеческой активности и интересов в области знания, в том числе и научных направлений. Следующий шаг в членении — 42 подразделения (римские цифры с независимой нумерацией по каждой части), которые в свою очередь делятся на 189 секций (большие арабские цифры), а внутри последних — свои части (заглавные буквы латинского алфавита с самостоятельной последовательностью в каждой части). Эти части делятся далее на разделы (малая арабская

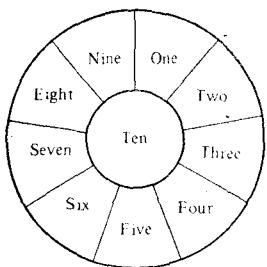


Рис. 2

нумерация), в которых вычленяются подразделы (строчные латинские буквы), а внутри последних — тематические зоны (снова римская нумерация, но малыми знаками). Первое членение включает такие части:

Часть один	Часть шесть
Материя и энергия	Искусство
Часть два	Часть семь
Земля	Технология
Часть три	Часть восемь
Жизнь на земле	Религия
Часть четыре	Часть девять
Человеческая жизнь	История человечества
Часть пять	Часть десять
Человеческое общество	Отрасли знания

В этой иерархии язык (предмет языкоznания) имеет шифр 442 (часть четыре, подразделение IV, секция 2), а семантика на фоне общей организации данной секции (прописные латинские буквы) представлена следующей схемой<sup>62</sup>:

А — Природа и функции языка  
Б — Структура речи и языка

4. Принципы семантики: философское и научное изучение значения

- a) философские концепции значений<sup>61</sup>: значение в его отношении с референтом, истиной и мыслью
- в) значение в лингвистике
  - I. семантика в лингвистической теории: изучение обозначения
  - II. семантика в преподавании языков (значение, структура, коннотация)
  - III. лексикография (значение как форма и адекватность словарных единиц)
  - IV. генеративная семантика: отношения между синтаксическими и семантическими компонентами

С — Письменные языки: системы нотации

D — Лингвистика

E — Классификация языков.

Таким образом, если нас интересует проблема соотношения семантики и синтаксиса, то сведения о ней мы найдем в иерархической классификации знаний на максимальной для данного тезауруса глубине 7 шагов: 4-4-2-В-4-В-IV.

Тезаурусный способ систематизации знания в силу своих преимуществ получает все большее распространение в разных науках, и предложение академика М. П. Алексеева строить Пушкинскую энциклопедию по тезаурусному принципу представляется весьма перспективным. Такая энциклопедия могла бы состоять из трех частей (трех томов) — пропедии, содержащей генеральную схему и указатель; микропедии, словник для которой уже подготовлен; макропедии, представляющей собой систематическое изложение результатов изучения творчества поэта. Самое сложное в этом случае — разработать классификационную схему, в которой конечными точками должны быть названия «дескрипторных» статей в макропедической части. Источниками для установления ключевых понятий, «концептов» пушкинианы могут служить монографические исследования, а также уже изданные Лермонтовская и Шевченковская энциклопедии (собственно «энциклопедические словари»), объем статей в которых служит косвенным показателем значимости (т. е. способности выступать в роли дескриптора) соответствующих понятий.

<sup>61</sup> В качестве инициатора, основного автора и руководителя разработки классификационной схемы назван председатель издательского комитета М. Дж. Адлер (Mortimer J. Adler) — Propaedia, р. X.

<sup>62</sup> При этом надо помнить, что за каждой большой латинской буквой скрывается разветвленная классификация соответствующей части с глубиной в три шага, подобная приведенной для «семантики».

## Тезаурус и лексикон человека

Тот факт, что тезаурус используется как адекватный способ систематизации знания, очевидно, не случаен. Как отмечает Г. А. Брутян, «знание имеет языковой характер не только в том смысле, что язык выступает как способ осуществления знания, но и в том, что язык оставляет свой специфический след па знании»<sup>63</sup>. Именно в этом ключе в «Общей и русской идеографии» мы развили мысль об известном отражении в идеографическом словаре (тезаурусе) «картины мира», «языковой модели мира». Содержание самого термина «тезаурус» эволюционировало в современном научном употреблении соответствующим образом, и теперь мы встречаем трактовку этого термина в расширительном понимании, которое, на наш взгляд, достаточно верно отражает существо дела и свидетельствует о развитии у слова «тезаурус» нового, третьего по счету значения<sup>64</sup>: «В словаре русского языка тезаурусу более всего соответствует одно из значений слова мир: например, мир подростка, мир робота»<sup>65</sup>. Раскрывая такое представление о тезаурусе, автор пишет далее: «Стратегия общения определяется тезаурусами людей, которые по принципу разноспектного сходства входят в разные социально-демографические группы и в своей совокупности образуют тезаурус человечества как глобальной информационной системы. . . В процессе жизнедеятельности динамической системы ее тезаурус может расти и сокращаться в объеме, а также перемещаться в информационном пространстве. . . Вырожденность тезауруса проявляется в таких личностных свойствах, как нелюбознательность, рутинность, детерминистичность мышления, педантизм, плохая приспособляемость к меняющимся условиям, ксенофобия, скептицизм, мнительность, повышенный самоконтроль. И тогда информативными чаще считаются сообщения, подтверждающие известные идеи, а не опровергающие их»<sup>66</sup>. Таким образом, отсюда уже один шаг остается до утверждения, что тезаурус-словарь если и не моделирует лексикон человека (т. е. тезаурус в третьем значении), то в сильной степени коррелирует с ним. Структура лексикона на базе трех родов данных: современных представлений о лексико-семантической системе, известных ныне закономерностей функционирования мозга и механизмов памяти, собственных экспериментальных исследований — пред-

ставлена А. А. Залевской. Вывод автора таков, что лексикон устроен иерархически, причем единицы высшего яруса по отношению к предыдущему низшему обладают интегративным характером; что принципы «вертикальной» и «горизонтальной» упорядоченности элементов взаимодействуют друг с другом, т. е. элементы каждого яруса «включаются» в линейные связи разной протяженности, обеспечивая тем самым контакты между различными иерархиями. Представление о лексиконе как многоярусной системе объемных многократно пересекающихся многопараметрных полей согласуется с топографической гипотезой кодирования и считывания информации»<sup>67</sup>. Такая картина, во всяком случае в вербальной ее части, довольно близка к общему представлению о тезаурусе как словаре; но, не поддаваясь соблазну установления слишком легкой поверхностной аналогии, попробуем проследить более основательное сходство словаря-тезауруса и лексикона человека в особенностях функционирования каждого из них в одинаковых ситуациях и в их отношении к психофизиологическим механизмам хранения информации в мозгу человека.

Из серии экспериментов, на которые опиралась в своих выводах А. А. Залевская: свободный ассоциативный, направленный ассоциативный, идентификации, поиска слова, — для нас наиболее интересен последний. После проведения одной из разновидностей эксперимента информантов просили записать по памяти предъявлявшиеся в ходе эксперимента слова, причем предварительной установки на их запоминание не было. Анализ случаев ошибочной записи слов убедительно показывает, что восприя-

<sup>63</sup> Грутяг А. Языковая картина мира и ее роль в познании. — В кн.: Методологические проблемы анализа языка. Ереван, 1976, с. 57—58.

<sup>64</sup> Первое значение, самое раннее, определяется древнегреческим этимоном этого слова θηράυρος 'хранилище, сокровищница', и именно в таком смысле употреблял его Щерба; второе развилось и распространилось, по-видимому, под влиянием тезауруса Роже и подражаний ему для обозначения способа организации материала в этом «хранилище»; и наконец, третье возникло как положительная оценка такого способа хранения, позволяющего экономным образом моделировать «мир».

<sup>65</sup> Воробьев Г. Г. Теория тезаурусов в анализе коммуникаций. — В кн.: Семиотика и информатика, вып. 11. М., 1979, с. 3.

<sup>66</sup> Там же, с. 4, 33.

<sup>67</sup> Залевская А. А. Проблемы организации внутреннего лексикона человека. Калинин, 1977, с. 73.

тие предъявлявшихся единиц и их запоминание шло в рамках соответствующих семантически связанных групп (полей):

Исходные слова	Ошибочно записанные слова	Исходные слова	Ошибочно записанные слова
ОРЕЛ	сокол	МАМА	папа
КРОЛИК	заяц	МАТЬ,	родители
СОБАКА	пес	БРАТ	сестра
ЛОШАДЬ	конь	СЫН	дочь
ЛЮДИ	народ	ДЕВОЧКА, ЖЕНЩИНА	девушка
ПУТЬ	дорога	ЖИВОЙ, ЖИЗНЬ	жить
ГОРЯЧИЙ	жаркий, теплый	СИДЕТЬ, СТОЯТЬ	лежать
КВАДРАТНЫЙ	четырехугольный	ПОЛУЧАТЬ	посылать
ГОВОРИТЬ	разговаривать		
УЧИТЬСЯ	изучать		и т. д.

Подводя итоги анализу приемов поиска информантами слов в памяти, автор пишет:

«...3) В условиях оперирования родным языком поиск слов в памяти шел по линии с м ѿ с л о в ѿ с с в я з ѿ м между словами, в то время как подобие формы слов оказалось лишь сопутствующим признаком.

4) Поиск слов в памяти шел по ряду смысловых признаков, перекрещивающихся между собой и служащих «мостиками» для объединения слов в разного рода группы.

5) Объединение слов по смыслу отражает привычные для информанта связи между объектами (т. е. его з на ие о мире)»<sup>68</sup>.

В принципе эти тезисы А. А. Залевской не требуют комментария, поскольку из предыдущего текста об устройстве тезауруса ясно, что, будучи сформулированными по отношению к лексикуону человека, они с равным успехом приложимы и к характеристике организации лексического материала в словаре-тезаурусе.

Это сходство прослежено нами в направлении «от лексикона — к тезаурусу». Для продолжения параллели между ними рассмотрим некоторые разновидности тезаурусов, структура которых как бы напоминает слепок со способа хранения лексической информации в сознании, эксплицируя многопараметровость смысловых связей и «обнажая» пути поиска слова в памяти, т. е. взглянем на их отношения «от тезауруса — к лексикуону». Речь идет

о специальных словарях, предназначенных для разного рода словесных игр и состязаний, типа «Les mots croisés» или «Competition Dictionary». Прежде всего такой словарь, сохраняя четыре основных тезаурусных входа, охарактеризованных выше, позволяет резко увеличить число точек доступа к лексике за счет широкого варьирования подходов в каждое из четырех измерений семантического пространства. Принципиальные возможности такого варьирования определяются следующими схемами, на которых стрелкам соответствуют подходы в словарь:

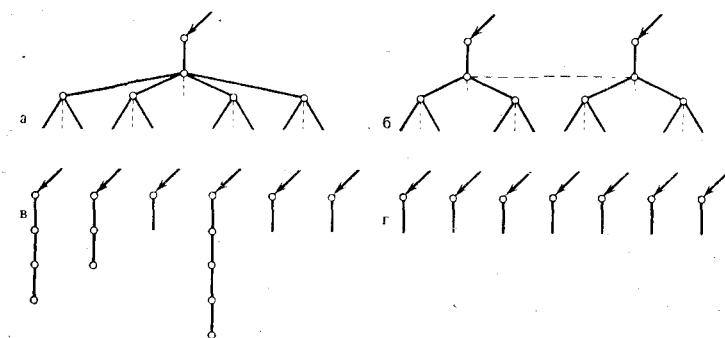


Рис. 3

где а) — моноиерархия, б) — полииерархия, в) — неполная иерархия, г) — отсутствие иерархии<sup>69</sup>. Сложная комбинация этих способов образует гибкую систему поиска слова. Моноиерархия обычно представлена дескрипторной (синоптической) схемой тезауруса, полииерархия задается разными приемами. Например, в «Competition Dictionary»<sup>70</sup> это, во-первых, разбиение универсума на 177 областей, которые неравномерно распределяются по лексикографическим параметрам: ареальному, тематическому, стилистическому, фразеологическому, грам-

<sup>68</sup> Залевская Л. А. Указ. соч., с. 49.

<sup>69</sup> Курбаков К. И. Кодирование и поиск информации в автоматическом словаре. М., 1968, с. 134.

<sup>70</sup> The indispensable competition dictionary. The Times of India press, Bombay, s. a.

матическому, ономастическому и группе системных параметров (синонимия, антонимия, омонимия, паронимия); во-вторых, сами эти области упорядочены и по алфавиту, и по своему положению друг относительно друга (т. е. по-странично в оглавлении); в-третьих, внутри областей принципом создания иерархии является число букв в слове (группы трехбуквенных, четырехбуквенных и т. д. слов); в-четвертых, внутри групп слов одинаковой длины в качестве основы иерархии опять используется алфавит. В этой конструкции первичными входами в словарь практически можно считать все 177 расположенных по алфавиту рубрик, хотя не за каждой из них стоит реальная группа слов, потому что около половины из этих 177 «дескрипторов» оказываются «ключевыми словами», служат для дополнительных перекрестных отсылок, как бы создавая ориентировочные вехи на пути поиска слова и усиливая системные связи внутри одной и той же группы параметров (тематических, системных, ономастических и т. д.): Sculptore (See «Artists»), Drinks (Alcoholic) (See «Wines»), Dyes (See «Colours»), Paintes (See «Colours»), Animals (See also «Birds», «Dogs», «Fish», «Insects», «Mollusks», «Reptiles» and «Worms»), Officials (See also «People», «Religion», «Military», «Nautical», and «Professional» — Including Ranks, Titles, and Religious Designations); Magazines (See «Principal Dayly Newspapers in Great Britain») и т. д. Кроме того, сами названия периодических изданий (Magazines, Periodicals, Newspapers), например, помимо уже приведенного, к которому дана отсылка от одного из общих имен, делятся внутри более подробно по тематическому, территориальному, времененному, географическому признакам: Periodical papers and reviews, Weekly papers religious, Professional and trade (periodical) papers, London offices of colonial and foreign Newspapers. Среди 98 «дескрипторов» этого тезауруса, которые имеют реальное наполнение словами, можно выделить три больших раздела.

Первый из них составляют тематические группы, общее число которых здесь — 40 (Cereals, Clothes, Flowers, Food, Fruit. . .). Следующий большой раздел — 42 ономастические группы, среди которых: имена известных лиц (Scientists, Authors, Musical Composers, Generals, Field-Marshals, Air-Marshals, etc., Explorers, Film-Stars. . .); марки товаров и названия кораблей, клубов,

колледжей, газет и журналов, фильмов (Motor-Cars, Film Titles, Beer, Some Public Schools (Boys). . .); личные имена; клички животных (породистых собак, скаковых лошадей), разделенные по географическому признаку; государственные и административно-политические названия (американские и индийские штаты, британские провинции, английские и ирландские графства, канадские провинции и т. п.); историко-географические названия (Battles-Naval and Military including Sieges); этнонимы. Наконец, третий раздел, включающий оставшиеся 16 рубрик, не имеет однозначно определяемого характера, и его удобно поэтому назвать «прочие». Сюда относятся четыре рубрики с системными параметрами (синонимы, антонимы, омонимы, паронимы); пять групп параметров, которые с некоторой долей условности можно назвать грамматическими: Common endings (т. е. наиболее распространенные вторые части сложных слов — наименование городов, железнодорожных станций и т. п. — Castle, Field, Stone, — опять-таки упорядоченные внутри по количеству букв и по алфавиту), произносительный параметр, префиксы (главным образом латинские и греческие), рифмующиеся слова, суффиксы; далее в этом разделе представлены ареальный (An Indian glossary), стилистический (Slang & Colloquialisms) и фразеологический (Trite phrases) параметры в виде словарей с алфавитным входом. Четыре последние рубрики в этом разделе содержат специальную «кроссвордную» информацию, т. е. такую формулировку вопросов по поводу слов языка и такие ответы на них, которые встречаются практически только в кроссвордах или в ассоциативных экспериментах.

В завершающей части «Competition Dictionary», как и положено в полном тезаурусе, помещен обычный толковый словарь, с той лишь разницей, что в данном случае он не включает с обязательностью всех тех слов, которые вошли в охарактеризованные рубрики (исключаются, естественно, ономастические, грамматические, стилистический, ареальный и некоторые другие разряды).

Поиск слова в таком словаре осуществляется одновременно по нескольким параметрам, например принадлежности к тематической сфере, числу букв и месту в алфавитном ряду или синонимическим отношениям с исходным словом, общему числу букв и некоторым уже известным буквам в нем и т. п. Таким образом, и в самих приемах поиска, и в его многопараметровости, или лучше ска-

зать — дивергентности, наблюдается известная перекличка, если не совпадение, с картиной функционирования лексикона человека в аналогичной ситуации. Тем, кто занимался лингвистическими основами словесных игр<sup>71</sup>, известно, что в играх с неполной информацией типа кроссворда, моделирующих некоторый творческий процесс, комплекс параметров включается одновременно и поиск происходит в скрытой форме, выходя на уровень явного перебора только в критических или тупиковых ситуациях. Интеллектуализация перебора является привилегией «нелевополушарного» мышления и основывается, в частности, на механизме «прыжка», механизме «и так далее», механизме «назначения вероятностей»<sup>72</sup>. Как известно, специализация работы полушарий мозга заключается в том, что левое контролирует логические, линейные и вербальные функции, а правое — эмоциональные, пространственные и интуитивные. Вообще деятельность левого полушария связана всегда с последовательным дискретным перебором, тогда как правого — с глобальной обработкой информации: «...по главным своим функциям правое недоминантное полушарие — это полушарие целостных («топологически связанных») единиц. Поэтому оно оперирует целостными зрительными и пространственными образами, предметами, иероглифами, жестами, музыкальными мелодиями и ритуализованными фразами и именами вещей, не членящимися на единицы («буквы») в самом этом полушарии. Но каждому целостному образу правого полушария может соответствовать его представление в виде последовательности дискретных символов в левом полушарии»<sup>73</sup>. Отсутствие надежных экспериментальных данных вынуждает нас остановить рассмотрение аналогии между устройством тезауруса, организацией внутреннего лексикона человека и механизмами мышления на этом утверждении, оставляющем читателя перед интересными и чреватыми далеко идущими выводами размышлений. Хотя в принципе возможны и более сильные предположения, не только переносящие тезаурус в сознание, но и локализующие его в мозгу человека: «...конкретная смысловая информация о внешнем мире, содержащаяся в толковых словарях естественного языка (и в аналогичных «тезаурусах» информационных машин), хранится и обрабатывается в правом полушарии»<sup>74</sup>.

## Тезаурус в процессах речепорождения и речевосприятия

Х. Касарес в предисловии к своему словарю, обосновывая его полезность и нужность, нарисовал такую картину: «Писатель остановился с застывшим в воздухе пером в надежде найти подходящее определение к некоему персонажу, который предстает в изношенном и покрытом пятнами платье. В какой части должен наш писатель открыть идеологический словарь, чтобы тот пришел ему на помощь? Пока еще ни в какой. Этот персонаж, о котором идет речь, вызывает в сознании прежде всего представления о *неряшлиности* и *грязности*; кроме того, он напаляет па мысль о *бедности*, если не выглядит лучше потому, что не имеет возможности, или о *скрупости*, если, напротив, сознательно носит плохую одежду, чтобы сэкономить то, что мог бы беззаботно потратить.

После такого краткого анализа, который при известной практике осуществляется автоматически и при котором, кстати сказать, актуализируется все, о чем бы пишущий хотел нам поведать, — наступает момент открыть идеологический словарь. Допустим, наш писатель решил, например, что характеристики, которые его интересуют, должны выявить *скрупость* и *грязь* персонажа. Остальное проще простого. Посмотрев соответствующие группы, он увидит, что в каждой из них предложены целые серии определений с различными оттенками, и найдет вдобавок,

---

<sup>71</sup> См., напр.: *Караулов Ю. П.* Семантические игры как прием обучения языку. — Русский язык в нац. школе, 1974, № 3.

<sup>72</sup> «Механизм «прыжка» . . . Содержательные производные правила чаще возникают при нацеленном «прыжке» к желательному выводимому объекту с последующим восстановлением «вставки» . . .

«Механизм «и так далее». Когда речь идет о порождении собственно допустимых правил, неизбежным и самым нетривиальным этапом является обнаружение того, что данные типы «вставок» укладываются в одну схему, что данный объект может служить метапеременной с такой-то областью задания. Обнаружение закономерности тоже является «прыжком» от наличного материала и еще более фундаментально противостоит последовательному перебору вариантов» (*Маслов С. Ю.* Теория поиска вывода и вопросы психологии творчества. — В кн.: Семиотика и информатика, вып. 13. М., 1979, с. 41).

<sup>73</sup> *Иванов В. В.* Чет и нечет: асимметрия мозга и знаковых систем. М., 1978, с. 73.

<sup>74</sup> Там же, с. 31.

что обе серии Пересекаются в слове *sórdido* (рус. ‘невероятно грязный, мерзкий и скаредный’ — Ю. К.), которое обозначает одновременно и грязного и жадного»<sup>75</sup>.

Эта иллюстрация как нельзя более просто и вместе с тем точно характеризует основную функцию тезауруса — быть инструментом для выбора слова, однозначным образом соответствующего мысли пишущего или говорящего, эксплуатируя вход концепт—знак. В этой практической функции тезаурусы естественных языков используются и художниками слова, и публицистами, находят применение в практике обучения языку, что неоднократно подчеркивалось в лингвистической литературе. Вырожденным случаем такого использования было бы развертывание в текст самой статьи тезауруса, поскольку получившийся текст отличался бы предельно возможной связностью (изотопией), составляющие его высказывания были бы максимально приближены к истинным (общизвестным, тривиальным), а значит, весь этот псевдотекст обладал бы минимальной информативностью, представляя бы собой то, что называют «информационной руминацией». На этом пути, развертывая, например, в текст дескрипторную статью «Музыка» (скажем, из словаря Х. Касареса), пишущий никогда не сможет «породить» такие сочетания, как *музыка гибели, благоденствие музыки, упадок музыки, греховая музыка, неистовая музыка, гегемония музыки* (Г. Гессе, Игра в бисер). Подобные сочетания возникают как раз на пересечении разных семантических полей, разных статей тезауруса. Всякий текст, если только он несет новую информацию, находится в оппозиции — скрытой или явной — со средним, стандартным (социально устойчивым) тезаурусом (здесь «тезаурус» в значении «мир» данной речевой общности).

Такое противопоставление могло бы показаться парадоксальным, поскольку, будучи источником, а точнее — одним из источников текста, тезаурус как будто не должен ему противостоять. Но надо помнить, что коль скоро тезаурус активно участвует в порождении текста, не менее существенна его роль и в восприятии текста или речи. Таким образом, говоря, что текст противостоит тезаурусу, мы сокращенно выражаем ту мысль, что один тезаурус (автора, говорящего), стоящий за некоторым текстом и порождающий его, противостоит другому тезаурусу (читателя, слушающего), на основе которого текст интерпретируется.

В своих рассуждениях мы незаметно переключились с одного понимания тезауруса (как материально существующего словаря) на другое его понимание, не материальное — как «мира», как отраженного и упорядоченного в сознании «лексикона» — в виде некоторой идеальной «картины мира». В каком отношении реальный тезаурус-словарь находится к идеальному тезаурусу-миру? Будем считать, что первый является просто удобной формой представления и изучения второго. Позволив себе метафорическую вольность, изобразим дело так, что тезаурус-мир существует в некотором фантастическом, постоянно флюкутирующем состоянии, но позволяет себя фиксировать, делать как бы мгновенный снимок, запечатлеваящий его в виде тезауруса-словаря. Если зафиксировать таким образом тезаурус в один и тот же момент у всех членов данной языковой общности, или наоборот, в разные моменты у одного и того же индивида (в последнем случае — с достаточным разрывом во времени), то снимки не получатся идентичными: индивидуальные тезаурусы накладываются один на другой с определенным «сдвигом». Тем не менее в них, вероятно, обнаружится общее ядро, инвариант, который и делает возможным взаимопонимание. Расхождения же могут касаться любой размерности семантического пространства тезауруса: и словесного (знако-вого) наполнения семантических (концептуальных) полей, и взаимоотношений (структур) самих концептов, и распределенности знаков по отношению одного к другому внутри полей.

Четыре входа и четыре размерности тезауруса, условно говоря, можно распределить так: там, где на противоположном от входа конце должен быть «знак», где результатом «работы» тезауруса должно быть — ни много ни мало — «приведение истины к выражению» (Ч. Пирс), мы имеем позицию говорящего (пишущего, автора); соответственно, когда итогом «работы» является «концепт», позиция характеризует воспринимающего (слушающего, читателя). Иными словами, размерности «концепт—знак» и «знак—знак» относятся к линии порождения текста, тогда как размерности «знак—концепт» и «концепт—концепт» — к линии его восприятия. Но так как каждый из нас попеременно выступает то в одной, то в другой роли,

---

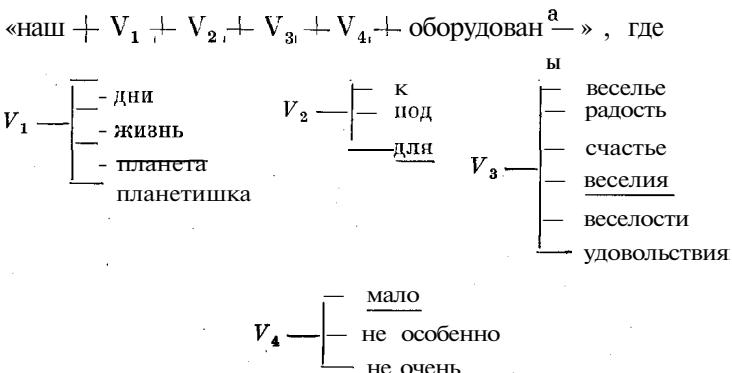
<sup>75</sup> Casares J. Diccionario ideológico de la lengua española. Barcelona, 1975, p. XXII.

предложенное различие не стоит расценивать как абсолютное и имеющее какой-то глубинный смысл, а следует использовать его лишь как прием для более подробной характеристики работы тезауруса.

Итак, позицию автора, ищущего максимально точное, максимально насыщенное, ничем не заменимое выражение (слово) и идущего по линии «концепт—знак», мы с помощью Касареса рассмотрели. Но понятно, что в иных ситуациях задача продуцирующего текст может оказаться прямо противоположной, и он, исходя из каких-то своих целей и используя комбинационные возможности по линии отношения «знак—знак», может опираться на другой принцип — «нет слов незаменимых». Замена обычно происходит либо в пределах одного семантического поля для достижения некоторого стилистического или риторического эффекта, либо с выходом за пределы поля, на пересечении полей — для получения эффекта концептуального, в известном смысле обратного тому, который описан Касаресом. С первым случаем, при котором порождаемый текст сохраняет соответствие эталонному, стандартному тезаурусу, мы имеем дело и в информационно-поисковых системах, где, собственно говоря, презумпция такого соответствия и есть условие работоспособности самой системы, и в обыденном повседневном общении, когда смысл замены не выходит за пределы метасемиотического уровня. Для иллюстрации этого тезиса сравним две следующие цитаты: «Одна из функций тезауруса может быть сведена к понятию неоднозначности между планом выражения и планом содержания в естественном языке: в тезаурусе задаются такие семантические отношения между понятиями (род—вид, синонимия—антонимия, часть—целое и др.), при которых замена одного понятия другим, связанным указанными отношениями, не вызывает качественных изменений в суждении о релевантности (нерелевантности) документа относительно информационного запроса»<sup>76</sup>. И другая: «На ключевое слово *hut* (избушка) один ребенок отвечает *загорелась*, а другой — *маленький бедный дом*. Обе реакции являются предикативными, но первая образует чисто повествовательный контекст, тогда как вторая устанавливает двойную связь с субъектом *hut*: с одной стороны, — позиционное равенство, с другой — семантическую непрерывность (связность). Этот словесный стимул *hut* допускает следующие замещающие реакции: тавтологию *hut*, синонимы *cabin* и *hovel* (из-

бушка), антоним *palace* (дворец) и метафоры *den* (пещера, логовище, берлога) и *burrow* (нора). Случай, когда два слова взаимозаменимы, представляет собой пример позиционного равенства; кроме того, все эти словесные реакции связаны со стимулом семантическим подобием (или соотв.—контрастом). Метонимические реакции на этот словесный стимул, как, напр., *thatch* (соломенная крыша), *litter* (подстилка для скота) или *poverty* (бедность), образуют комбинацию между позиционным равенством и семантической связностью»<sup>77</sup>.

Развивая мысли В. В. Маяковского о творческом процессе, высказанные им в статье «Как делать стихи», И. Левый предложил схему, которая довольно точно, на наш взгляд, отражает эту часть работы тезауруса при рождении текста. И. Левый анализирует все 12 прочтений (вариантов) первой строчки стихотворения о С. Есенине (1. Наши дни к веселью мало оборудованы. . . 12. Для веселия планета наша мало оборудована) и строит следующую рамку:



(слова окончательного варианта подчеркнуты)<sup>78</sup>.

Другой случай, предполагающий введение в действие «отклоняющихся словесных групп» и получение преднамеренного сдвига в тезаурусе, связан с созданием асим-

<sup>76</sup> Пшеничная Л. Э. Тезаурус в документальной ИПС. Киев, 1977, с. 3.

<sup>77</sup> Jakobson R. Der Doppelcharakter der Sprache. Die Polarität zwischen Metaphorik und Metonymik. — In: Literaturwissenschaft und Linguistik, Bd. 1. Frankfurt-am-Main, 1972, S. 328.

<sup>78</sup> Levy I. Generative poetik. — In: Literaturwissenschaft und Linguistik, Bd. 2, T. 2. Frankfurt-am-Main, 1972, S. 558.

метрии между ожидаемым и реализованным в тексте, с конструированием ситуации «обманутого ожидания» читателя (слушающего). На этом пути пересечения и перестановки одного в другое семантических полей и зиждется так называемый «семантический жест» (Я. Мукаржовский), т. е. художественный образ в широком смысле, под которым мы имеем в виду и обычную метафору (*minden fenynél édesebb* — слаше всякого света, *péta iPat* — немой аромат, *űvolt a csend a fülemben* — ревела тишина в моих ушах<sup>79</sup>); и рифму, которую Р. Барт называл «структурным скандалом»<sup>80</sup>, а Р. Якобсон и другие исследователи рассматривали как один из способов организации движения смысла за счет семантического сближения рифмующихся слов (*kind* — *mind*, *dove* — *love*, *light* — *bright*, *place* — *space*, *name* — *fame*)<sup>81</sup>; и «паронимическую атракцию» («В солдатском сердце замер зуммер»)<sup>82</sup>. Однако «семантический жест» или эффект «обманутого ожидания» могут возникать и за счет иных комбинаций по линии «знак—знак», которые не всегда укладываются в лексический тезаурус в разбираемом здесь смысле: «М. Пруст называл Флобера „человеком, который с помощью совершенно нового и индивидуального употребления *passé défini*, *passé indefini*, причастия настоящего времени, определительных местоимений и некоторых предлогов почти так же обновил паши представления о вещах, как Кант своими категориями — теорию познания и реальность внешнего мира"»<sup>83</sup>.

Переходя к анализу размерностей «знак—концепт» и «коинцепт—концепт», характеризующих позицию воспринимающего (слушателя, читателя), мы должны констатировать, что современная наука не располагает достаточно надежными данными (ни экспериментальными, ни теоретическими) о том, каким образом линейная последовательность единиц преобразуется в процессе восприятия и понимания в иерархическую систему зависимостей в мозгу слушающего. Уже по поводу восприятия предложения, например, существуют прямо противоположные точки зрения: согласно одной из них, предложение воспринимается слово за словом, слева направо в один шаг, так что каждый последующий сокращает вероятности прогнозируемых слушающим продолжений; согласно второй, более правдоподобной представляется модель восприятия сверху вниз, согласующаяся с гипотезой глубинных структур. Разнообразие трактовок существенно возрастает при

переходе к вопросу о восприятии текста. Отдавая себе отчет в исключительной сложности проблемы и допуская множественность ее возможных решений, мы и в этом случае попытаемся развивать нашу метафору о «флуктуирующем идеальном тезаурусе», изложенную выше. Причем мы будем опираться на представления о восприятии именно художественного текста, полагая эту ситуацию наиболее сложной и остальные случаи — выводимыми из нее.

Функционирование литературного текста, как известно, не исчерпывается, а только начинается процессом его чтения, ибо чтение не может быть самоцелью. Чтение линейно, синтагматично, а существование и функционирование литературного произведения в читательском сознании систематично (парадигматично). Таким образом, в процессе чтения происходит переход от синтагматики к парадигматике, от горизонтальной оси развертывания текста к вертикальной оси его организации в иерархии семантических, эстетических, идеологических ценностей читательского тезауруса. «В произведении, основанном на вымысле, осуществляется переход, привычность которого заслоняет его значение и исключительность, — от последовательности фраз к целому воображаемому миру. Пере-вернув последнюю страницу „Мадам Бовари“, мы можем сказать, что познакомились со многими персонажами и узнали довольно много об их жизни; а ведь у нас в руках был всего-навсего текст, вытянутый в линейную последовательность. . .»<sup>84</sup>.

Существующие лингвистические способы анализа текста, которые претендуют на то, чтобы реконструировать процесс его восприятия, оперируют как традиционными языковыми единицами (фонема, морфема, слово, троп,

<sup>79</sup> Petöfi J. Bemerkungen zur semantischen Interpretation von sprachlichen Künstenwerken. — In: Literaturwissenschaft und Linguistik, Bd. 3. Frankfurt-am-Main, 1972, S. 135.

<sup>80</sup> Барт Р. Основы семиологии. — В кн.: Структурализм: «за» и «против». М., 1975, с. 156.

<sup>81</sup> Якобсон Р. Лингвистика и поэтика. — В кн.: Структурализм: «за» и «против», с. 216—217. См. также: Симонов П. В. Категории сознания, подсознания и сверхсознания. — В кн.: Бессознательное, т. 2. Тбилиси, 1978.

<sup>82</sup> Григорьев В. П. Поэтика слова. М., 1979, с. 251 и след.

<sup>83</sup> Todorov Tz. Die Beschreibung der Bedeutung in der Literatur. — In: Literaturwissenschaft und Linguistik, Bd. 3. Frankfurt-am-Main, 1972, S. 122.

<sup>84</sup> Тодоров Ц. Поэтика. — В кн.: Структурализм: «за» и «против». М., 1975, с. 62–63.

предложение), так и специально текстовыми общностями и понятиями, как правило, большей чем предложение протяженности — *iiber die Satzgrenzhinaus, beyond the sentence* (высказывание, сверхфразовое единство, абзац, дискурс, текст, тема, прием выразительности, стимул-реакция). В большинстве своем эти способы позволяют выявить некоторую макроструктуру текста, получить своеобразный его реферат, или квазитекст, составленный из упорядоченных списков высказываний, — «текстовую базу»<sup>85</sup>. Однако на этом пути мы не получаем прямого ответа на вопрос, как воспринимается и как хранится текст в памяти. Есть попытки оперировать при анализе понятиями иного порядка, иного характера — незнакомого. Так, Ю. Кристева вслед за русской формальной школой в литературоведении использует понятие «идеологемы»: «Назовем идеологемой ту общую функцию, которая связывает в интертекстуальном пространстве некоторую структуру (скажем, роман) с другими структурами (скажем, дискурсом науки). Идеологему некоторого текста можно определить через его отношения с другими текстами. Итак, мы говорим, что покрытие данной текстуальной организации высказываниями (последовательностями), которые принадлежат к ее пространству или на которые она (эта организация) указывает вне ее пространства (семиотическая работа), называется идеологемой. Идеологема представляет собой интертекстуальную функцию, которую на разных структурных уровнях всякого текста можно «материализовать» в чтении и которая распространяется сверх своей длины, причем она задает исторические и социальные координаты текста»<sup>86</sup>. Но помимо того, что это понятие (даже при условии различения идеологемы текста, идеологемы символа и идеологемы знака) остается фундаментально неясным, автор нигде не показывает его в действии, в работе, не использует его в конкретном анализе текстов, почему оно и остается многозначительным, но туманным намеком, благим, но не конструктивным пожеланием.

Мы будем исходить из того, что за восприятием текста стоит тезаурус читающего (слушающего), и нормально восприятие происходит по каналу «знак—концепт»: знаки текста (слова и словосочетания), рассредоточиваясь и комбинируясь в процессе чтения привычным образом (стандартный тезаурус), формируют некоторый заданный в тексте и тоже относящийся к числу стандартных (реже —

совершенно новый) концепт. Возможности введения новизны исчерпываются либо новым знаковым наполнением (в том числе новым комбинированием знаков на синтагматической оси) старого концепта, либо введением оригинальных концептов и смещением акцентов путем придания статуса концептов знакам, которые в узусе данной речевой общности, в ее эталонном тезаурусе такого статуса не имели; либо, наконец, изменением отношений между существующими в стандартном тезаурусе концептами. Соответственно при восприятии непривычная сортировка знаков образует сдвиг по отношению к читательскому тезаурусу, сигналом которого является «эстетическое недовольство» (Я. Мукаржовский). Если в процессе чтения, при наполнении знаками традиционного концепта, выхода за рамки стандартного тезауруса не происходит, читатель имеет дело с псевдотекстом, несущим минимум информации. И наоборот, если при формировании концептов в процессе чтения не остается никакой корреляции со стандартным тезаурусом, текст вообще не воспринимается. Суггестивное воздействие произведения на читателя, если не учитывать эмоционального момента, заключается в том, что оно может оказывать давление на формирование его тезауруса, приводить к перестройке ужо сложившейся «картины мира».

Рассмотрим приложимость нашей точки зрения к восприятию некоторых идей романа Г. Гессе «Игра в бисер». Сложную структуру романа образуют несколько кардинальных концептов, между которыми при аналитическом чтении обнаруживаются отношения антонимии (отнюдь не стандартной) и взаимодополнительности. Особенность изложения в том, что эти противопоставления — между формирующимися у читателя концептами — в тексте романа нигде явным образом не названы, за исключением символического и всепоглощающего Инь и Ян (темное и светлое, доброе и злое, теплое и холодное, женское и мужское и т. д.). Эта «неназываемость» производит на воспринимающего тем большее впечатление, что он по ходу

---

<sup>85</sup> Dijk T. A. van. Some aspects of text grammars. The Hague, 1972; Kintsch W. The representation of meaning in memory. Hillsdale, 1974.

<sup>86</sup> Kristeva J. Probleme der Textstrukturation. — In: Literaturwissenschaft und Linguistik, Bd. 2, T. 1. Frankfurt-am-Main, 1972, S. 502.

изложения все больше и больше ощущает глубокую мотивированность изображаемого, чувствует справедливость и обоснованность направления, в котором развертываются предложенные автором концепты и отношения между ними, но в то же время не может и не в а ть ситуацию, однозначно соотнести ее с тем, что его повседневно окружает, с общезвестным, со «стандартным тезаурусом». Это, на наш взгляд, и есть художественное открытие, ведущее к суггестии, к последующей перестройке читательского тезауруса.

У каждого человека есть свое представление о таких понятиях, как «судьба», «личность», «история», «иерархия», «музыка», «воспитание» (наставничество), «пробуждение» (*trascendere*), «медитация». Именно эти концепты составляют основной костяк мини-мира, построенного Г. Гессе. Формирование этих концептов у читателя происходит путем извлечения из соответствующих фрагментов текста знаков, характеризующих эти понятия по законам семантического поля — через синонимы, антонимы, родовые, видовые, ассоциативные отношения. Рассмотрим несколько примеров, в которых знаки, раскрывающие данное семантическое поле, задающие читателю содержание соответствующего концепта, выделены курсивом.

### ИЕРАРХИЯ

«. . . именно изгнание индивидуального и возможно более полное включение личности в иерархию Воспитательной коллегии и научного мира есть один из высших принципов нашей духовной жизни. И принцип этот укоренился и стал традицией столь давно, что теперь крайне затруднительно, а зачастую даже невозможно выяснить подробности жизни и черты характера отдельных лиц, имеющих перед этой *иерархией* особые заслуги. . . . иерархическая организация исповедует идеал безымянности. . .» (с. 36). И далее: «порядок . . . наложенная организация . . . не машина, но живое тело. . .» (с. 37).

### МЕДИТАЦИЯ

«. . . Братство паломников в страну Востока, сочлены которого заботились не столько о культивировании интеллекта, сколько о культивировании души, о воспитании благоговения и благочестия — отсюда наша современная форма духовности и Игра в бисер восприняли важные импульсы, особенно это касается приемов контемпляции — созерцания . . . благодаря своей способности, развитой стаинными тайными приемами, магически входить

в минувшие эпохи и состояния духа. . . (с. 48); . . . из духовного инвентаря воспитательной традиции . . . из обычаяев и преданий паломников в страну Востока, в Игру было привнесено понятие *контемпляции* (с. 59); . . . возникло и требование более глубокой и более душевной самоотдачи. . . Технические навыки процесса *созерцания* члены Ордена и игровых братств выносили из школ элиты, где *искусству контемпляции и медитации* обучали с великим тщанием. . . играющие и слушатели живут. . . *аскетической отрешенной жизнью абсолютного самоуглубления. . .* (с. 60).

AMOR FATI (что лучше, очевидно, не переводить, а передать словами самого Гессе — «приятие человеческого удела»).

«... у пас, потомков, сложилось совсем иное отношение к классической музыке, чем у представителей творческих эпох, наше *одухотворенное и не всегда* в достаточной мере *свободное от резиньющей меланхолии* почитание подлинной музыки. . . Главным итогом новой позиции или, вернее сказать, нового *подчинения ритму* культурного процесса был далеко заходящий *отказ от создания* новых произведений искусства, постепенный *отход* служителей культуры *от мирской предприимчивости* и, что не менее важно и как бы венчает все, — рождение Игры в бисер, или Игры стеклянных бус (с. 49); . . . Основные черты классической музыки: *знание о трагизме человеческого бытия, приятие человеческого удела, мужество и ясность!* Будь то. . . или *тихая, сосредоточенная готовность к смерти*, как у Баха — это неизменно некое *противление*, некая *неустрешимость*, некое *рыцарство*, и во всем этом отзвук *сверхчеловеческого смеха, бессмертной ясности*» (с. 64).

Далее, восходя от комплексов знаков к тем или иным характеризуемым ими концептам, читатель начинает замечать, что семантические поля пересекаются, концепты взаимодействуют, образуя как будто парные противопоставления. Так, в приведенных отрывках *иерархия* противостоит *личности, медитация* влечет за собой идею *воспитания, обучения, amor fati* неизбежно взаимодействует с *музыкой, trascendere* характеризует ступени движения и совершенствования *истории*. При последующем углублении в текст оказывается, что некоторые концепты всегда определяются в связи с другими: таковы *amor fati* и *Инь-Ян*, которые, будучи общезначимыми и вполне

самостоятельными, приводятся к выражению лишь в контексте музыки.

### ИНЬ-ЯН

«Истоки музыки лежат далеко. Она рождается из меры, и корни ее в великом Едином. Великое Единое рождает два полюса; они рождают силу темного и светлого. . . Музыка зиждется на гармонии неба и земли, на соразмерности темного и светлого. . .» (с. 51).

Постепенная кристаллизация названных концептов в тезаурусе читателя, по мере развертывания соответствующих тем в тексте, подводит его к пониманию того, что сама внутренняя структура (т. е. знаковое наполнение) концептов, так же как и их взаимоотношения, пронизана противоречиями, предстает часто в виде парадоксов. Так, в Кастальской «Воспитательной коллегии» воспитание зиждется на *истории*. Однако мы тут же узнаем, что в Кастилии *традицией* стало «забвение истории» и «полное включение личности в иерархию», когда нельзя узнать ни черт характера, ни имен. Как же тогда можно воспитывать? Ведь воспитание немыслимо без опоры на историю и на личность. Но этот парадокс пока зашифрован, и только на следующей странице (с. 36) эта внутренняя противоречивость вскрывается и объясняется автором: «... четче мы формулируем тезис, тем неумолимей он требует своего антитезиса». Итак, тезис: «безымянность наших коллегий и нашей духовной жизни»; антитезис: «каждая фаза этого развития. . . наиболее ярко предстает перед нами именно в самой личности преобразователя. . .» И здесь же далее возникает синтезис, который как будто и не был предусмотрен авторским объяснением, включающим лишь тезис и антитезис: «... личности того, кто был лишь неким инструментом данного изменения и усовершенствования» (т. е. Кнехта. — Ю. К.). Именно так, через диалектическое разрешение противоречия тезиса и антитезиса в синтезисе, разворачивается каждый из названных выше концептов, ко все более глубокому пониманию и характеристике которых подводит автор своего читателя. И теперь уже знаки, раскрывающие содержание концепта *личность*, например, естественным образом укладываются в читательском тезаурусе в три группы, во внутреннем напряжении между которыми и в возникающем за счет этого напряжения движении угадывается символика жизни и судьбы главного героя романа, Магистра Игры — Кнехта.

## ЛИЧНОСТЬ

— тезис: культ великих людей, биография, единственный и главный инспиратор, преобразователь, отклонение личности от нормы, аномалии, неповторимость, патологическое, выдающаяся, почитать. . .

— антитезис: человек без оригинальничанья, без при чуды, найти себя в общности, служить сверхличному, идеальный представитель некоего типа, а не индивидуальность, великий образ истории, анонимность. . .

— синтезис: герой, свежесть, удивительная энергия, конфликт с иерархией, мятежник, жертва, трагическое. . .

Но поскольку к этому моменту читатель уже осознает, что стержнем композиции всего романа является контрапункт, он ищет ритмического соответствия этой триаде также во взаимоотношениях самих концептов. Происходит перестройка читательского тезауруса, и на месте бинарных противоположений возникают замкнутые триады образов-идей: «музыка»—«амор fati»—«Инь и Ян»; «амор fati»—«личность»—«иерархия»; «личность»—«медитация»—«воспитание»; «иерархия»—«история»—«trascendere». Сложное переплетение и движение этих тем создает впечатление, что над всем этим стоит «музыка», как бы определяющая своими законами и само развитие каждого образа, и многослойную композицию романа, строящуюся на контрапунктирующих мотивах образов-идей и образов-людей. Но в конце концов становится ясно, что и сама «музыка» поглощается более универсальным, всепроникающим и как будто всеобъясняющим фантастическим образом «Игры». Это легко обнаружить и в приводившихся примерах, так как Игра — явно или скрыто (т. е. своими синонимами или атрибутами) — участвует в выкристаллизовывании всех основных концептов, всех образов-идей романа. С другой стороны, ее ключевая позиция становится очевидной в тех фрагментах текста, которые посвящены целиком этому образу, потому что в этих фрагментах мы снова встречаем все основные концепты романа, но уже в роли определителей, раскрывающих смысл этого начала начал в жизни Касталии:

«. . . Игра игр развилаась до некоего универсального языка, посредством которого оказывается возможным выражать ценности духа в осмысленных знаках и сопрягать их между собой.

Игра. . . тесно связана с музыкой и. . . велась по музыкальным и математическим правилам. При этом назы-

вались одна, две, три темы, затем их разыгрывали, варьировали, развивая подобно теме фуги или музыкальной фразе концерта. Игра могла, например, отправляться от определенной астрономической конфигурации или от темы баховской фуги, от фразы Лейбница или Упанишад, далее же, в зависимости от намерения и способностей играющего, вызванная к жизни главная мысль могла развиваться и шириться или обогащаться в своей выразительности через отзвуки родственных ей представлений. Если начинающему удавалось, например, в знаках Игры установить параллели между классической музыкой и формулой физического закона, то у искушенного мастера Игра, начиная с исходной темы, развивалась свободно в безграничных комбинациях. . . Сопоставления, противопоставления и, наконец, гармонические сочетания враждебных друг другу тем или идей, например, закона и свободы, индивидуума и общности, причем особое внимание уделялось тому, чтобы обе эти темы или оба тезиса развивались на абсолютно равных правах, беспристрастно, и из тезиса и антитезиса в наивозможно чистом виде был выведен синтез. . . Игры, оканчивающиеся негативным или пессимистическим, дисгармоническим аккордом, не пользовались любовью. . . [Игра] . . . означала изысканную, символически многозначительную форму исканий совершенства, высокую алхимию, приближение к единому в себе и превыше всех образов и множеств духу, стало быть — к богу. . . . так строились, музицировали и любомудрствовали фигуры и формулы Игры стеклянных бус на вселенском языке, вскормленном всеми науками и искусствами, в танце устремляясь к Совершенству, к чистому Бытию, к глубокой и полной Действительности . . . Да и сама Игра, с тех пор как Игры стали открытыми и был назначен Magister Ludi, находилась под покровительством Ордена и Воспитательной коллегии. . .

. . . в каждой стране были созданы Комиссии Игры в бисер и назначены верховные руководители со званием Магистра, а официальные Игры, проходившие под личным руководством Магистра, были возведены в ранг духовного празднества. . . Магистр, как и все высокие и высшие функционеры интеллектуальных сфер, сохранял анонимность. . . Помимо руководства публичными Играми, в обязанности Магистра входило попечение о мастерах и школах Игры. . .» (с. 60—63).

В итоге в восприятии читателя взаимоотношения тем, образов, идей, т. е. концептов романа, выстраиваются в некую иерархическую схему триад, подобную приводимой ниже;



Рис. 4

Однако на этом читательская концептуализация не кончается: вся схема триад и Игра как высший синтез, как символ, вбирающий в себя все содержание указанных противопоставлений, включаются в качестве отдельного члена в новую, более крупную и обобщающую триаду идей — «Игра»—«Касталия»—«внешняя жизнь», которая в свою очередь по закону контрапункта оказывается полностью изоморфной соответствующей триаде образов-людей — Тегулярис—Кнехт—Дезиньори. Причем средний член в каждой из обеих триад, призванный символизировать очередную ступень синтеза между полярными крайностями, приводится автором к краху, к трагическому концу, но читатель со своим теперь обогащенным и, возможно, несколько реформированным тезаурусом выходит из этих перипетий оптимистически и по-боевому настроенным (amor fati), вынося твердое убеждение в разрешимости всякого противоречия и одновременно в потенциальной разложимости всякого синтеза, в его чреватости своим собственным Инь и Ян, лежащим в основе неуклонного движения Действительности «вперед и выше».

Конечно, изложенная трактовка не выходит за рамки наивно-непосредственного читательского восприятия, и будь на его месте критик, он не преминул бы углубиться также и в авторский тезаурус, в котором обнаружил бы, вероятно, «феноменологическое понятие о мире», а в永恒ном напряженном противоборстве жизни и духа, торжестве их синтеза через сублимацию тезиса и антитезиса —

элементы шелеровской философской антропологии<sup>87</sup>, так же как в идее «пробуждения» Европы к «подлинной рефлексии» соблазнительным кажется усмотреть влияние на Гессе философии истории К. Ясперса с его взглядом на Европу как на «полуостров Азиатского материка»<sup>88</sup>. Но хочется надеяться, что сам механизм формирования концептов, который нас в данном случае главным образом и интересует, в тезаурусе критика и наивного читателя в основных своих чертах остается одним и тем же.

Ведь критик — тот же читатель, и различие их позиций касается прежде всего глубины реконструкции, глубины «достраивания», «доопределения» изображаемого в тексте. Всякий текст, согласно Р. Ипгардену<sup>89</sup>, принципиально неполон, и процесс чтения и понимания есть одновременно процесс реконструкции, восполнения недостающего. К этому надо добавить, что и тезаурус, построенный на основе любого текста, будет неизбежно неполон. Однако было бы ошибкой полагать, что восполнение текста в процессе чтения может изменить сам тезаурус, заложенный автором в этот текст при его производстве. Восполнение всегда идет только в пределах генерирующего текст тезауруса. Причем «наивный» читатель в этом процессе ограничивается восполнением текста, выявлением текстового мини-тезауруса и последующим соотнесением его со своим читательским тезаурусом. Критик же идет дальше: он не только восполняет текст, что необходимо для его понимания, не только строит по тексту мини-тезаурус, но и достраивает последний до полного, привлекая другие тексты того же автора, обращаясь к контексту эпохи, используя биографические данные, т. е. стремится по возможности адекватно воспроизвести весь авторский тезаурус.

Если основываться на предпосылке о существовании корреляции между идеальным тезаурусом-миром и реальным тезаурусом-словарем, то последний может стать эффективным инструментом критического анализа литературного текста. В таком своем качестве эталона, модели тезауруса-мира, тезаурус-словарь используется, например, в анализе текстов двух близких по тематике античных авторов. На базе словарей Роже и Дорнзайфа была составлена классификационная схема из 11 классов (с разветвленной системой разрядов внутри каждого класса), т. е. некоторый сокращенный тезаурус. Вся лексика (**существительные и прилагательные**) басен Авиана и Бабрия

распределялась по этим классам, причем учитывалась частота употребления слова в отдельном значении для каждого автора. На основе этих данных определялся удельный вес каждого тематического класса для того и другого автора, что позволило установить, каким классам (темам, концептам) автор отдает предпочтение, а в итоге сопоставительного анализа сделать идеологическое заключение об отношении «баснописцев к окружающему их миру, который у Авиана рисуется довольно смутным и неопределенным миром знати и черни, а у Бабрия отличается значительной конкретностью и разнообразием»<sup>90</sup>.

Последний аспект семантического пространства тезауруса — «концепт—концепт», тоже относящийся к характеристике позиции воспринимающего (слушающего, читателя), по принципу своей работы не отличается, на наш взгляд, от рассмотренного канала «знак—концепт». Его психологическим аналогом можно считать процесс так называемого «быстрого чтения», конструктивными, научно-методическими разновидностями которого являются индексирование и контент-анализ. Хотя успех «быстрого чтения» зависит только от установки, от ясного осознания читателем «чего он хочет», сам механизм его представляет общий интерес. «Быстрое чтение», как известно, представляет собой чтение «по понятиям», при котором информация кодируется не цифрами, не словами, не предложениями и не текстами, а наиболее экономным и емким способом — с помощью идей. При этом различают два этапа ее переработки: сам процесс восприятия, понимания, и процесс запоминания, которые по своей структуре практически оказываются тождественными. Картина преобразования текстового содержания в «смысловые вехи» «внутренней речи» при быстром чтении хорошо согласуется с развивающей здесь метафорой об идеальном читательском (слушательском) тезаурусе как организующей основе читательского (слушательского) восприятия: «Таким обра-

---

<sup>87</sup> Scheler M. Die Stellung des Menschen im Kosmos. Darmstadt, 1928, S. 46—53.

<sup>88</sup> Jaspers K. Vom Ursprung und Ziel der Geschichte. Zürich, 1949, S. 83—88.

<sup>89</sup> Ингарден Р. Исследования по эстетике. М., 1962.

<sup>90</sup> Борецкий М. И. Опыт использования систематического словаря для сопоставительного анализа басен Авиана и Бабрия. — Іноземна філологія, вип. 49. Питання класичної філології, № 15. Львів, 1978, с. 71.

зом при чтении слова распознаются благодаря их связям с понятиями и общей концептуальной моделью, т. е. некоторой алгоритмической схемой или совокупностью соотношений информативных единиц смысла, которые конструирует сам читающий. Таким алгоритмическим распознаванием вербальных речевых образов люди занимаются постоянно: по-видимому, оно основано на процессе, в котором слова служат лишь своего рода «ключами», помогающими конструировать представление смысла того сообщения, которое программируется мозгом, или, другими словами, конструировать информационную модель ситуации по тексту»<sup>91</sup>. Следовательно, уже на первом этапе «быстрого чтения», па этапе понимания мы минуем описанный выше процесс обобщения знаков в концепты и оперируем непосредственно самими концептами. Этап запоминания предполагает те же процедуры — «смысловую группировку материала» и выделение «смысловых опорных пунктов», а затем «соотнесение полученных результатов с уже известным»<sup>92</sup>.

Характеризуя процессы понимания текстовой информации и хранения ее в памяти, некоторые авторы говорят о «табличном чтении», которое в этом случае противопоставляется «линейному», т. е. обычному, чтению. Однако речь в этом случае идет не о направлении движения глаз по тексту — вертикальном при «быстром» и построчно-линейном при «медленном» чтении, а о способе фиксирования самой информации. Линейное — это чтение «случайное»: при движении по тексту вдоль строчек, слева направо и сверху вниз, читатель сталкивается с определенными количествами несистематизированной, неупорядоченной информации, с набором случайностей, которые сводятся затем в мысленную «таблицу». Таким образом, в этой модели<sup>93</sup> табличное чтение следует за линейным, является его результатом, тогда как в нашем представлении восприятие информации читательским тезаурусом как по каналу «знак—концепт», так и по каналу «концепт—концепт» происходит с известной антиципацией, с некоторым опережением ее текстового отражения в мозгу реципиента<sup>94</sup>: читательский тезаурус как бы предписывает, предопределяет способ отражения и обобщения. Отсюда так называемый «эффект страусового крыла», когда «в первую очередь принимается информация, отвечающая установке (т. е. стандартному тезаурусу. — Ю. К.), во вторую очередь — противоречащая ей»<sup>95</sup>.

Процедура индексирования, применяемая при обработке научно-технической информации в разного рода информационных системах, предполагает выбор из текста ключевых понятий данной области знания, выбор, который осуществляется, как правило, с помощью соответствующего отраслевого тезауруса. «Под индексированием... понимается интеллектуальный процесс формирования поискового образа документа (ПОД) или запроса (ПОЗ), выражаемого на некотором информационно-поисковом языке любого типа (дескрипторном или классификационном). Такое определение индексирования является более широким, чем то, которое дается, например, в „Терминологическом словаре по научной информации“: „Индексирование — классифицирование и простановка индексов классов, к которым отнесены классифицируемые документы“»<sup>96</sup>. Постановка в один ряд индексирования и «чтения по понятиям» не приводит, на наш взгляд, к дальнейшему расширению трактовки самого индексирования, поскольку в обоих случаях мы имеем: а) целевую установку в процессе чтения; б) отбор из текста кандидатов в дескрипторы или ключевые слова; в) базовый тезаурус, с которым соотносятся понятия, претендующие на роль ключевых. Тем не менее при сопоставлении индексирования с «быстрым чтением» или просто чтением создается парадоксальное впечатление о разном направлении обоих процессов: для понимания при чтении необходимо до-составление, т. е. известное расширение текста, тогда как индексирование в качестве первого этапа требует, наоборот, сжатия текста, приведения изложения к телеграфному стилю, особенно если учесть, что «аннотирование и рефериование являются этапом, составной частью индексирования»<sup>97</sup>. Думается, что такое впечатление чисто внешнее, поскольку направляющим и ограничивающим

---

<sup>91</sup> Кузнецов О. А., Хромов Л. П. Быстрое чтение. М., 1976, с. 28.

<sup>92</sup> Там же, с. 46.

<sup>93</sup> Rhétorique de la poésie: lecture linéaire, lecture tabulaire. Paris, 1977.

<sup>94</sup> Караплов Ю. Н. Асимметрия языкового знака во времени. — В кн.: Современные проблемы литературоведения и языкознания. М., 1974.

<sup>95</sup> Воробьев Г. Г. Теория тезаурусов в анализе коммуникаций. — В кн.: Семиотика и информатика, вып. 11. М., 1979, с. 9.

<sup>96</sup> Якушин Б. В. Алгоритмическое индексирование в информационных системах. М., 1978, с. 29.

<sup>97</sup> Там же, с. 34.

Моментом обоих процессов служит тезаурус — идеальный «читательский» или вполне реальный, отраслевой, — через который можно определить любую форму фиксации содержания текста: «Различая три степени сокращения:  $\emptyset$  — полный тезаурус,  $\{\emptyset\}$  — неполный,  $\{\{\emptyset\}\}$  — сокращенный и три уровня формализации:  $\alpha$  — отсутствие,  $\beta$  — частичная,  $\gamma$  — полная, мы можем классифицировать вторичные источники информации:  $\alpha \{\emptyset\}$  — резюме,  $\beta \{\emptyset\}$  — аннотация,  $\alpha + \beta \{\emptyset\}$  — реферат,  $\gamma \{\emptyset\}$  — при книжный указатель,  $\alpha \{\{0\}\}$  — синтетический заголовок,  $\beta \{\{\emptyset\}\}$  — аналитический заголовок,  $\gamma \{\{\emptyset\}\}$  — ключевые слова»<sup>98</sup>.

Контент-анализ, как и только что описанная процедура, ставит своей задачей конденсированное выражение смысла текстов, но от «быстрого чтения» и от индексирования его отличает то, что он предназначен для эксплицирования латентного содержания источников, отражающего либо социальную, идеологическую тенденцию, либо мировоззренческую картину мира определенных слоев общества. В условиях современной буржуазной культуры контент-анализ одинаково применим и к средствам массовой коммуникации (газета, радио, телевидение), и к произведениям массовой литературы (комиксы, реклама, приключенческие книжки, литература серии know how: как разбогатеть, как добиться успеха, как получить власть над людьми, как завоевывать друзей и оказывать влияние на других и т. п.).

Если согласиться с представлением об идеальном, постоянно флуктуирующем тезаурусе-мире отдельной личности, то контент-анализ предстает как способ единовременной, единоразовой фиксации фрагмента этого тезауруса на материале текстов массовых коммуникаций, т. е. в его статистически и социально усредненном варианте. Сополагаемость контент-анализа с рассмотренными выше процедурами существует, в частности, из следующей его характеристики: «. . . Контент-анализ действительно оказывается специфическим способом научного „прочтения“ текстов. Это есть: а) „чтение“ множества текстов как некоего целого, безотносительно к индивидуальным (的独特的) особенностям каждого текста; б) „чтение“, предполагающее использование формализованного наблюдения и — как правило — статистических процедур; в) „чтение“, имеющее конечной целью выявление или измерение некоторой социальной тенденции, ре-

презентированной данной совокупностью текстов. Фигурально говоря, контент-анализ является исследованием „мира“ текстов в целях познания „мира“ социальных отношений<sup>98</sup>. Таким образом, это тоже своего рода «чтение по понятиям», которое роднит с индексированием использование формальных процедур и за которым стоят те же три общих момента, названных выше: целевой параметр, отбор единиц из текста и базовый стандартный тезаурус. Роль последнего в контент-анализе выполняет специальная «понятийная схема», или «стандартизованный категориальный аппарат», структура которого может варьироваться в зависимости от направленности, методики и объекта анализа. Так, в одной из школ<sup>100</sup> этот аппарат состоит из 100 ключевых слов, пригодных к анализу неполитических текстов, и 25 слов, необходимых в анализе политического или иного материала, выявляющего общественное мнение. Эти 125 концептов охватывают наиболее распространенные способы выражения оценочных суждений и образуют базисный словарь анализа оценок. В этом случае контентный словарь представляет собой неполный, или мини-тезаурус. Компьютеризация и универсализация контент-анализа потребовали создания соответствующего полного словаря, который был построен, например, в программе General Inquirer, реализованной в Массачусетском технологическом институте<sup>101</sup>. Особенность такого словаря в том, что он по сути дела представляет собой перевернутый тезаурус, в котором единственным входом является «знак—концепт», поскольку семантика отдельного слова характеризуется здесь набором дескрипторов, в семантические поля которых это слово входит, т. е. дескрипторы, или концепты «понятийной схемы» контент-анализа, выступают как своего рода маркеры, классификаторы, в своем комплексе отражающие функционирование данного слова в исходной

---

<sup>98</sup> Воробьев Г. Г. Указ. соч., с. 30.

<sup>99</sup> Смысловое восприятие речевого сообщения (в условиях массовой коммуникации). М., 1976, с. 145—146.

<sup>100</sup> Andersson B. The quantifier as qualifier. Some notes on qualitative elements in quantitative content analysis. Gothenburg, 1974.

<sup>101</sup> Kelly F., Stone Ph. Computer recognition of English word senses. Amsterdam—Oxford, 1975, p. 1, 5, 82, 112—113; Holsti O. R. Content analysis for the social sciences and humanities. Addison-Wesley publishing company, 1969, p. 150—168,

совокупности текстов<sup>102</sup>. Известную перекличку с такого рода словарями можно видеть в полученном на первом этапе нашего эксперимента по машинному построению русского тезауруса распределении дескрипторов по словам. Фрагмент этого распределения, т. е. слово с указанием всех дескрипторов, в дефинициях которых обнаружились общие с ним семантические множители, приводится здесь в § 7. В таком виде это тоже перевернутый тезаурус, отличающийся от контентного словаря тем, что маркеры (дескрипторы) в данном случае характеризуют не функционирование слова в текстах, а семантику его лексикографического инварианта, поскольку источником служат словарные статьи толковых словарей.

В заключение этого раздела, подводя итог рассмотрению взаимоотношений тезауруса и текста, можно сказать, что всякий текст либо есть покушение на стандартный тезаурус, коль скоро он содержит потенциальную попытку изменить его, либо, наоборот, имеет целью активно противостоять таким попыткам, сохранить тезаурус без изменения. Надо сказать, что попытки изменить тезаурус весьма редко — во всяком случае, на протяжении жизни одного поколения — бывают удачными. Известная консервативность коллективного тезауруса-мира обусловлена, во-первых, значительной устойчивостью семантических или смысловых полей в тезаурусе личности<sup>103</sup>, во-вторых, относительной социальной стабильностью личности.

---

<sup>102</sup> Исходная совокупность текстов в этом случае была достаточно представительной и включала 6 млн. словоупотреблений, но тем не менее надежность контентного словаря была невысока из-за трудностей адекватного выбора из перечня, характеризующего каждое слово, дескрипторов, полностью отвечающих конкретному употреблению данного слова в конкретном контексте. Дело в том, что в комплект дескрипторов буквенного слова па первом этапе включались недифференцированно все классификаторы полисемантического или омонимичного слова, характеризующие все его контексты независимо от того, в каком значении фигурировало это слово в том или ином из них. Другими словами, полисемия и омонимия слов естественного языка порождала трудности в идентификации отдельного значения, вела к размытости, неопределенности смысла (*ambiguity*). Эти трудности потребовали создания в дополнение к обычному контентному словарю (который в принципе является словарем тезаурунского типа) еще «дизамбигуэнтного» словаря (*the disambiguation dictionary*), т. е. словаря идентификации. См., напр.: *Kelly E., Stone Ph. Op. cit.*, р. 123—267.

<sup>103</sup> См. об этом, напр.: *Лурия А. Р. Язык и сознание*. М., 1979, с. 91—114.

## Глава III

# ОПЫТ КОНСТРУИРОВАНИЯ РУССКОГО ТЕЗАУРУСА С ПОМОЩЬЮ ЭВМ

---

### § 6. ФЕНОМЕНОЛОГИЧЕСКАЯ МОДЕЛЬ ТЕЗАУРУСА И ПОСТАВЛЕННЫЕ ЕЮ ПРОБЛЕМЫ

In der Beschränkung zeigt sich erst der Meister  
Und das Gesetz nur kann uns Freiheit geben.

J. XV. Goethe\*

#### Основные параметры общезыковых и специальных тезаурусов

Словарь-тезаурус, как ясно из предыдущего параграфа, представляет собой систематическое упорядочение лексики какой-нибудь научной или технической области, ориентированное на информационный поиск<sup>1</sup>, а в применении к общелитературному языку словарь такого типа чаще называют идеологическим или идеографическим<sup>2</sup>. В основе систематизации лежит иерархически организованная совокупность конституирующих понятий данной области, и вся относящаяся к ней лексика (как терминологическая, так и нетерминологическая) без остатка распределяется

\* Как в тесных рамках лишь себя проявят мастер,  
Так нам свободу дать сумеет лишь закон (*И.-В. Гете*).

<sup>1</sup> См., напр.: Словарь дескрипторов по химии и химической промышленности. М., 1967; Thesaurus of engineering and scientific terms. N. Y., 1967; Тезаурус по автомобилестроению. М., 1971; Thesaurus Okonomie. Berlin, 1971; Тезаурус по книжной торговле. М., 1971; Тезаурус научно-технических терминов. М., 1972; Информационно-поисковый тезаурус по информатике. М., 1973; Macrothesaurus. A basic list of economic and social development terms. 1-st English ed. Organisation for economic co-operation and development. Paris, 1972; ERIC. Thesaurus of ERIC descriptors. 7-th ed. (With a special chapter on the role and function of the thesaurus in education). N. Y.—London, 1977.

<sup>2</sup> Roget P. M. Thesaurus of English words and phrases. Harmondsworth, 1964; Robertson T. Dictionnaire idéologique. Paris, 1859; Casares J. Diccionario ideológico de la lengua española. Barcelona, 1959; Dornseiff F. Der deutsche Wortschatz nach Sachgruppen. Berlin, 1959.

по этим понятиям, образуя вокруг каждого из них особую группу, или множество (семантическое поле, дескрипторную статью), которое содержательно характеризует, задает это понятие, представляя собой его «неявное определение». По сравнению с обычными объяснительными словарями — толковыми, терминологическими, энциклопедическими, которые исходят из закрытого списка слов и где упорядоченность достигается за счет «установленного беспорядка», т. е. алфавита, являющегося единственным «входом» в словарь, — тезаурусы имеют два преимущества: во-первых, они увеличивают количество входов в лексику данной отрасли знаний; во-вторых, благодаря свойству открытости (каждой словарной статьи, а значит — тезауруса в целом) и возможности постоянного пополнения новыми словами без перестройки общей схемы, тезаурус легко адаптируется к происходящим в языке (или системе знаний определенной области) изменениям и на каждом этапе развития той или иной отрасли, таким образом, способен адекватно отражать ее состояние.

Структура хорошо разработанного тезауруса состоит из следующих взаимосвязанных частей:

а) классификационной (либо синоптической, т. е. прогнозирующей, априорной, либо обобщающей, итоговой) схемы, делящей всю совокупность понятий на крупные понятийные зоны или области (фасеты) и системно упорядочивающей их; число таких зон обычно невелико — от 10 в отраслевом тезаурусе до 40—50 в универсальном, общеязыковом;

б) последовательного перечисления всех понятий (дескрипторов), характеризующих каждую понятийную зону (число дескрипторов может сильно колебаться от 200 в тезаурусе по субатомной физике до нескольких тысяч в общетехнических тезаурусах); внутри зон понятия расположены по логико-семантическому (синонимы, род, виды, ассоциаты), тематическому, фабульльному<sup>3</sup> либо какому-нибудь другому, но обязательно содержательному принципу;

в) собственно идеографической части, в которой под каждым понятием сгруппирована вся относящаяся к нему лексика; т. е. дескрипторная статья представляет собой семантическое поле, в качестве названия которого и используется понятие, или дескриптор; сами понятия в этой части упорядочены либо по алфавиту, либо в соответствии с общей классификационной схемой;

г) пермутационного (алфавитного) указателя понятий (дескрипторов) с адресом соответствующей зоны или зон;

д) пермутационного указателя слов (терминов и терминосочетаний) с адресами соответствующих дескрипторов. Эта последняя часть выполняется иногда в виде обычного толкового словаря, сама дефиниция в котором служит средством адресации слова, т. е. указания понятий (дескрипторов), к которым оно относится.

Таким образом, вместо одного входа в словарь в тезаурусе мы имеем как минимум четыре: при идентификации или поиске конкретного слова (термина) можно идти либо от общей схемы, либо от известной понятийной зоны (дескрипторной области), либо от заранее выбранных дескрипторов путем перебора множества относящихся к ним слов, либо, наконец, от пермутационного указателя слов путем подбора синонимов, эквивалентов, установления соответствующих дескрипторов, с тем чтобы найти адекватное данной задаче слово среди множества характеризующих тот или иной дескриптор слов.

Специальные тезаурусы по сравнению с общезыкими, т. е. идеографическими, словарями имеют то преимущество, что в них в явном, четком виде фиксирована собственно семантическая иерархия и указаны типы семантических отношений между всеми единицами иерархии. Но при такой глубокой разработанности «семантического метаязыка», они, с другой стороны, проигрывают общезыковым тезаурусам в объеме отражаемой ими другой языковой информации, в количестве реализованных в них лексикографических параметров. В информационно-поисковом тезаурусе все направлено на выявление семантики подъязыка некоторой отрасли в целом. «Семантическая

---

<sup>3</sup> «Под ф а б у л о й информационно-поискового тезауруса понимается схема категорий (фасетов) и лексических блоков, раскрывающая наполнение данного тезауруса. Целесообразно различать три уровня фабул: 1) универсальную фабулу, включающую категории, общие для всех тезаурусов, независимо от их тематической принадлежности; 2) типовые фабулы, представляющие собой детализацию универсальной фабулы применительно к комплексам или циклам отраслей знания, например, для технических, общественных или геологических наук; 3) частные фабулы, являющиеся детализацией типовой фабулы соответствующего комплекса и учитывающие специфические категории данной тематической области» (Соколов А. В. Некоторые проблемы типового проектирования информационно-поисковых тезаурусов. — В кн.: Структурная и прикладная лингвистика, вып. 1. Л., 1978, с. 172).

характеристика термина не ограничена дескрипторной словарной статьей. Своёобразие ИПТ состоит в том, что он заключает в себе весьма значительную семантическую информацию, которая, однако, содержится не только и не столько в самой дескрипторной словарной статье, сколько в тезаурусе, его рубриках и разделах в целом. Основное назначение ИПТ — максимальное раскрытие семантических отношений каждого термина. Все ИПТ различаются: а) характером и объемом включаемой в тезаурус лексики; б) характером и глубиной иерархии; в) наличием или отсутствием дополнительной аспектизации иерархии; г) числом и типом используемых парадигматических отношений; д) способом представления словаря»<sup>4</sup>.

Что касается глубины иерархии, то в специальных тезаурусах она достигает в ряде случаев 9 уровней<sup>5</sup>, что делает структуру словаря уже громоздкой, и такой тезаурус оказывается неудобным в эксплуатации. В практике же идеографических словарей глубина не превышает 6—7 уровней, и это находит теоретическое оправдание, поскольку, как показывают исследования, 6—7 семантических множителей с избытком хватает для дефиниции любого слова естественного языка и любого термина специального подъязыка<sup>6</sup>.

Вместе с тем специальный тезаурус, будучи средством систематизации лексики (терминологии) определенной отрасли, выступает и как достаточно компактный способ обобщения и упорядоченного представления уровня развития данной отрасли науки или техники. Иными словами, тезаурус как бы подводит итог некоторому этапу развития науки и отражает ее современное состояние. Именно это качество делает тезаурус основным элементом информационно-поисковых систем, позволяет использовать его для автоматического поиска информации, автоматического индексирования или реферирования текстов соответствующей отрасли знаний.

Вслед за техническими и естественными науками к идеи «тезаурусостроения» обратились и специалисты по общественным наукам. Известен опыт построения тезаурусов по лингвистике<sup>7</sup>, делаются первые шаги по созданию тезауруса в психологии<sup>8</sup>, в отраслевых институтах разрабатываются тезаурусы по юриспруденции и управлению. Актуальной задачей является построение социологического тезауруса, который будет знаменовать собой серьезный прогресс этой важной отрасли общественных наук и

создаст предпосылки для углубления и расширения автоматизации социологических исследований, для применения автоматического контент-анализа публицистических, политических, научных, социальных — вообще текстов массовых коммуникаций.

Необходимость разработки общеязыкового тезауруса, или тезауруса естественного языка, диктуется в настоящее время рядом причин как чисто научного, так и прикладного характера. Прежде всего, отсутствие русского тезауруса — на фоне довольно богатой мировой практики «тезаурусостроения»<sup>9</sup> — значительно обедняет отечественную лексикографию, сужает исследовательские возможности в области лексико-семантической системы русского языка, тормозит развитие словарного дела в стране, затрудняя создание словарей новых типов, препятствует, наконец, внедрению эффективных методов обучения русскому языку в национальной средней и высшей школе. Далее, как показали практика применения специальных тезаурусов, а также теоретические исследования последнего времени, отказ от использования естественного языка, его наиболее общих, внетематических, связей и естественных семантических полей является причиной недостатков отраслевых тезаурусов, их неполного соответствия задачам, для которых они предназначаются<sup>10</sup>. Дело здесь не

<sup>4</sup> Герд А. С. и др. Информационно-поисковый тезаурус как объект лексикографии. — В кн.: Структурная и прикладная лингвистика, вып. 1. Л., 1978, с. 165.

<sup>5</sup> Там же, с. 167.

(<sup>6</sup>) Мадоян Р. Теория идеографии и методические приемы идеографической классификации лексики. — Изв. Ереванского ун-та. Сер. обществ. наук, 1979, № 30.

<sup>7</sup> Никитина С. Е. Тезаурус по теоретической и прикладной лингвистике. Автоматическая обработка текста. М., 1978; Тезаурус информационно-поисковый по языкознанию. М., 1977.

<sup>8</sup> Мириманова М. С. Тезаурус как метод выявления системы понятий психологии. — Вопросы психологии, 1976, № 5.

<sup>9</sup> Помимо сноски 2, см., напр., библиографию в кн.: В. В. Морковкин. Идеографические словари. М., 1970.

<sup>10</sup> Сидорченко В. Д. Семантическая структура тезауруса: современное состояние и направления ее совершенствования. — НТИ, сер. 2, 1976, № 9, с. 10; Арапов М. В., Шрейдер Ю. А. Интегральная система и диалог. — НТИ, сер. 2, 1977, № 1, с. 3—4; Соколов А. В. Об одном возможном подходе к обеспечению совместности ИПТ (универсальная фабула информационно-поисковых тезаурусов). — Там же, с. 21; Маршакова И. В. Построение информационно-поискового тезауруса методом дистрибутивно-статистического анализа. — НТИ, сер. 2, 1977, № 5, с. 11.

только в том, что общеязыковой тезаурус, фиксируя в структуре и взаимоотношениях своих рубрик, разделов, зон, областей широкие возможности «невербального соединения идей» (Локк), обеспечивает учет человеческих ценностей и оказывается тем самым основой для разнообразных эвристик. Как подчеркивают специалисты, формализованные понятия, относящиеся к определенным научным областям, зачастую выглядят обедненными и выхолощенными на фоне понятий общеязыковых. «Одной из наиболее важных черт развития современной физики является уяснение того, что понятия естественного языка при всей их слабой определенности оказываются в ходе расширения знания более устойчивыми, нежели точные термины научного языка, полученные как идеализация при изучении ограниченных групп явлений. И это в общем не удивительно, поскольку понятия естественного языка формируются в непосредственной связи с реальностью, репрезентируют саму реальность. Верно, конечно, что они не очень хорошо определены и могут поэтому изменяться во времени, как изменяется и соответствующая им реальность. Однако они никогда не теряют непосредственной связи с этой реальностью. Научные же понятия представляют собой идеализации: они выводятся из опыта, полученного с помощью изощренного экспериментального оборудования, они точно задаются через аксиомы и определения. Благодаря этим точным определениям можно связать данные понятия с математическим аппаратом, математическим путем вывести громадное разнообразие возможных явлений в соответствующей области. Но в ходе этого процесса идеализации и точного определения утрачивается непосредственная связь с реальностью. Хотя понятия продолжают довольно точно соответствовать реальности в той части природы, которая была объектом научного исследования, но их соответствие может быть потеряно в иных частях, характеризующихся иными типами явлений»<sup>11</sup>.

Таким образом, дальнейшее совершенствование структуры и эффективности применения даже специальных тезаурусов оказывается невозможным без создания тезауруса естественного языка, без общеязыкового русского тезауруса. В этом отношении проведенная в Институте социологических исследований АН СССР работа по машинному построению и анализу семантических полей идеографической части русского тезауруса является необходимой ба-

зой и важной предпосылкой для разработки тезауруса по социологии. Преимущества машинной обработки словарного материала при построении тезауруса не нуждаются, очевидно, в особых доказательствах: ЭВМ не только обеспечивает скорость и максимальную полноту анализа, но и создает возможности для параллельных дополнительных «промеров» словаря, выявления многих существенных семантических характеристик, установление которых ручным способом невозможно или крайне затруднительно (частотность слов и семантических компонентов; количественная оценка силы связи слов друг с другом, слов с дескрипторами и дескрипторов друг с другом; измерение степени семантической непрерывности слова и наполнения семантических полей и др.).

### **Обработка параметров компилятивного уровня**

В составе исходных данных при построении тезаурусов в специальных областях и при построении общезыкового тезауруса имеются принципиальные различия. В специальных областях состав дескрипторов установлен довольно точно и полно, а хорошо известная структура самой отрасли однозначно определяет базисные отношения между ними, позволяя группировать их в дескрипторные области, фасеты, строить фабулы. Таким образом, при составлении отраслевых тезаурусов исследователи исходят из трех групп данных: набора дескрипторов, семантических базисных отношений между ними, т. е. классификационной схемы, и набора лексики, т. е. слова и терминов отрасли. Семантические же связи, семантические отношения на нижнем уровне — и терминов друг с другом внутри одного поля, и каждого термина со своим дескриптором (синонимия, антонимия, аллонимия, род—вид, часть—целое, ассоциаты) — являются как бы результатом тезаурунского построения; в структуре тезауруса они фиксируются, а источником наблюдения над ними служат тексты. В случае общезыкового тезауруса в распоряжении составителя лишь две группы достаточно надежных данных: словник, т. е. набор лексических единиц, подлежащих классификации, и семантические связи между этими единицами, заданные толковыми словарями языка (подробнее см. ниже). Здесь нет ни удовлетворительно

---

<sup>11</sup> Heisenberg W. Physics and philosophy: the revolution in modern science. N. Y., 1958, p. 200.

определенного набора дескрипторов, ни тем более базисных отношений между ними, под которыми понимается разбиение множества дескрипторов на дескрипторные области и систематизация взаимосвязей между дескрипторами внутри областей. Трудность для составителя заключается в том, что за общенациональным языком нет такой прозрачной для понимания и строго упорядоченной схемы, какой характеризуется структура той или иной научной, технической отрасли, хотя в идеале в общенациональном литературном языке должны быть отражены, представлены в сжатом виде результаты всех наук, технических областей, искусств, и идеально мыслимая классификационная схема общеязыкового тезауруса, будучи обобщением наших знаний о мире, должна в этом случае представлять собой некий универсальный «философский алфавит», высшее всеобъемлющее знание. Попытки построения универсального философского языка известны с глубокой древности. Интересное воплощение эти попытки нашли в философских системах Декарта и Лейбница. Вообще говоря, намеченный ими путь дедуктивного построения понятийной картины мира, запечатленной в языке, не является, на наш взгляд, невозможным. Именно такой путь лежит, например, в основе программы общеязыкового тезауруса и разработанного его фрагмента — «Время»<sup>12</sup>. Предпосылкой подобного подхода служит тот факт, что набор основных понятий (дескрипторов) в естественном языке и связей между ними в известном смысле определяется мировоззренческими установками составителя, его «концептуальной моделью мира» (КММ). Однако принципиальное несовпадение КММ с «языковой моделью мира» (ЯММ)<sup>13</sup> не позволяет жестко задать подробную (т. е. вплоть до дескрипторов) классификационную схему для лексики общенационального языка без заведомых исажений в его семантике. Кроме того, такой путь, как всякий путь дедуктивных суждений, допускает множественность результатов.

Альтернативный подход, предполагающий выведение дескрипторов естественного языка из анализа семантических отношений в лексике, реализованный автором этих строк<sup>14</sup>, при большей однозначности и степени соответствия результатов собственно языковым закономерностям (а не отраслевой структуре) страдает тем недостатком, что не позволяет одновременно установить также и базисные системные отношения между выделенными дескрипторами.

Как известно, помимо дедукции и индукции, в логике возможен и третий путь — путь абдуктивных суждений, оставляющий место для эвристики: «Все люди смертны. Сократ смертен. → Сократ человек»<sup>15</sup>. Для нашей задачи такой путь состоит в принятии некоторого априорно полученного набора дескрипторов с последующим наведением системных отношений в этом наборе, построением классификационной схемы тезауруса — уже по результатам идеографического распределения лексики — и коррекцией самого исходного набора. Таким образом, исходными для нас должны быть три массива параметров: перечень дескрипторов (основных «идей», понятий) русского языка; словник, т. е. перечень основных элементов словаря — слов, подлежащих распределению по семантическим полям дескрипторов; генерализованные, постоянные семантические связи между словами, па основе которых и осуществляется это распределение. Перейдем к характеристике способов формирования каждого из названных параметров и их свойств.

Каталог дескрипторов формируется методом компиляции: понятийные наборы наиболее полных идеографических словарей с хорошо разработанными классификационными схемами<sup>16</sup> накладываются один на другой и, взаимодополняя друг друга, суммируются. В принципе при таком наложении возможна и противоположная операция, т. е. сравниваемые наборы могут «вычитаться» друг из друга, в результате чего мы получим некоторое концентрированное инвариантное понятийное ядро, общее для сопоставляемых языков, своеобразную универсальную фабулу (ср. выше). Но в данном случае целесообразнее пойти по пути максимализации числа дескрипторов, заранее допуская некоторую их избыточность, поскольку

---

<sup>12</sup> Морковкин В. В. Опыт идеографического описания лексики (апализ слов со значением времени в русском языке). М., 1977.

<sup>13</sup> Бритян Г. А. Очерки по анализу философского знания. Ереван, 1979, с. 102—134.

<sup>14</sup> Карапулов И. Д. Общая и русская идеография. М., 1976, с. 58—68, 119—142.

<sup>15</sup> Александров Е. А. Основы теории эвристических решений. М., 1975, с. 17—22.

<sup>16</sup> Помимо названных в сноске 2 словарей Роже, Касареса и Дорнзайфа, в сравнении использованы: Бабов К., Въргулев А. Тематичен руско-български речник. София, 1961; Hallig R., Wartburg W. von. Begriffssystem als Grundlage für die Lexikographie. Berlin, 1963.

она гарантирует от появления лакун, а на этапе объединения их в дескрипторные области эта избыточность автоматически будет снята./ Теоретическим основанием для приравнивания друг к другу и суммирования основных понятий разных языков (английского, болгарского, испанского, немецкого и французского) служит наблюдение автора, согласно которому различия в лексической семантике между языками уменьшаются, если при сравнении переходит на уровень все большей абстракции<sup>17</sup>. На уровне слов, т. е. когда объектами сравнения служат отдельные слова, эти межъязыковые различия максимальны. Именно этот уровень дает материал для гипотезы лингвистической относительности, согласно которой семантические языки принципиально несводимы один к другому. Следующий уровень — сравнение семантических полей, которые тоже носят в значительной степени национальный характер, — показывает еще заметные расхождения между языками. Однако относящиеся к более высокому, абстрактному уровню сами имена полей, т. е. дескрипторы, сохраняя частично национальную окраску, обнаруживают уже и значительную общность. Наконец, максимальные сходства между языками обнаруживаются в понятийных зонах и далее — в фабулах. Среди множества дескрипторов, полученных в результате суммирования, естественно, окажутся и чуждые системе русского языка. Однако после процедуры наполнения дескрипторов словами, т. е. распределения слов из словаря по дескрипторным статьям, те из них, которые не свойственны русской языковой системе, будут отсеяны, так как получат неудовлетворительное, недостаточное наполнение<sup>18</sup>.

Принцип максимализации числа дескрипторов предполагает также при формировании каталога дополнительное расширение их первоначального списка, полученного суммированием. Расширение осуществляется за счет включения понятий, которые, отсутствуя в сравниваемых словарях, являются вместе с тем необходимыми для русского тезауруса, поскольку отражают нашу современность, советскую действительность, нынешний уровень развития науки и техники. Так, в добавление к готовому списку составителем вводятся дескрипторы СОВЕТСКИЙ, КОЛХОЗ, АТОМНЫЙ, РАКЕТА и под. Источником пополнения списка дескрипторов являются, во-первых, имена полей из «Минимального идеографического словаря» русского языка<sup>19</sup>, во-вторых, проделанный автором выбороч-

ный анализ передовых статей в 30 номерах газеты «Правда» за 1976 г. (контент-анализ).

Таким образом был сформирован каталог дескрипторов, содержащий 1621 понятие и по определению целиком покрывающий семантическую сферу общелитературного русского языка. Этот список не является закрытым: как уже упоминалось выше, процесс построения семантических полей должен внести корректизы в набор дескрипторов, приведя к исключению полей неполноценных. С другой стороны, последующая процедура объединения дескрипторов в более крупные единицы — дескрипторные области и построения из них общей классификационной схемы тезауруса может привести к выявлению лакун в этой схеме, что потребует в ходе отладки тезауруса введения некоторых новых дескрипторов.

В каталоге дескрипторов помимо шифра и порядкового номера составляющих его единиц, расположенных по алфавиту, нашла также отражение предварительная информация о дескрипторных областях. С известной долей условности использована классификация Касареса, которая состоит из 38 областей:

- |   |                                      |
|---|--------------------------------------|
| 1. Религия                                    | 18. Форма                            |
| 2. Физика и химия                             | 19. Движение                         |
| 3. География,<br>астрономия,<br>метеорология  | 20. Положение                        |
| 4. Геология, мине-<br>ралогия,<br>горное дело | 21. Время                            |
| 5. Ботаника                                   | 22. Количество                       |
| 6. Зоология                                   | 23. Разум                            |
| 7. Анатомия                                   | 24. Оценка                           |
| 8. Физиология                                 | 25. Воля                             |
| 9. Питание                                    | 26. Поведение                        |
| 10. Одежда                                    | 27. Действие                         |
| 11. Жилище                                    | 28. Коммуникация                     |
| 12. Медицина                                  | 29. Искусство                        |
| 13. Ощущение                                  | 30. Государство                      |
| 14. Чувство                                   | 31. Обычай                           |
| 15. Бытие и изменение                         | 32. Закон и право                    |
| 16. Отношение, по-<br>рядок, причин-<br>ность | 33. Собственность,<br>имущество      |
| 17. Пространство,<br>геометрия                | 34. Армия,<br>военное искус-<br>ство |
|   | 35. Торговля                         |
|   | 36. Земледелие                       |
|   | 37. Животноводство                   |
|   | 38. Транспорт                        |

<sup>17</sup> Караполов Ю. Н. Указ. соч., с. 267—274.

<sup>18</sup> О критериях полноты семантического поля, или статьи идеографического словаря, см.: Караполов Ю. Н. Указ. соч., с. 211—214.

<sup>19</sup> Там же, с. 276—298 и 327—329.

Как легко видеть, в этой классификации отсутствует в качестве самостоятельной такая важная область, как «промышленность»; соответствующие дескрипторы рассеиваются по другим областям. Нет здесь и общего понятия «наука», хотя в отдельные области выделены названия естественных наук (2—8); дескрипторы, отражающие понятия точных наук, распределяются по областям 16, 17, 22, а связанные с общественными науками — по областям 28—33, 35. Намеченная в нашей программе переклассификация призвана привести в соответствие число, состав и взаимосвязи дескрипторных областей — с результатами распределения слов из словаря по дескрипторам, т. е. с результатами построения семантических полей, или собственно дескрипторных статей тезауруса.

Итак, с учетом сказанного, каталог дескрипторов выглядит следующим образом:

Номер	Шифр	Область	Текст
1160	12262	26	РАСПУСТВО
1161	12264	20	РАСПУТЫВАТЬ
1162	12266	23	РАССЕЯННОСТЬ
1163	12268	17	РАССТОЯНИЕ
1164	12270	23	РАССУЖДАТЬ
1165	12272	2	РАСТВОР

**Словарник**, т. е. набор слов, составляющих всякий словарь, комплектуется обычно либо исходя из функциональной направленности создаваемого словаря, либо, в случае его полифункциональности, универсальности, — по принципу наибольшей полноты охвата лексики языка. В нашем случае ни один из этих принципов не подходит, так как, с одной стороны, словарь-тезаурус является полифункциональным, что требует как будто максимализации словарника. С другой стороны, опытный характер и условия нашей экспериментальной работы, сроки подготовки материала и его машинной обработки требуют сокращения, минимизации словарного материала и не позволяют использовать готовые словарники многотомных ( $\approx 120$  тыс. слов) и даже однотомных ( $\approx 60$  тыс. слов) русских словарей. Тем не менее и минимальный словарник должен отвечать следующим основным требованиям: быть репрезентантом всей русской лексики, т. е. содержать адекватное отражение «языковой модели мира», отражать коллективный опыт носителей русского языка, а значит, быть основой для выполнения языком его главной функции — коммуникативной и служить целям общения и

взаимопонимания. Кроме того, объем словаря должен быть достаточным для проявления, реализации внутренних закономерностей организации всякого словаря<sup>20</sup>: он должен обладать семантической непрерывностью и давать возможность развернуть грамматику языка.

Первое из этих свойств выражается количественно в «правиле шести шагов», согласно которому расстояние между двумя произвольно выбранными словами, определяемое по наличию общих элементов в дефинициях, не превышает шести шагов, т. е. пяти промежуточных слов, через которые и устанавливается связь двух исходных. Если это правило не соблюдается и семантическая непрерывность нарушена, то требуется расширение словарика, введение в него дополнительных слов, восстанавливающих отсутствующие звенья. Это правило может иметь социально-психологическую и семиотическую интерпретацию. Психологический смысл его в том, что оно, очевидно, отражает некоторые фундаментальные особенности человеческой психики, использующей для организации и хранения в памяти больших массивов информации ассоциативный принцип<sup>21</sup>, причем глубина этих ассоциаций, как и глубина предложения в соответствии с гипотезой Ицгве, не превышает в кратковременной памяти  $7+2$  единиц. Это предположение имеет и экспериментальное обоснование. В исследовании семантической организации ассоциативных сетей для определения количества переходов между понятиями, совершенно не связанными друг с другом по смыслу, использовались тесты, в которых испытуемым предлагалось установить естественную ассоциативную связь в парах понятий типа «война—лопесток» или «лежащий—энергия»<sup>22</sup> и т. п. Эксперименты показали, что максимальная длина перехода между отдаленными по смыслу понятиями не превышает 5—6 шагов (например, война — солдат — идти — земля — цветы — лопесток; лежащий — загорает — солнце — водород — молекула — распад — энергия), что, с точки зрения семантической организации

---

<sup>20</sup> Каразулов Ю. И. Указ. соч., с. 67—70, 75—78.

<sup>21</sup> См., напр.: Берков Б. П. Ассоциативные связи слова и двуязычная лексикография. — В кн.: Вопросы романо-германской и славянской лексикографии. Минск, 1976.

<sup>22</sup> Старинец В. С., Агабабян К. Г., Недялкова Г. И. Экспериментальные исследования семантической организации ассоциативных сетей. — В кн.: Моделирование в биологии и медицине, вып. III. Киев, 1968, с. 6—11.

памяти, подтверждает установленное нами на словаре «правило шести шагов».

Семиотический же смысл «правила шести шагов» связан со статусом самого словаря в иерархии уровней языковой структуры: сам словарь может трактоваться как своего рода языковая память, в которой с помощью (и на материале) лексики закодированы системы правил, регулирующих применение в речи единиц фонологического, морфологического, синтаксического, семантического и pragmatischenkого уровней<sup>23</sup>.

Что касается второго требования, которому должен удовлетворять словарь, — возможности развернуть полностью грамматику языка, то оно связано с первым и интерпретируется как потенциальное свойство строить из составляющих его единиц связные тексты.

Словарь, обладающий указанными свойствами, представляет собой самостоятельную замкнутую систему, презентирующую достаточно адекватно весь лексический состав языка. Количественный объем такого словарника наиболее удачно определяется статистически: известно, что примерно 4000 наиболее частотных слов покрывают около 95% текстов общего, неспециального содержания<sup>24</sup>. Таким образом, казалось бы, материал для тезауруса можно взять как начальную часть какого-нибудь русского частотного словаря или провести усреднение списка по нескольким словарям, применив опять метод компиляции. Но из опыта обучения языку, из анализа речеупотребления в различных коммуникативных ситуациях установлено, что такой механический отбор не дает еще списка слов, достаточного для осуществления коммуникативной функции, и его приходится пополнить рядом менее частотных, но совершенно необходимых для общения единиц. Кроме того, полученный таким путем словарь не обладает свойствами замкнутой системы, так как в нем нарушена семантическая непрерывность, и вводить в него новые слова поверх частотного списка нужно также и для «замыкания» системы, восстановления отсутствующих звеньев в семантических связях слов друг с другом.

Практически комплектование нашего словарника проходило в несколько этапов, соответствовавших эволюции представлений автора об общей задаче работы. В итоге получено множество слов, отвечающее сформулированным выше требованиям и составившееся как сумма трех разных списков; словаря-минимума для иностранцев<sup>25</sup>,

Включающего приблизительно 4,5—5 тыс. слов; словника из 2,5 тыс. слов на букву «П», представляющего собой часть планировавшегося вначале для обработки словаря на 25—30 тыс. слов<sup>26</sup>, и, наконец, списка дескрипторов, которые на этом уровне, войдя в словарь, выступают здесь уже как равноправные его единицы, т. е. как обычные слова<sup>27</sup>. Таким образом, второй массив исходных данных — набор слов, подлежащих идеографической классификации, т. е. распределению по набору дескрипторов, составил словарь на 9 с лишним тысяч единиц.

Третий, последний массив исходных данных должен составлять набор семантических связей между словами в русском языке. Практика построения специальных тезаурусов прибегает здесь либо к методу экспериментальных оценок, либо к выявлению этих связей из текстов; иногда эти два метода соединяются. Анализ текстуальных связей дает наиболее достоверные данные о семантических отношениях слов, но он представляется и наиболее трудоемким. Одним из путей такого анализа является дистрибутивно-статистический метод: «Реальная совместная встречаемость двух лексических единиц (двух ключевых слов) и ожидаемая (теоретическая) встречае-

<sup>23</sup> Theoretische Probleme der Sprachwissenschaft. Berlin, 1976, с. 591 и след. Ср. с этим также нашу идею о «лексикографической параметризации языка», которая по сути дела и представляет собой совокупность этих языковых правил (см. гл. I).

<sup>24</sup> Фрумкин Ф. М. Некоторые практические рекомендации по составлению частотных словарей. — Русский язык в нач. школе, 1963, № 5, с. 77.

<sup>25</sup> Этот минимум, отвечающий требованию обеспечения возможности общения и взаимопонимания, был в свое время взят из готовившегося тогда к изданию, а теперь уже опубликованного пособия: Краткий толковый словарь русского языка (для иностранцев). М., 1978 — с любезного разрешения его редактора В. В. Розановой.

<sup>26</sup> Слова на букву «П», самые многочисленные в русском языке, отобранные из расчета общего объема словаря в 30 тыс., были уже кодированы для машинной обработки, и поэтому автор считал целесообразным включить их все в предназначавшийся для такого кодирования словарь-минимум (см. предыдущую сноскую), в котором они, естественно, поглотили 700—800 слов, составлявших здесь эту букву.

<sup>27</sup> Эта операция, гарантирующая установление семантической непрерывности в словаре, осуществляется в соответствии с основным принципом семантического кодирования (см. ниже), согласно которому единицы высшего уровня (в данном случае — дескрипторы) включаются в состав единиц низшего уровня (в данном случае — слов), но не наоборот.

Мость этих же единиц дает представление о силе связи этих элементов текста. Оценка расхождений между эмпирическими значениями и статистическими расчетами является основой метода дистрибутивно-статистического анализа»<sup>28</sup>. Степень обобщенности и постоянства установленных таким путем связей зависит от объема проанализированных текстов. Единичный текст всегда беднее, чем тезаурус. По тезаурусу можно построить любой текст, а по одному тексту нельзя построить тезауруса. Чем большее число текстов мы возьмем для анализа, тем более надежные сведения о семантических связях мы получим. В идеале для достижения абсолютной надежности и точности следовало бы проанализировать все тексты на данном языке. Понятно, что эта задача практически невыполнима.

Другая трудность при установлении этого набора заключается в неясности самого понятия «семантическая связь»: до последнего времени оно не лишено в лингвистике некоторого мистического налета. В самом распространении смысле оно трактуется как интуитивно ощущаемая близость двух слов, которая для синонимов, например, выражается в возможности замены их друг другом в одном и том же контексте с сохранением его мысли, но изменением стилистической окраски; для антонимов — это возможность их употребления в тождественном окружении при изменении смысла всего высказывания на противоположный, и т. д. При дистрибутивно-статистическом анализе, как мы видели выше, семантическая связь понимается как совместная встречаемость.

Наконец, в исследованиях сочетаемости слов семантическая связь выступает как синтагматическая их совместимость: «Таким образом, проводя границу между языковыми и внеязыковыми аспектами комбинаторики слов, можно заключить, что наиболее регулярны и, следовательно, наиболее „лингвистичны“ синтаксические правила, управляющие сочетанием слов. Именно синтаксическим правилам принадлежит решающая роль в аранжировке словосочетаний. Что же касается лексической совместимости, то она наиболее подвержена экстралингвистическим влияниям, в связи с чем конкретную лексическую совместимость/несовместимость нельзя считать абсолютной. То, что несовместимо в одних условиях (птица катится; книги препятствуют распространению знаний), может оказаться совместимым в других условиях («птица катится», если ее подстрелил охотник, и она, упав на

возвышенное место, катится вниз; «книги препятствуют распространению знаний», например, в том случае, если речь идет о пошлой, бессодержательной литературе»<sup>29</sup>. Для решения нашей задачи понятие семантической связи должно не только получить операциональное определение и четкое конструктивное воплощение, но и допускать количественную оценку ее силы.

Составитель общеязыкового тезауруса, находясь по отношению к составителям специальных тезаурусов в более трудном положении из-за неопределенности общей классификационной схемы понятий для языка в целом, обладает все же и одним преимуществом: в его распоряжении находится информация, которой зачастую лишены исследователи, работающие над отраслевыми тезаурусами (скажем, социологическим или психологическим), он располагает толковыми словарями общелитературного языка. В толковых словарях зафиксированы статистически усредненные (поскольку они выявлены из анализа значительного числа текстов) и социально апробированные (поскольку они отражают литературную норму), т. е. в достаточной степени генерализованные семантические связи в лексике языка. На возможность и желательность привлечения толковых словарей — в частности, в качестве средства, дополняющего результаты дистрибутивно-статистического анализа при установлении смысловых связей слов — указывают и некоторые составители специальных тезаурусов<sup>30</sup>.

В толковом словаре семантика слова, его смысл, значение заданы словарной дефиницией в правой части словаря. Сравнивая дефиниции двух слов друг с другом, мы можем установить, связаны ли эти слова по смыслу, и если да, то какова сила этой связи. Например, интуитивно кажется ясным, что слова *терпеть* и *выносливый* имеют нечто общее в их значениях, т. е. семантически связаны:

Терпеть 1. Безропотно и *стойко* переносить что-н. 2. Мириться с наличием, существованием кого-чего-н., допускать что-н. 3. Испытывать что-н. (неприятное, тяжелое). 4. Быть таким, что можно отложить, подождать.

---

<sup>28</sup> Маршакова И. В. Указ. соч., с. И.

<sup>29</sup> Зотова Т. Я. Сочетаемость русских глаголов преобразующего физического воздействия на объект (на материале языка научной литературы). Автореф. канд. дис. М., 1979, с. 7—8.

<sup>30</sup> Глотов Б. А. Об использовании лексикографических данных при построении тезауруса. — НТИ, сер. 2, 1975, № 9.

**Выносливый.** Физически сильный, *стойкий*, способный много вынести. Будем считать показателем этой связи наличие общих, т. е. одинаковых слов в правой части словаря <sup>31</sup>. В приведенном случае таким общим для сравниваемых слов компонентом является *стойко, стойкий*. Если число общих компонентов у сравниваемых слов увеличивается, значит, растет и сила семантической связи между ними. Так, с одной стороны, связь в паре *наука и теория*, как показывает количество выделенных единиц, сильнее, чем в паре *терпеть и выносливый*; с другой стороны, между двумя рассматриваемыми парами нет никакой связи, поскольку не обнаруживается ни одного общего компонента:

**Наука** 1. Система знаний о *закономерностях* в развитии *природы, общества и мышления*, а также отдельная отрасль таких знаний. 2. То, что получает, дает *опыт*.

**Теория** 1. Учение, система научных принципов, идей, обобщающих практический *опыт* и отражающих *закономерности природы, общества, мышления*. 2. Совокупность обобщенных положений, образующих *науку* или раздел какой-н. науки, а также совокупность правил в области какого-н. мастерства. 3. Сложившееся у кого-н. мнение, суждение, взгляд на что-н.

Итак, семантическую связь между словами, так же как между дескриптором и словами, будем определять на основании дефиниции в толковом словаре, считая показателем этой связи наличие в дефинициях общих элементов (т. е. совпадений слов с точностью до класса вокабул, или гиперлексемы) и производя измерение силы связи по количеству этих общих элементов. Схему распределения слов из словаря по словарным статьям дескрипторов, основанную на таком понимании семантической связи, будем называть феноменологической (т. е. реализуемой на уровне явлений семантической связности) моделью общеязыкового тезауруса.

### **Семантическое кодирование по квазиосновам как аналитическая процедура**

Смысл этой операции — в обработке третьего массива исходных данных, т. е. массива, содержащего информацию о семантических связях, с целью представления его, как и первых двух, тоже в виде каталога, в виде своеобразного «семантического алфавита». Кодирование можно

представить себе как процесс, обратный порождению: если при порождении текста, речи происходит подстановка лексем в соответствующих формах на ветви синтаксического дерева для получения целостной структуры предложения, то в нашем случае, наоборот, осуществляется свертывание синтаксических структур, представленных в предложении-дефиниции, их семантизация, лек- (Сикализация, превращение исходного текста в неупорядоченный набор смыслов, «семантических множителей»).

Такой подход к представлению предложения-дефиниции имеет некоторую аналогию с трактовкой семантики в системе, понимающей естественный язык, Т. Винограда: «В этом методе анализа СЛОВО является основным „строительным блоком“. Имеются различные классы слов — „существительное“, „прилагательное“, „глагол“ — и каждое слово есть интегральная единица: оно не расщепляется на гипотетические атомы (например, „dogs“ представлялось бы состоящим из „dog“ и „-s“, или „dog“ и „множественное число“). Вместо этого мы считаем, что каждое слово обладает признаками. Слово „dogs“ есть та же самая словарная единица, что и „dog“, но имеющая признак „множественное число“ вместо „единственного числа“. Слова „took“, „take“, «taken», „taking“ все являются одним и тем же базовым словом, но с разными признаками, такими как „причастие И“ (EN), „инфinitив“ (INF), „-ing“ (ING) и т. д.»<sup>32</sup> В отличие от охарактеризованного, ваш подход предписывает отказаться от некоторых «признаков» (в смысле Т. Винограда), но при этом мы далеко не полностью лишаемся возможности «интегрировать семантику с синтаксисом», поскольку, эlimинируя сведения о синтаксической сочетаемости, мы сохраняем в нашей записи информацию о лексической сочетаемости элементов дефиниции<sup>33</sup>.

Семантическая запись значения (дефиниции) слова *наука* в нашей системе выглядит следующим образом:

---

<sup>31</sup> Караплов Ю. Н. Структура лексико-семантического поля. — НДВШ, филол. науки, 1972, № 1; *Он же*. Общая и русская идеография, с. 181—186.

<sup>32</sup> Виноград Т. Программа, понимающая естественный язык. М., 1976, с. 34.

<sup>33</sup> Ср. у того же Т. Винограда: «Тем не менее уместно потребовать, чтобы определения содергательных слов, таких как «snake» («змея») или «walk» («гулять»), не предполагали детального знания синтаксиса языка» (указ. соч., с. 54).

*наука, система, знание, закономерность, природа, общество, мышление, отрасль, поучать, опыт.* Приравнивая подобную семантическую запись приведенной выше дефиниции того же слова, мы исходим из предположения, что объяснительные тексты, разновидностью которых являются толкования значений слов в словарях, дают минимальный прирост смысла за счет синтагматических «зазоров», т. е. за счет «синтаксических признаков» слов. Таким образом, используемый способ представления смысла эффективен лишь для текстов дефинитивных, объяснительных и имеет силу не для фрагментарных исследований, а только в случае глобального описания семантики на всем словаре.

Получение семантической записи — это только первый этап кодирования, который по своему содержанию совпадает с процессом лемматизации, т. е. сведёния текстовых форм слова к исходной, лексикографической — при составлении словников и конкордансов, отличаясь от него тем, что не все слова, входящие в дефиницию, подвергаются этому процессу. Дело в том, что некоторые составляющие дефиницию компоненты обладают настолько высокой степенью общности своего значения, что могут характеризовать любые слова языка, в том числе и совершенно между собой не соотносящиеся, т. е. не связанные по смыслу. Поэтому в семантическую запись слова *наука* не включены компоненты *развитие, такой, то, дать, отдельный*, поскольку *развитие*, например, входит в дефиницию слова *жизнь* и многих других, *дать* встречается в дефинициях слов *показательный, помочь, обращать* и др. Подобные компоненты называются нулевыми и называются списком<sup>34</sup>. Компонент *отдельный* не включается в семантическую запись потому, что в данном случае он является просто избыточным («отдельная отрасль»), поскольку *отрасль* уже содержит в себе этот смысл (ср. дефиницию «отрасль»). Избыточные семантические множители отличаются от пустевых тем, что не во всех дефинициях обладают нулевым содержанием и потому в ряде случаев не подвергаются исключению при кодировании, а выступают как носители тех или иных «признаков». Они тоже задаются списком, хотя провести четкую границу между нулевыми и избыточными не всегда удается. Вопрос этот решается однозначным образом уже не в рамках общей феноменологической модели построения тезауруса, а на уровне уточненной «асимптотической модели»,

Ё которой решающим для отнесения того или иного множителя к нулевым или содергательным является селективный частотный критерий (см. ниже).

Итак, первый этап кодирования — это обычная лемматизация с исключением из числа компонентов нулевых и присоединением к ним • — в виде самостоятельного семантического множителя — самого кодируемого слова в соответствии с общим принципом, согласно которому элементы высшего уровня всегда входят в состав элементов ближайшего низшего, но не наоборот: слова включаются в состав своих семантических множителей; дескрипторы — в состав слов, т. е. в словарь (ср. выше); единицы классификационной схемы входят в состав дескрипторов (ср. список дескрипторных областей, в котором название области совпадает с одним—тремя дескрипторами из нашего каталога).

Однако результат лемматизации еще недостаточен для выявления семантических связей. Хотя, вообще говоря, можно было бы ограничиться этим, и такой путь в принципе не отрицается, но он приводит к резкому увеличению как количества сравниваемых компонентов, так и абсолютной длины ( побуквенно) каждого из них, что серьезно затрудняет машинную обработку материала. Кроме того, лемматизация нейтрализует лишь словоизменительные отношения, т. е. приводит некоторую текстовую форму, употребленную в каком-то числе, падеже, времени, роде, и т. п., к исходной, словарной. Если мы ограничимся ею, то не сможем приравнять друг к другу такие компоненты, как *наука* и *научный*, *система* и *систематизация*, *природа* и *природный*, *знание* и *знать*, которые явно связаны семантически очень тесно. Значит, надо найти такую форму кода, чтобы в пей нейтрализовалась, помимо словоизменительных, и часть словообразовательных отношений.

Самым простым решением было бы выделение в качестве кода корня слова. Однако словообразовательные отношения настолько сложны, и приставочные модификации от одного и того же корня могут в такой степени изменить значение слова, что семантическая связь нарушится (ср. *казаться* — *оказаться*, *сказать* — *присказка*, *гадать* — *загадка* и т. п.). Поэтому был принят такой принцип, что

---

<sup>34</sup> См.: Караплов Ю. Н. Частотный словарь семантических множителей русского языка. М., 1980, с. 32—33.

начало слова оставлялось без изменения, а сокращению, усечению подвергалась правая часть; для приведенных слов коды выглядят так: «каз», «оказ», «сказ», «присказ», «гада», «загад». При таком кодировании происходит как бы сдвиг смысла влево, концентрация его на начале слова, что и составляет характерную черту «лексикалистического» подхода в противоположность подходу «синтаксическому», при котором релевантной является как раз другая, правая часть слова, и сдвиг смысла происходит вправо. Такой прием кодирования позволяет не считать релевантными различия между частями речи, образованными от одного корня, и некоторые другие словообразовательные отноше-

ния: гада<sup>ть</sup><sub>ние</sub>, загад<sup>ка</sup><sub>ывать</sub>, и т. д. Правая граница

кода никак не связана с морфемным членением слова,

она может проходить и внутри корня (поздр<sup>авить</sup><sub>авление</sub>),

и по его конечной букве (лад<sup>но</sup><sub>ть</sub>), и справа от нее

(представи<sup>тель(ство)</sup><sub>тель(ный)</sub>). Таким образом, получаемый сег-

мент не является ни корнем слова, ни его основой, которые представляют собой по сути дела прежде всего единицы (и понятия), обусловленные генетически, исторически и уж в последнюю очередь — семантически. Полученный сегмент — это своего рода квазиоснова, исключительно семантический идентификатор слова, и способ его получения именно таков, что он выполняет однозначную функцию — быть единицей смысла, единственным средством для опознания семантики слова. В этом отношении и можно утверждать, что план выражения квазиосновы ( побуквенный состав) равен, тождествен его плану содержания <sup>35</sup>. Иными словами, квазиоснова обладает свойством одноплановости, каким и должен обладать семантический множитель.

Правила кодирования семантической информации определяются тремя в известном смысле противоборствующими, разнонаправленными тенденциями: во-первых, стремлением к минимизации числа букв в коде (квазиоснове); во-вторых, необходимостью получения сегмента

с максимальной способностью к дифференциации, с максимальной дифференцирующей силой, что определяется самим статусом квазиосновы как семантического идентификатора слова и что ведет, естественно, к удлинению кода, увеличению числа букв; наконец, в-третьих, условием однозначного декодирования, т. е. сохранением возможности для того, кто работает с кодом, с каталогом семантических множителей, в любой момент восстановить слово (или гнездо слов), которое послужило исходным для данного сегмента и которое им семантически идентифицируется. Последнее условие также ведет к удлинению кода.

В соответствии с этими правилами строится алгоритм кодирования, который включает следующие шаги:

1. Внесение самого кодируемого слова («входного слова») в состав его собственных семантических множителей;

2. Исключение из дефиниции нулевых семантических множителей, в число которых, помимо слов очень широкой семантики (типа *явление*, *совокупность*, *состояние*, *место*, *величина*, *свойство*), входят грамматические слова, характеризующие способы действия (*повести*=*н а ч а т ь вести*, *придержать*=*л е г к а удержать*, *искать*=*п р о в е с т и* и *некоторое в ре м я в поисках*), а также предлоги, союзы, местоимения.

3. Отсечение правой части каждого семантического множителя, т. е. получение квазиоснов, которое осуществляется в несколько этапов:

— правая граница проводится по конечной букве корня независимо от того, приставочное это слово или нет: *крик-*, *наук-*, *утверд-*, *двиг-*, *навстреч-*, *вызов-*, *рассмотреть-*, *пропаганд-*;

— далее проводится сокращение числа букв за счет отсечения чередующихся согласных и беглых гласных корня (кри, нау, утвер, выз, дви), а при отсутствии чередования — за счет отбрасывания тех букв справа, которые представляются избыточными при декодировании, т. е. без которых слово может быть однозначным образом восстановлено по квазиоснове (навст, рассм, пропаг);

— проверка по словарю на омонимию сегментов, преследующая цель избежать возможности совпадения квазиоснов разных слов или возможности неоднозначного декодирования какого-то сегмента; так, из числа приве-

---

<sup>35</sup> Караулов Ю. Н. Общая и русская идеография, с. 186.

Данных рассм, как показывает словарь, допускает три атрибута: аттрактор, аттрактор и аттрактор. Поэтому его окончательный вид требует уточнения, и в картотеку (каталог) квазиоснов он попадает в виде рассмо.

4. Некоторые дополнительные процедуры доопределения дефиниции, в том числе:

— отрицательные трансформации (*не дающий полезной работы* → неполез нераб, *не имеющий отопления* → неотапл, *перестать спать* → неспа, *из-за отсутствия боязни* → безбоя, *сбивать с толку* → бестол и др.);

— создание нескольких искусственных образований для кодирования устойчивых (идиоматических) словосочетаний, вообще говоря, не характерных для объяснительных, дефинитивных текстов (*образ жизни* → образжи, *иметь в виду* → иvv, *точка зрения* → тз, *обмен веществ* → обмев и некот. др.). Таких семантических множителей-аббревиатур не более 10 единиц и все они, как показывает частотный словарь, обладают минимальной употребительностью в дефинициях<sup>36</sup>,

— пополнение числа семантических множителей в недостаточных, неполных дефинициях (доопределение) за счет раскрытия скобок и использования поясняющих примеров (такие случаи немногочисленны). Так, *прикоснуться* определяется в словаре С. И. Ожегова следующим образом: прикоснуться, -путь, -пешься, сов., к кому-чему. Дотронуться слегка. П. к руке. Большой не прикоснулся к еде (нисколько не съел). // несов. прикасаться, -аюсь, -аешься. Я к твоим вещам не прикасался (ничего не трогал, не брал). // сущ. прикосновение, -я, ср. Легкое п.

В соответствии с алгоритмом кодирования семантическая запись этого слова включает только два множителя — прикос дотр, что интуитивно представляется недостаточным. Однако в разъяснениях к примерам употребления слова *прикоснуться* (в скобках) встречается слово *трогать*, квазиоснова которого — тро — тоже включается в коды. Другой случай: дефиниция слова приклад II (вспомогательный материал для швейных изделий, обуви) слишком абстрактна и содержит почти все семантические множители высокой степени общности — прикла вспомог мате шв издел обу. Поэтому семантическая запись «доопределен», и к исходным добавлены компоненты из

той же дефиниции, конкретизирующие значение этого слова в скобках (подкладка, пуговицы) → подкла пого<sup>37</sup>.

После того, как каждое слово из словарника подверглось операции кодирования, получаем массив слов с семантическими кодами, который выглядит следующим образом:

Номер	Шифр	Число						
		множ.						
923	56358	2	<i>распутство</i>					
			распу	образжи				
924	56359	10	<i>распутать</i>					
			распута	раздел	разъед	спут	неспут	
			освоб	пут	разобр	рассмо	ясп	
925	56362	6	<i>распухнуть</i>					
			распух	взду		воспал	бол	чрезмер
			увел					
926	56363	8	<i>рассвет</i>					
			рассв	врем	восход	соли	нач	
			утр	рано	период			
927	56367	4	<i>рассеянный</i>					
			рассеян	неуме	несосред	невним		
928	56368	9	<i>рассказ</i>					
			расск	слов	излож	событ	худож	
			повест	проз	произв	небол		
929	56369	4	<i>рассказывать</i>					
			расск	слов	сообщ	излож		
930	56372	5	<i>расследовать</i>					
			рассле	рассмо	изуч	следс	всесто	
931	56373	3	<i>рассматри-</i>					
			вать					
			рассмо	опе	воспри			
932	56374	7	<i>рассмотреть</i>					
			рассмо	всмо	распоз	осмо	вник	
			разобр	обсу				
933	56375	7	<i>рассольник</i>					
			рассоль	мяс	рыб	суп	свар	
			сол	огур				
934	56377	7	<i>расспросить</i>					
			расспрос	зада	вопр	цель	узина	
			выяс	расспраш				
935	56379	8	<i>расстаться</i>					
			расста	уй	переста	невид	невстре	
			разой	иопрош	лиш			

<sup>36</sup> Карапулов Ю. П. Частотный словарь семантических множителей русского языка.

<sup>37</sup> Используемый нами прием имеет аналогию и в других системах анализа лексической семантики. Так, в рассмотренной в I главе системе SOLAR (см. сноска 179) «раскрытие скобок» в дефиниции позволяет извлекать так называемые «дополнительные признаки значения».

Окончание

Номер	Шифр	Число	множ.						
936	56380	6	расстояние	рассто два	пространс	раздел	пун	промежут	
937	56381	9	рассуждать	рассу излаг	мысл сужд	строи обсу	умозак бесе	последов	
938	56382	11	рассчитать	расчи предусм	исчис задум	учиты опред	данн увол	возмо расчет	
939	56383	7	рассчитывать	расчи надеж	надея полага	счи	возмо	возл	

Приписывание семантической информации дескрипторам, т. е. их кодирование, имеет некоторые особенности. Понятно, что эти особенности не касаются правил кодирования, т. е. того, «как» кодировать: алгоритм применяется один и тот же и для слов, и для дескрипторов. Особенности же связаны с тем, «что» кодировать, т. е. касаются отбора релевантных семантических множителей. При кодировании словарника, как мы видели выше, все значения полисемантического слова (рубрики словарной статьи толкового словаря) считаются равнозначными, и соответствующие множители просто суммируются. Например, слово *голос* в словаре С. И. Ожегова определяется следующим образом (иллюстрации опущены): 1. Звучание, производимое колебанием связок, находящихся в горле. 2. Одна из двух или нескольких мелодий в музыкальной пьесе, партия в вокальном ансамбле. 3. Перец. Мнение, высказывание. 4. Право заявлять свое мнение при решении вопросов в государственных, общественных учреждениях, а также само такое выраженное мнение.

В соответствующем машинном массиве это слово со своими семантическими множителями записано так:  
*голос* зву колеб связк горл мелод муз пьес  
 парт вокал анс мнен высказ заяв право решен  
 вопр гос общес учрежд выра

Однако то же слово, выступающее в роли дескриптора, должно удовлетворять ряду требований<sup>38</sup>, основными среди которых являются его принадлежность к базовым понятиям языка и, по возможности, однозначность. Последнее, фактически невыполнимое для слов естественного языка, интерпретируется здесь как способность слова вы-

делять «главное значение», что и позволяет практически воспринимать его как однозначное. Это хорошо видно на примере рассмотренного слова *голос*, у которого второе значение выступает как естественное продолжение, развитие первого, и поэтому в кодах дескриптора составляющие оба значения семантические множители должны быть суммированы, тогда как третье и четвертое значения не связаны с первым прямо, а выводятся из него путем импликации (если результатом «звучания, производимого колебанием связок, находящихся в горле» являются слова, фразы, «высказывания», которые несут некоторую мысль, то эта мысль передает определенное «мнение». . .), и поэтому их компоненты опускаются. Но чтобы выполнить роль базового понятия, слово должно давать, хотя и концентрированный («однозначность»), но достаточно емкий спектр множителей, обеспечивающий максимальный охват оттенков значения, т. е. характеризоваться довольно значительным набором близких по семантике компонентов. Такое «распространение однозначности» достигается за счет расширения состава семантических множителей как «вглубь» (диахронический промер языка), с использованием словарей, охватывающих более отдаленные от сегодняшнего дня периоды развития русского литературного языка<sup>38</sup>, так и «вширь» — с использованием, в случае необходимости, компонентов от синонимов и от слов того же словообразовательного гнезда (синхрония и близкая диахрония). В итоге та же единица, но на другом уровне — в каталоге дескрипторов с их семантическими множителями характеризуется более компактным набором элементарных смыслов, сем:

ГОЛОС голос зву колеб связк горл мелод муз вокал кри  
реч пе произп манер част диапа искусс высо  
Другой пример — кодирование дескриптора ПОКУПАТЬ.

<sup>38</sup> Караплов Ю. И. Общая и русская идеография, с. 138—139.

<sup>39</sup> Схема использования словарей при кодировании дескрипторов была такой: за основу был взят словарь Ушакова (Толковый словарь русского языка, т. I—IV. М., 1934—1940), который пополнялся компонентами из дефиниций словаря Ожегова (С. И. Ожегов. Словарь русского языка. М., 1972); если количество семантических множителей, полученных таким путем, представлялось все же недостаточным, то использовался словарь Даля (В. И. Даля. Толковый словарь живого великорусского языка, т. I—IV. М., 1955); в редких случаях, особенно при кодировании отрицательных дескрипторов, приходилось обращаться к словарю синонимов (Словарь синонимов русского языка, т. I—II. Л., 1970—1971).

Словарь Ушакова дает отсылку к совершенному виду *купить*, который толкуется «приобрести за деньги», т. е. получаем всего четыре семантических множителя: покупка приобретение денег. Обращение к словарю Ожегова позволяет добавить компонент собс («приобрести в собственность»); далее, из дефиниций слов, производных от того же корня, по словарю Ушакова находим дополнительные множители: *покупатель* → торг, *покупательный* → товар клие (в словаре Ожегова этих множителей нет). В итоге указанный дескриптор характеризуется восемью компонентами, что существенно расширяет возможности наполнения соответствующей словарной статьи в тезаурусе.

Наибольшие затруднения, как отмечалось выше, вызывает кодирование отрицательных дескрипторов. Так, *непроточный* отсутствует среди входных слов и в словаре Ожегова, и в словаре Ушакова; нет этого слова и в академическом словаре синонимов. Его кодирование осуществлялось по словарю Даля: *непроточный* — непроточ засто стоя невпад озер пруд теря песок болот нетеч вод. Как видим, особенность далевского определения в том, что в нем есть элементы семантического поля, включение которых в перечень множителей данного дескриптора соответствует нашему принципу максимального расширения семантического спектра дескрипторов при их кодировании.

Использование словаря синонимов проиллюстрируем следующим примером. *Бесполезный* определяется в словаре Ожегова как «не приносящий пользы, напрасный». После трансформации (*не приносящий пользы* → неполез) и включения самого кодируемого слова в число его множителей семантическая запись его выглядит так: бесполе неполез напрас. Для дескриптора, каким является рассматриваемое слово, число множителей представляется явно недостаточным, поскольку не позволяет при формировании поля учесть оттенки значений, особенности ситуативного и контекстного употребления слов, входящих в поле. Расширение числа множителей достигается за счет обращения к синонимическому словарю, который дает следующие добавления: ненуж тщет безусп бесплод бессмы.

Расширение набора семантических множителей в соответствии с нашей задачей построения общеязыкового тезауруса осуществляется, естественно, в пределах филологических словарей, без выхода в терминологические области. Понятно, что кодирование, например, дескриптора

ГОСУДАРСТВО по политическому или философскому словарю, дескриптора ГОЛОС — по музыкальному, а КОНФИСКАЦИЯ — по юридическому заметно увеличило бы число компонентов каждого из них, но в то же время привело бы и к искажению общей картины в итоговых статьях тезауруса, поскольку наш исходный словник содержит минимум терминологической лексики и рассчитан не на специальные, а на так называемые «языковые понятия».

В результате кодирования дескрипторов в машине формируется следующий массив — каталог дескрипторов с семантическими множителями:

Номер	Шифр	Число						
			множ.					
54	12272	14	РАСТВОР					
			раств	разводи	размеш	разжиж	жидк	
			равн	смес	однород	соотн	вод	
			тверд	газоо	веществ	смеш	разой	
759	12268	5	РАССТОЯ-					
			НИЕ					
			рассто	пространс	раздел	пут	промежут	
877	12264	18	РАСПУТЬ-					
			ВАТЬ					
			распуть	размот	поряд	неспут	доиск	
			сущн	ясн	разреш	освоб	разъезд	
			раздел	разобр	сня	развяз	поля	
			непут	спут	пут			
993	12266	5	РАССЕЯН-					
			НОСТЬ					
			рассеян	невним	несосред	ослаб	незамет	
1022	12270	14	РАССУЖ-					
			ДАТЬ					
			рассу	умозак	последов	излаг	сужд	
			бесе	мысл	логи	форм	обсу	
			излож	высказ	возра	говор		
1146	12262	5	РАСПУТ-					
			СТВО					
			распу	образжи	разгу	развр	безир	

Приведенные фрагменты двух массивов — каталога слов с семантическими множителями и каталога дескрипторов со своими множителями — выбраны таким образом, что у читателя возникает возможность еще раз сравнить результаты кодирования одних и тех же единиц, но принадлежащих разным массивам: распутьство—РАСПУТСТВО, распутывать—РАСПУТЬВАТЬ, рассеянность—РАССЕЯННОСТЬ, расстояние—РАССТОЯНИЕ, раствор—РАСТВОР.

Параллельно формированию двух последних массивов шло обобщение показателей семантических связей, т. е. создание единого для дескрипторов и слов банка семантических множителей (квазиоснов), который в окончательном виде включает 8750 единиц:

Номер	Частота сл	Частота дес	Семантич. множитель	Номер	Частота сл	Частота дес	Семантич. множитель
1001	3	4	глот	1046	0	1	горлы
1002	2	0	глу	1047	1	1	горн
1003	13	9	глуб	1048	1	0	горни
1004	2	1	глубоком	1049	2	1	горнопр
1005	1	0	глум	1050	1	1	горо
1006	4	1	глуп	1051	24	6	город
1007	2	2	глух	1052	2	1	городк
1008	1	1	гляд	1053	1	0	горс
1009	1	1	глянц	1054	2	2	горта
1010	9	1	гна	1055	1	2	горч
1011	5	3	гнев	1056	8	8	горь
1012	1	1	гнез	1057	4	3	горьк
1013	1	0	гнет	1058	8	4	горю
1014	6	2	гни	1059	18	4	горяч
1015	2	2	гио	1060	1	1	горячи
1016	5	2	гну	1061	68	30	гос
1017	O	1	гнусн	1062	15	5	госп
1018	26	4	говор	1063	0	1	господ
1019	0	1	гоя	1064	3	1	гост
1020	39	11	год	1065	4	0	гостин
1021	16	2	годн	1066	2	2	госуд
1022	5	1	годовщ	1067	26	8	готов
1023	1	3	гол	1068	3	1	граб
1024	2	2	годен	1069	1	0	грабл
1025	31	13	голов	1070	1	1	грав
1026	3	1	головк	1071	2	3	град
1027	O	1	головокр	1072	3	1	граду
1028	3	2	голод	1073	11	6	граж
1029	17	4	голос	1074	1	0	грам
1030	0	1	голосем	1075	40	2	грамма
1031	7	1	голосов	1076	1	0	гранди
1032	1	1	голу	1077	22	11	грани
1033	3	1	голуб	1078	7	4	грань
1034	5	1	гон	1079	3	0	графа
1035	1	1	гонк	1080	6	3	графи
1036	2	3	гончар	1081	1	1	графит
1037	27	13	гор	1082	0	1	грач
1038	3	1	горб	1083	2	2	гре
1039	7	3	горд	1084	2	1	греб
1040	18	3	горе	1085	2	1	гребн
1041	3	1	горес	1086	1	2	грез
1042	1	1	гореч	1087	1	0	грек
1043	14	7	гориз	1088	1	0	грэм
1044	0	1	горк	1089	1	1	грес
1045	6	7	горл	1090	9	2	гриб

## § 7. АСИМПТОТИЧЕСКАЯ МОДЕЛЬ ТЕЗАУРУСА: РЕАЛИЗАЦИЯ И ИНТЕРПРЕТАЦИЯ

На первый взгляд кажется, что фигуры могут двигаться и только, но в действительности это не так. Как человек может суетиться, но не работать, так и фигура может двигаться, но не действовать. Движение — это необходимое условие, действие — достаточное условие, чтобы оценить полезную роль фигуры в игре. Поэтому все программы, основанные на полном переборе, не могут привести к достижению цели: в основном они занимаются суетой сует.

*M. M. Ботвинник. О кибернетической цели игры*

### Системные отношения между множителями и статистика

Каталог семантических множителей по своему объему для нашей задачи явно избыточен и сделан с расчетом на некоторую перспективу. Избыточность его в том, что, согласно феноменологической модели тезауруса, фактически работающими при распределении являются семантические связи, заданные дескрипторами: именно на основе общности множителя (а на уровне феноменологической модели, напомним, все множители, за исключением нулевых, считаются имеющими одинаковую значимость, одинаковый вес) у дескриптора и какого-то слова решается вопрос о наличии семантической связи между ними, а значит, о принадлежности слова к семантическому полю данного дескриптора. В этих условиях все множители, которые встречаются только на массиве словаря, т. е. в каталоге слов с семантическими множителями, и не встречаются в дефинициях дескрипторов, оказываются как бы излишними, неработающими. Последнее не означает тем не менее, что такие множители вообще бесполезны: они будут необходимы при решении важных дополнительных задач, сопутствующих построению русского тезауруса.

Во-первых, они могут понадобиться при уточнении и изменении состава дескрипторов и расширении словаря. Во-вторых, с помощью этих множителей решается задача установления семантических связей и измерения их интенсивности в готовой уже статье тезауруса (семантическом поле), но связей не между дескриптором и составом статьи, а только включенных в статью слов между собой. Наконец, в-третьих, при решении некоторых других, помимо идеографической классификации, семантических задач, например построения ассоциативного словаря, выделения и анализа гнезда слов с одним или несколькими

(двумя, тремя и т. д.) общими семантическими множителями.

Но все же в данный момент, для задачи получения идеографической части тезауруса главную роль играет множество семантических множителей, относящихся к дескрипторам. Можно полагать, что и среди них окажутся неработающие (как увидим ниже, это или малочастотные, или, наоборот, слишком употребительные, сверхвысокочастотные элементы), что должно свидетельствовать о неудачном выборе дескриптора или — с другой стороны — о несоответствии словарника некоторым намеченным дескрипторам, его неспособности с достаточной полнотой охарактеризовать сферу того или иного понятия. Поэтому статистический анализ единиц этого массива оказывается весьма информативным<sup>40</sup>: частота семантических множителей дескрипторов на своем собственном — дескрипторном массиве, так же, как частота множителей слов па словнике дает возможность оценить семантический вес компонентов и установить их роль (практически каждого из них, или любого) в характеристике семантического ядра русского языка; далее частота множителей дескрипторов на массиве слов позволяет сделать выводы как о степени полноты и точности кодирования дескрипторов, так и шире — об адекватности их выбора. Кроме того, частотный анализ этого массива приводит к уточнению состава и установлению окончательного числа семантических множителей, что, во-первых, создает предпосылки для более глубокого теоретического осмысления их статуса и некоторых практических выводов, а во-вторых — для автоматизации самого процесса кодирования при необходимости увеличить состав словарника до 50 или 100 тыс. слов.

В идеале уточненный состав дескрипторных множителей содержит тот минимум семантической информации, который необходим для характеристики всей русской лексики, иными словами, представляет собой семантическое ядро русского языка, которое при своем развертывании порождает словарь-минимум, обеспечивающий нормальную коммуникацию на русском языке. Развертывание в данном случае совпадает с декодированием<sup>41</sup> и предполагает приписывание каждой квазиоснове всех тех слов (гнезда), семантику которых она представляет: хром → хромать, хрометь, хромец, хромой, хромоногий, хромота; монах → монах, монастырь, монашество, монаший и т. п.

По своему статусу и по принципу использования квазиоснова представляет собой некоторый элемент смысла, семантический множитель, т. е. имеет строго одно значение, однако развертываться и коррелировать она может с целой группой слов, лексическим гнездом, в котором каждое из составляющих обязательно содержит данный элемент смысла. В процессе кодирования, как было показано выше, полисемии как бы не существовало, она была снята самой процедурой за счет того, что при семантической обработке дескрипторов происходило развитие, расширение лишь одного, главного значения, а при кодировании слов множители разных значений недифференцированно суммировались, т. е. осуществлялся охват семантики слова во всей совокупности его проявлений<sup>40</sup>.

Такое искусственное, методическое устранение (усреднение) многозначности привело к тому, что при развертывании массива семантических множителей на ее месте выступают словообразовательные отношения, восполняющие, очевидно, тот дефицит расчлененности и детализации в замкнутой системе, который образовался в результате интегрирования значений полисемантического слова и снятия многозначности. Это явление компенсации средств лексической семантики морфологическими, а точнее, словообразовательными средствами языка, с одной стороны, отчетливо показывает тесную связь, проницаемость и взаимопроникновение уровней языковой структуры, а с другой, свидетельствует о наличии во всякой системе некоторого критического показателя расчлененности, диф-

---

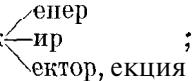
<sup>40</sup> См.: Караплов Ю. Н. Частотный словарь семантических множителей русского языка.

<sup>41</sup> Подробнее о приемах декодирования см. там же, с. 14–17.

<sup>42</sup> Этот методический прием не только оправдывается конструктивной целесообразностью, поскольку облегчает формализацию процедур, но и имеет содержательное обоснование. Во-первых, из лексикографической практики известно, что провести границу как между отдельными значениями многозначного слова, так и между значением и оттенком подчас бывает трудно, и поэтому такая граница носит, как правило, условный характер. Во-вторых, как показывают результаты психолингвистических экспериментов, слово реализуется в речи всегда во всей совокупности своих значений одновременно, но при этом доминирующим, актуальным, выдвинутым на первый план в каждом употреблении бывает какое-то одно значение, обусловленное ситуацией, контекстом, коммуникативным намерением говорящего при неявном, потенциальном участии всех других.

ференцированности, нарушение которого может привести, очевидно, к разрушению самой системности.

Требование строгой однозначности не только предполагает отсутствие полисемии, но и исключает, естественно, омонимию квазиоснов. Это достигается тем, что при кодировании лексических омонимов квазиосновы конструируются разными; например, гнезду слов *волнение* (душевное), *волновать(ся)*, *волнувший* приписывается в качестве квазиосновы искусственный сегмент *волне*, тогда как однокорневое гнездо — *волнение* (на море), *волна*, *волноваться* (о воде) — кодируется сегментом *волн*; аналогично *доля* (судьба) → *доль*, а *доля* (часть) → *доля* и т. п. Однако проблема омонимии, а скорее, «псевдоомонимии» квазиоснов, может все же возникнуть иногда при декодировании, когда дешифровщику, например, понадобится восстановить по квазиоснове исходное слово без обращения к картотеке семантических множителей. Дело в том, что в некоторых (впрочем, немногочисленных) случаях требованию минимизации числа букв в квазиоснове отдавалось предпочтение перед требованием однозначности декодирования (см. выше алгоритм кодирования). Поэтому множитель *ска*, например, нельзя декодировать без картотеки, поскольку он с равным успехом может развертываться и как *скакать*, *скакоч* (т. е. до морфонемы *к/ч*), и как *сказать*, *скажут* (морфонема *з/ж*). Множитель инж создает при декодировании впечатление омонимии, так как до-

пускает три возможности развертывания: *инж*  ;  
 ;  
 ;

выбор такой квазиосновы именно для слова *инженер* обусловлен тем, что *инжир* для русского языка является почти экзотизмом, а *инжектор* представляет собой узкоспециальный термин, так что вероятность встретить в общелитературном языке два последних слова гораздо меньшая, чем слово *инженер*. Естественно, что в случае необходимости оба этих слова получат в каталоге семантических множителей свои собственные идентификаторы-квазиосновы *инжи* и *инжек*.

Требование строгой однозначности квазиоснов, т. е. их статус семантических множителей, практически исключает возможность синонимии. Однако при декодировании и последующем развертывании квазиоснов в лексические гнезда синонимические отношения могут снова возникать уже на уровне слов, проявляясь в виде эквивалентности,

близости значений элементов двух разных лексических гнезд. Источников такой синонимии три: во-первых, это обычные синонимы, включенные в исходный словарь (таких синонимов немного); во-вторых, синонимы, использованные в дефинициях разных словарей при толковании одного и того же дескриптора; наконец, в-третьих, синонимы иногда возникают как результат отрицательных трансформаций при кодировании. Например, при кодировании дескриптора КОМПАС, в набор его семантических множителей попали указ (из *указывать* (*о стрелке*) — по словарю Ожегова) и показ (тоже о стрелке компаса, но по словарю Ушакова), опред (из *определять страны света* — по Ожегову) и распоз (для *распознавания стран света* — по Ушакову). В итоге, при развертывании полученных квазиоснов в их лексических гнездах появляются синонимические пары: *указывать* — *показывать* и *определять* — *распознавать*.

В результате трансформаций чаще всего, как было сказано, образуются синонимы-отрицания: в каталоге уже имелся семантический множитель бесцер (из *бесцеремонно*, см., например, дефиницию слова *примазаться* и др.), когда путем соединения с отрицанием был получен практически синонимичный ему нецерем (из *не церемонясь*, см., например, дефиницию слова *попросту*).

Наряду с синонимами и, вероятно, даже шире синонимии представлены на массиве квазиоснов антонимические отношения, что объясняется их универсальностью и известной логической простотой. Если синонимы и омонимы запрещены среди дескрипторов уже по самому статусу последних, то антонимия представляет собой, пожалуй, очень распространенное отношение даже и в классификационной схеме, т. е. на уровне дескрипторных областей и фабулы тезауруса: СИЛА — СЛАБОСТЬ (соответственно в семантических множителях — сил ~ слаб, РЕЛИГИЯ ~ АТЕИЗМ (соотв. религ ~ атеи), РАДОСТЬ ~ ГОРЕ (рад ~ горь) и т. п. Далее, в исходный словарь, репрезентирующий, по определению, весь словарь русского языка, вошло некоторое количество исконных антонимов, т. е. антонимов на уровне слов, которые тоже дали антонимичные квазиосновы: *порядок* ~ *беспорядок* (поряд — беспор), *гаснуть* ~ *загораться* (гас ~ загоре), *молодой* ~ *старый* (молод — стар) и т. д. Но антонимия, как оказалось, широко используется и в дефинициях, на уровне семантических множителей, поскольку противопоставление — один из

самых распространенных логических приемов определения действия, объекта или его свойств; правда, словарная дефиниция при этом приобретает иногда противоречивый характер: *качать* . . . 4. Вытягивать, нагнетать насосом (антонимичные семантические множители — вытя  $\sim$  нагне); *объяснение*. . . 2. Письменное или устное изложение в оправдание чего-н. (антонимичные множители — пис  $\sim$  устн); *оправдать*. . . 2. Признать. . . не заслуживающим наказания, кары. 3. Признать. . . заслуживающим снисходительного отношения (антонимичные множители — незаслуж  $\sim$  заслуж); *масло* 1. Жидкое или твердое жировое вещество. . . (антонимичные множители — жидк  $\sim$  тверд); *мяч* — сплошной или полый внутри шар из какого-н. упругого материала, отскакивающий от твердой поверхности. . . (антонимичные компоненты — сплош  $\sim$  полый, упру  $\sim$  тверд); *соль*. 1. . . продукт полного или частичного замещения водорода кислоты металлом (антонимичные множители — полн — частич). Наконец, источником антонимии квазиоснов, как и в случае синонимических отношений, могут быть отрицательные трансформации, осуществлявшиеся в соответствии с алгоритмом кодирования: лишенный свободы (при кодировании слова *plen*)  $\rightarrow$  несвоб (антоним к квазиоснове своб), освободиться от грязи (*очищать*)  $\rightarrow$  негряз (антоним к гряз), избегающий движений (*неподвижны*)  $\rightarrow$  педви (антоним к дви), прекратить общение (*отставать*)  $\rightarrow$  необща (антоним к обща или общен), утратить чувствительность, гибкость (*онеметь*)  $\rightarrow$  нечув, негибк (антонимы соответственно к семантическим множителям чувст и гибк).

Таким образом, предварительный анализ некоторых свойств каталога семантических множителей показывает, что этому множеству присущи системные отношения словообразовательного характера, синонимии и антонимии, иными словами, массив множителей обладает почти всеми (за исключением полисемии) свойствами самостоятельного лексикона. Для превращения этого массива в реальный перечень слов к составляющим его единицам должна быть применена операция декодирования и развертывания. Ее целесообразно осуществить прежде всего для семантических множителей, относящихся к дескрипторам. Полученный в результате такого развертывания набор слов будет несколько расходиться с исходным словарником нашего тезауруса — эмпирически отобранным словарем-минимумом, но значительную часть двух словников соста-

вит все же общая лексика. Невозможность полного совпадения объясняется, во-первых, фундаментальной асимметрией всякого словаря и действием в нем закона притирания смысла<sup>43</sup>, из-за чего словарь и не поддается механическому «переворачиванию», а во-вторых, разными принципами отбора исходного материала: дескрипторы комплектовались с таким расчетом, чтобы с максимальной полнотой отразить идеографическую классификацию, тогда как отбор слов в словарь должен в первую очередь обеспечить удовлетворительную коммуникацию. Таким образом, два сравниваемых словарника находятся в отношении взаимодополнительности. Суммируя их, мы получим в некотором смысле идеальный набор единиц, который характеризуется всеми свойствами словаря (ср. выше) и в то же время апробирован идеографической классификацией. Такой научно обоснованный словарник — важный промежуточный результат нашей работы, поскольку до сих пор не существует сколько-нибудь последовательных и удовлетворительных с точки зрения соответствия целям словаря принципов комплектования его словарника. Полученный нами словарник может быть использован как базовый словарь для преподавания русского языка иноязычным пользователям, для составления малых и средних национально-русских, русско-национальных и толковых словарей русского языка.

Однако возможности использования каталога семантических множителей этим не ограничиваются. Если осуществить теперь обратный переход — от полученного базового словарника к множителям дескрипторов и пополнить состав последних недостающими в списке множителями самих слов из словаря (уже не обращаясь к дефинициям и не прибегая к семантической записи составляющих их компонентов, а закодировав только каждое входное слово), то такой список дескрипторных множителей и будет тем константным семантическим ядром русского языка, которое необходимо и достаточно для характеристики всей русской лексики.

Теперь, имея в своем распоряжении такие данные, исследователь может упростить и ускорить (а в перспективе, возможно, и автоматизировать) процедуру кодирования новых порций слов, практически при любом расширении словарника, за счет включения в этот процесс машины. Поскольку расширенный каталог дескрипторных

---

<sup>43</sup> Караполов Ю. П. Общая и русская идеография, с. 80—94.

множителей является константным и расти не должен, задача кодировщика будет состоять лишь в сравнении компонентов (гипотетических множителей) добавляемого массива слов с множителем постоянного каталога, содержащегося в памяти машины. Последний, являясь семантическим ядром, по самому своему статусу должен содержать для любого слова русского языка хотя бы один множитель, характеризующий его значение; и наоборот, теоретически не существует такого слова в русском языке, которое нельзя было бы привязать через тот или иной компонент его словарной дефиниции к данному множеству семантических множителей, а значит — к одному или нескольким дескрипторам.

Это свойство уточненного и расширенного каталога семантических множителей позволяет рассматривать его как универсальную базу для составления не только филологических, но и специальных словарей.

### **Эвристические приемы конструирования нового лексикографического объекта**

После того как сформированы три массива: каталог дескрипторов, словарник и каталог семантических множителей — можно приступить к решению основной задачи — получению дескрипторных статей, т. е. идеографической части тезауруса. Одновременно с этой решаются некоторые дополнительные задачи (см. здесь же, ниже), имеющие значение для уточнения структуры тезауруса и, шире, вообще для изучения русской лексической семантики.

При реализации феноменологической модели работают два множества — дескрипторы со своими семантическими множителями и слова с соответствующими множителями; цель программы — распределить слова по дескрипторам. Основанием для включения слова (например, *поблажка*) в словарную статью дескриптора (например, ПРОЩАТЬ) служит тождественность, т. е. побуквенное совпадение хотя бы одного множителя в семантических кодах того и другого (в данном случае снисх). В результате такого распределения вокруг каждого дескриптора образуется некоторое семантическое поле, составленное из относящихся к нему слов, из слов, находящихся с данным дескриптором в семантической связи. Естественно при этом, что одно слово попадает в несколько семантических полей, и с одним дескриптором то или иное слово может быть свя-

зано не единственным, а несколькими общими семантическими множителями. Полученные таким образом дескрипторные статьи и представляют собой идеографическую основу русского тезауруса.

Для оценки эффективности феноменологической модели тезауруса было осуществлено пробное распределение, правда, не слов по дескрипторам, а наоборот, дескрипторов по словам<sup>44</sup>. Иначе говоря, для каждого из 400 слов выбранного фрагмента словаря были перечислены все дескрипторы, в статьи которых это слово должно войти по правилам феноменологической модели. Согласно условиям задачи, в выходных данных содержалась также информация о тех общих в каждой паре семантических множителях, которые послужили основанием для отнесения слова к соответствующему дескриптору, их количество для той или иной пары и частотности каждого из них на массиве дескрипторов. Результат распределения выглядел следующим образом (слово, которое подлежит распределению, подчеркнуто; после каждого из относящихся к нему дескрипторов приведены в скобках общие для данной пары семантические множители; частота множителя указана один раз при первом его появлении):

аптека		уступка	
фармакология	(апт 1 лекар 4)	убыток	
никель		моноополия	
литъе		деньги	
виноград		продавать	
шить	(изгот 14)	торговля	
шелк		магазин	
ткань		дом	
кирпич		больница	
производство		часть	
скульптура		полномочия	
инструмент		должностной	
ювелирный	(изгот прода 9)	служба	
работать		милиция	
фотография	(изгот учрежд 19)	таможня	
малина		опека	
медицина	(лекар)	суд	
впрыскивание		вклад	
		бухгалтерия	
		учреждение	
		сберкасса	

<sup>44</sup> Вообще говоря, эти отношения (дескриптор—слово И слово—дескриптор) не обладают полной обратимостью, однако для предварительных интуитивных оценок степени связности слов со своим дескриптором этой разницей можно пренебречь.

банк		стебель	
почта		смородина	(раст слад)
арбуз		смола	
тыквенные	(арбуз 1 плод 20	палеонтология	
раст 67 тыкв 1)	(однолет 1 раст)	кожа	
трава	(плод раст)	волосы	
дуб	(плод раст слад 13)	рост	
виноград	(плод раст)	чай	
бобовые	(плод раст)	кофе	
вишня	(плод)	лен	
зерно	(плод раст)	орнамент	
овощные	(плод)	паразиты	
груша	(плод слад)	пупок	
яблоня	(плод)	возраст	
фрукты	(плод слад)	посадить	
семя	(плод раст)	возделывать	
слива	(плод)	вянуть	
оплодотворение		произрастать	
сад		посев	
плодородие	[плод]	жатва	
неурожай		прививка	
урожай		земледелие	
спелый	] (плод)	гибридизация	(раст)
плодиться		торт	
масло		кондитерский	
жир		мед	
роса		сладкий	
клубень		сахар	
пальма		аргументировать	
листва		[ (аргум 1 кора	довод 3
деревянный		аргументировать	доказ 5
ботаника		основ 13	привод 3)
бурьян		безрассудство 1 (довод)	
почка		отговаривать J	
ветка		убеждение ] (доказ)	
вьющиеся		документ	
злаковые		опровержение	
капуста		учть ] (доказ)	
лес		каркас	
водоросль		фундамент	
земляника		существенный	
крыжовник		правило	
куст	] (раст)	причина	
корневище		основа	
гриб (раст)		начало	
малина (раст	слад)	нижний	
мох		цилиндрический	
луковица		пирамида	
луг		колесо	
роза		восстанавливать	
тростник		(привод)	
цветок			

Примеры можно продолжить, но в этом, очевидно, нет необходимости, поскольку уже из знакомства с приведенными иллюстрациями читателю ясно, что такое распределение зачастую не согласуется с интуитивным представлением о смысловой связности слова и перечисленных под ним дескрипторов, обладающих общими с этим словом семантическими множителями. Кроме того, и внутри групп самих дескрипторов, объединенных одним семантическим множителем, не всегда обнаруживается смысловая общность. Вместе с тем нельзя сказать, чтобы полученное предварительное распределение было совершенно бессмысленным. На деле оно дает интересный материал для изучения семантики отдельных слов, для выделения и анализа лексических групп разной степени семантической плотности, их внутреннего смыслового единства, для систематизации и улучшения словарных определений. Однако по отношению к основной задаче вывод должен быть такой: феноменологическая модель и положенный в ее основу критерий связности оказываются недостаточным, не адекватным приемом получения идеографической классификации.

Для разрешения выявившихся трудностей и построения модели тезауруса, в большей степени приближающейся к точному отражению действующих в нем семантических закономерностей, были осуществлены поиски таких характеристик (параметров, критериев), которые позволили бы уточнить первоначально принятые принципы распределения слов по дескрипторам. Новая, уточненная модель семантической связности получила название асимптотической, что должно подчеркнуть ее несовершенный, но поддающийся улучшению характер, а также ее направленность ко все более полному учету реальных семантических отношений в тезаурусе.

Естественно, что показатели, углубляющие оценку степени связности дескриптора со словом, выбирались из наличной информации (см. приведенный выше фрагмент распределения). Каждая из перечисленных далее характеристик испытывалась на некотором множестве слов и соответствующих дескрипторов, и результаты испытания каждый раз сравнивались с интуитивным представлением носителя языка (автора) об обоснованности включения исходного слова в ту или иную дескрипторную статью, а также с результатами испытаний других показателей с целью определения наиболее эффективного из них.

Среди подвергшихся подобным испытаниям и оценкам были следующие критерии:

- 1) показателем связности считается не один общий семантический множитель, а два и более;
- 2) слово и дескриптор считаются связанными, если в их семантических кодах совпадают как минимум два семантических множителя, причем частота<sup>45</sup> одного из них не превышает величины 2 (затем то же — для частоты 3, 4, 5);
- 3) показателем связности является сумма частот одинаковых для дескриптора и слова семантических множителей; эта сумма частот не превышает 20 (затем то же — для суммы в 18, 15, 14 употреблений множителей на массиве дескрипторов);
- 4) показателем связности является совпадение хотя бы одного семантического множителя, частота которого не превышает некоторого порога (4, 5, 6, 7, 8, 9 употреблений).

Для каждого из охарактеризованных случаев строились специальные таблицы, позволяющие наглядно представить результат распределения при принятии того или иного критерия и сравнить друг с другом результаты разных распределений. Эти таблицы ввиду их большого объема и незначительной информативности здесь не приводятся, за исключением последней (см. ниже). Читателю предоставляется самому оценить, хотя бы в первом приближении — на нескольких примерах, каждый из остальных критериев с использованием при необходимости частотного словаря семантических множителей.

Таким эмпирическим путем, методом проб и ошибок было установлено, что число совпадений, т. е. количество общих семантических множителей у слова и дескриптора, не может служить надежным показателем их смысловой связи. Решающей характеристикой оказалась частотность совпадающих множителей. Показатель частотности был назван селективным параметром асимптотической модели тезауруса. Для данного объема материала (количества слов в словаре, числа дескрипторов и совокупного числа их семантических множителей) этот селективный параметр, т. е. пороговая частота множителя, должен быть менее 7. При употребительности семантического множителя, превышающей критическое число 6, он теряет свою релевантность, перестает быть показателем семан-

тической связи. Понятно, что характер множителя не меняется скачком при достижении его частотностью критической величины, а изменяется в зоне последней асимптотически, поэтому возможны и некоторые отклонения от установленного правила — по ту или другую сторону указанной границы. Такие отклонения читатель обнаружит и в табл. 3; комментарии к ним по возможности даны в последней графе.

Сравнивая заполнение клеток таблицы дескрипторами, читатель увидит, что их количество в одной строке (т. е. для одного слова) меньше их числа для того же слова в приведенных выше иллюстрациях предварительного распределения. Объяснение этого факта состоит в том, что четвертая логическая возможность («дескриптор и слово не имеют общего семантического множителя — с частотой ниже пороговой, и одновременно носитель языка не фиксирует между ними интуитивно ощущаемой смысловой соотнесенности») ввиду тривиальности ее результата в таблице не учитывается. А как легко заметить, подавляющая часть дескрипторов в группах предварительного распределения характеризуется именно таким отношением с исходным словом, т. е. в соответствии с правилами асимптотической модели исключается из рассмотрения. Таким образом, таблица служит обоснованием выбора именно данной величины селективного параметра. Цифры после дескриптора указывают частотность множителей, общих у него с исходным словом, а здесь возможна и минимальная употребительность — в definicции только одного дескриптора. Отметим при этом, что частотность, равная единице, вовсе не означает ошибки в расчетах или противоречивости в рассуждениях. Ведь, казалось бы, если данный множитель является общим для слова и хотя бы одного дескриптора, то его частотность уже не может быть равной единице, так как он употребляется минимум дважды. Однако напомним, что релевантной для асимптотической модели признана частотность лишь на одном массиве — массиве семантических множителей, относящихся к дескрипторам, а число цифр соответствует числу

---

<sup>45</sup> Здесь и далее по изложенным в § 6 соображениям в качестве релевантной принимается частота семантических множителей на массиве дескрипторов (см. также упомянутый «Частотный словарь»).

Этих совпадающих множителей. Однаковые слова в первой и второй графах означают, что данное слово является одновременно и дескриптором. В остальном структура таблицы кажется достаточно прозрачной и потому не требующей дополнительных комментариев.

Из анализа табл. 3 можно сделать два принципиально важных вывода. Во-первых, установленная величина селективного параметра (частотность  $\leq 6$ ) обеспечивает максимальный учет асимптотической моделью связности всех дескрипторов, имеющих хотя бы отдаленное касательство к исходному слову (или, наоборот, всех слов, которые, пусть даже в отдельном употреблении, обнаруживают смысловую близость с данным дескриптором). Это означает, что при распределении практически ничего не пропускается: внимательное прочтение графы 3 в таблице показывает, что ее заполнение (т. е. те дескрипторы, которые как бы не учтены, пропущены, но которые потенциально могут проявить семантическую связность с исходным словом) почти всегда осуществляется с некоторой натяжкой, включение дескрипторов в эту графу всегда не безусловно, проблематично. Во-вторых, как видно из рассмотрения графы 4, принятый критерий связности придает асимптотической модели также и известный запас надежности, поскольку обуславливает отнесение к типу семантически связанных с данным словом некоторого избыточного числа дескрипторов. При более глубоком анализе оказывается, что далеко не все из них характеризуются отсутствием интуитивно устанавливаемой семантической связности с исходным словом. Это становится очевидным при переменах местами, так сказать, «стимула и реакции»: если в качестве исходного принять не слово, а дескриптор (как реально и обстоит дело при конструировании нашего тезауруса), то включенность многих слов в соответствующие дескрипторные статьи, особенно с учетом возможного словарного наполнения последних, представляется достаточно обоснованной. Ср., например: осада — *артиллерия*, вогнутый — *арка*, голос — *артист*, башня — *архитектор*, сомнение — *атеизм* и т. п. Понятно, что это не исключает некоторого (но уже небольшого) процента ошибок в распределении, который потребует — для своего устранения — постмашинного редактирования словаря.

Дополнительные задачи, решаемые в ходе распределения слов по дескрипторам, а также по итогам этого рас-

пределения, распадаются на две группы. Первый круг задач, затрагивая вопросы организации лексического материала в тезаурусе, вопросы его структуры, связан с поисками путей наилучшего представления результатов на выходе из машины. Во-первых, это определение объема данных (помимо самих слов) и нахождение такого способа их компоновки в дескрипторной статье тезауруса (т. е. прямо в выходе «на печать»), который обладал бы максимальной информативностью и оптимально отвечал бы целям словаря; во-вторых, встает вопрос об упорядочении самих слов, составляющих дескрипторную статью, т. е. о порядке следования, порядке расположения слов в семантическом поле каждого дескриптора; наконец, в-третьих, на этом этапе возникает необходимость вернуться к уточнению классификационной схемы, т. е. рассмотреть возможности упорядочения дескрипторов, построения фабулы — уже по результатам самой идеографической классификации.

Что касается первой задачи, то она не имеет пока окончательного решения, но предварительные обсуждения позволяют отдать предпочтение следующему варианту:

- 1) для слова, включенного в дескрипторную статью, будут указаны все общие у него с данным дескриптором семантические множители;
- 2) каждый множитель снабжается индексом его частотности на массиве дескрипторов;
- 3) для каждого слова будут вычислены и приведены тут же коэффициенты силы связи его с дескриптором по формулам, учитывающим критерий пороговой частоты семантических множителей (селективный параметр).

Вторая задача имеет содержательный смысл, и пути возможных ее решений подробно рассмотрены в одной упоминавшейся нашей работе<sup>46</sup>. Наиболее приемлемым из приведенных там вариантов — и с точки зрения выполнимости его на машине, и с точки зрения информативности —казалось вначале алфавитное упорядочение с указанием возле каждого слова того множителя (тех множителей), который является показателем семантической связи с основанием для включения слова в статью. Однако такой порядок не отражает внутренних закономерностей организации семантического поля и не обладает той наглядностью, которая позволила бы без дополнительного

---

<sup>46</sup> Караулов Ю. Н. Общая и русская идеография, с. 229—241.

Таблица 3

<b>Слово</b>	Дескриптор			Оценка возможного источника ошибки
	Есть интуитивная отнесенность слова к дескриптору и показатель связи	Есть интуитивная отнесенность, <b>нет</b> показателя связи	Нет интуитивной отнесенности, есть показатель связи	
1	2	3	4	5
арбуз	тыквенный 1, 20, 67, 1	спелый 20 сладкий 13	трава 1, 67	Непоследовательное использование в дефинициях толковых словарей семантич. множ. однолет
аргументировать	аргументировать 1, 3, 5, 13, 3 убеждение 5 опровержение 5 отговаривать 3	причина 13 основа 13	безрассудство 3 документ 5 учить 5 колесо 3 восстанавливать 3	1) Некорректность дефиниций дескрипторов в графе 4, включающих множ. довод 3, доказ 5, привод 3; 2) включение дескрипторов ПРИЧИНА и ОСНОВА в графу 3 представляется проблематичным
арендовать	арендовать 1, 2, 6, 1 кредит 6 долги 6 брать 2, 6		выбор 6 заменять 6 буржуазный 4 заработка 4 такси 4	Непоследовательное применение правил кодирования (множ. ая 6 и брать 5 являются нулевыми по определению)

арестовать	заточение 6		заплата 3 наполнять 3 обязательный 3	Неточность кодирования слова <i>арестовать</i> (множ. налож 3 избыточен)
арифметика	счет 4, 47, 2 цифра 3, 2 вычитание 4, 47, 12 сложение 4, 47, 12 умножение 4, 47, 12	деление 47,12 решать 47 действие 47	дата 3 жест 3	<i>Вычитание, сложение, умножение</i> определяются через семантич. множ. ариф, а <i>деление</i> не имеет такого компонента в дефиниции; 2) множитель обозн 3, общий с дескрипторами ДАТА и ЖЕСТ, является избыточным в дефиниции слова <i>арифметика</i>
арка	арка 1, 1, 5, 13, 2, 3, 1, 22, 12, 36 проем 3, 12 выпуклый 5 стена 2, 22, 12 дверь 3, 12	строение 22	сгибать 5, 36 вогнутый 5 бровь 5	Вероятно, множитель дугообр 5 в дефиниции дескриптора БРОВЬ избыточен
армия	армия 3, 9, 6, 30, 1, 2, 15, 1, 2 военный 3, 9, 30 вооружение 3, 19, 6 война 9, 6, 30 артиллерия 9, 6 бой 6 караул 6	знамя 30 победа 9 государство 30	почта 30, 5 бухгалтерия 5 сберкасса 5 банк 5 работать 5 гитара 2 берег 2	1) Множитель операц 5 следовало бы исключить из дефиниции слова <i>армия</i> ; как показывает графа 4, он достаточно последовательно характеризует группу дескрипторов другого типа; 2) с дескриптором ГИТАРА общий множ. корпус 2, а с дескриптором БЕРЕГ — множ. сухопу 2

Таблица 3 (продолжение)

Слово	Дескриптор			Оценка возможного источника ошибки
	Есть интуитивная отнесенность слова к дескриптору и показатель связи	Есть интуитивная отнесенность, нет показателя связи	Нет интуитивной отнесенности, есть показатель связи	
1	2	3	4	5
аромат	аромат 1, 5, 19 обоняние 5	приятный 19 зловоние 19	след 3 портиться 5 тянуть 5 штемпель 3 фотография 3	1) Семантич. множ. отпеч 3 не должен входить в дефиницию слова 'аромат', 2) дескрипторы ПОРТИТЬСЯ и ТЯНУТЬ связаны с исходным словом через семантич. множ. запах 5. Связность не кажется невозможной
артиллерия	артиллерия 1, 3, 6, 1, 1, 2, 12, 1 вооружение 3, 6, 11 война 3, 6 армия 3, 6 огнестрельный 2 бой 6	военный 9	засада 3 осада 3 победа 3 караул 6	войск 3 } общие множ. вооруж 6 }
артист	актер 2, 2, 23 цирк 2, 21 образ 2, 21, 31 пение 21, 23, 14, 31, 3	искусство 21, 12 театр 21, 31	голос 21, 14, 3 воспроизвести 31, 4 литература 31, 4 жаворонок 3 слово 4	пе 3 публ 4 } общие множ.

археология	археология 1, 5, 2, 12, 10, 13, 36, 9, 14 древний 2, 9, 14	памятник 10, 9 культура 10, 13 история 10, 36, 9, 14	электричество 5 падение 5 вежливость 5 обряд 5	Множитель быт 5, очевидно, не должен входить в дефиницию слова и дескриптора АРХЕОЛОГИЯ
архитектор	архитектура 6, 1, 18, 22 памятник 6, 22	храм 18 дом 18 проект 22	колонна 6, 18, 22 арка 6, 22 башня 6, 22	Общий множ. — архитек. 6
архитектура	архитектура 6, 1, 21, 9, 4, 15 колонна 6 арка 6 башня 6 памятник 6	искусство 21 проект 9 дом 15	танцевать 21, 4 язык 4	Общий множ. — стиль 4
ассистент	хирургия 3, 36, 8 помощь 4, 21	звание 13	гинекология 3, 36 резать 3 кастрация 3 тянуть 4 щедрость 4 опека 4	врач 3 ] опера 3 } общие множ. помо 4 ]
астрология	астрология 1, 8 23, 5, 9, 1, 8 астрономия 5 звезда 5	предвидение 8, 23 суеверие 8, 9 судьба 8 будущий 8 теология 8 предрассудок 9	карта 5 снег 5	Общий множ. — зvez 5

Таблица 3 (окончание)

Слово	Дескриптор			Оценка возможного источника ошибки
	Есть интуитивная отнесенность слова к дескриптору и показатель связи	Есть интуитивная отнесенность, нет показателя связи	Нет интуитивной отнесенности, есть показатель связи	
1	2	3	4	5
астрономия	астрономия 5, 36, 10, 32 космический 3	физика 36, 32 луна 10, 32 метеорит 10, 32 комета 10, 32 затмение 10, 32 звезда 10, 32 планета 10, 32 солнце 10, 32 наука 36 оптика 36 карта 10	месяц 3	Общий множ. — астрон 3
атака	нападение 1, 5, 14 быстрый 11, 3 оборона 5, 14 сопротивление 5 поражение 3	победа 9, 10 война 9 спорт 15	засада 9, 5, 14 грабить 5 замедление 6 зима 6 преждевременность 6 срок 6 сумерки 6 спор 3 суд 3 проток 3 водопад 3	напа 5 насту 6

<b>атаковать</b>	нападение 1				
<b>атеизм</b>	атеизм 1, 5, 4, 7, 6, 9, 22 религия 5, 22  теология 5, 22 верование 5, 22 набожность 5, 22 бог 5, 22	отрицать 9 ересг. 22	молитва 5, 22 ангел 5, 22 богослужение 5, 22 сомнение 4 невежество 4 несправедли- вость 4 отказ 6, 9 уступка 6 непослушный 6 отговорка 6	бог 5  невер 4  отказ 6	общие множ.
<b>ателье</b>	живопись 4, 13 фотография 4	шить 25 одежда 25	кузница 4 слесарная 4 переплетная 4	Общий множ.— мастерс 4	

анализа вычленить ядро поля, т. е. группу слов, наиболее тесно связанных с дескриптором, и отделить его периферию. Вообще говоря, в идеале каждое поле должно структурироваться как своего рода микротезаурус, имеющий собственную классификационную схему и самостоятельное деление на подобласти, но трудоемкость такого структурирования превышает, видимо, его эффективность. Поэтому предлагается оптимальный, как нам представляется в данных условиях, вариант упорядочения слов в семантическом поле — по силе семантической связи их с дескриптором. Для определения силы связи эмпирически подобраны коэффициенты, в формуле которых учитываются данные о количестве совпадений в множителях и частоте последних. С помощью таких коэффициентов можно не только автоматически располагать слова в поле по убыванию силы связи и тем самым легко отделять ядро от периферии, но и сравнивать поля друг с другом.

Наконец, последняя из названных здесь задач — уточнение классификационной схемы — предполагает новую группировку дескрипторов в дескрипторные области на основании словарного наполнения семантических полей: общность такого наполнения, т. е. совпадение до определенной глубины состава словарных статей, служит критерием для отнесения соответствующих дескрипторов к одной области на классификационной схеме. Разрабатываемый вариант программного решения этой задачи основывается на кластерном анализе.

Второй круг задач нацелен на изучение некоторых аспектов русской лексической семантики теми средствами, которые дает в распоряжение исследователя опыт и результаты построения тезауруса. Об одном из этих результатов — получении оптимального для изучающих русский язык и пользующихся им словарника, состав которого установлен на базе различных, в том числе и семантических, критерий — рассказано выше. Такой словарник, на наш взгляд, имеет серьезное значение, поскольку очерчивает тот минимум лексики, который необходим для практического владения русским языком. Теоретическое же осмысление этого результата может заключаться в следующем.

Как отмечалось выше, показателем семантической непрерывности словаря является соблюдение на его словарнике «правила шести шагов», согласно которому расстояние между двумя произвольно выбранными словами, опре-

деляемое по общности множителей в дефинициях, не превышает шести шагов. Иначе говоря, во всяком словаре, адекватно репрезентирующем лексику языка, два случайным образом выбранных слова разделяет от нуля (они могут быть связаны непосредственно) до пяти, но не более, других слов; эти пять слов и образуют ту максимальную цепочку в шесть шагов, через которую можно проследить связь двух исходных единиц. В этой ситуации представляется логически возможным из множества, скажем, в 100 тыс. слов выбрать п-е число некоторых базовых единиц таким образом, чтобы все остальные слова из этих 100 тыс. были связаны с базовыми непосредственно, в один шаг — через семантические множители последних. Можно полагать, что наш оптимальный словарник, или семантическое ядро русской лексики, как раз и обладает таким свойством. Однако это положение, так же как и содержательный смысл «правила шести шагов», еще нуждаются в дополнительной проверке.

Среди других задач дополнительного цикла можно назвать следующие:

— Замена или исключение нерелевантных семантических множителей из числа последних на массиве дескрипторов. Этим достигается несколько целей: во-первых, уменьшается общий объем материала, которым оперирует машина; во-вторых, мы тем самым вносим уточнения в состав «семантического ядра» русской лексики; наконец, в-третьих, такая замена (скажем, синонимами, релевантными с точки зрения их частотности) ведет к необходимой корректировке и улучшению словарных дефиниций, на основе которых осуществлялось первоначальное кодирование.

— Использование полученных результатов как банка данных для семантического анализа лексики. Это предполагает, например, построение гнезд слов по одному, двум или более общим семантическим множителям с целью последующего изучения полученных таким путем ассоциативных групп, их сравнения с коррелирующими семантическими полями дескрипторов и выработки методических следствий для унификации и систематизации толкований слов, составляющих эти группы, в филологических словарях. Примеры подобных достаточно компактных групп можно наблюдать уже по результатам предварительного распределения, фрагменты которого приводились выше. Так, множитель приятн объединяет: *удовольствие*,

*приятный, сладкий, вкусный, привлекательность, надежда, восхищение, развлечение, радость, обольстить, прелесть, счастье, легкость, ласка, музыка, праздник, поздравление; множитель войск: артиллерия, война, вооружение, разведка, армия, военный, засада, осада, победа; множитель слад: торт, кондитерский, мед, сладкий, сахар; и т. п.*

Операция постредактирования машинного результата сводится к исключению, вычеркиванию из состава семантических полей тех слов, которые интуитивно представляются носителю языка (в данном случае — составителю словаря) не имеющими отношения к данному дескриптору. Можно назвать два наиболее вероятных источника таких ошибок: во-первых, неточности кодирования, когда в число семантических множителей случайно попадает, например, нулевой или избыточный компонент из словарного определения; во-вторых, неправильности в самих definициях, связанные как с искажением предметно-вещных отношений реального мира (например, *валуй* определяется в словаре Ожегова как «гриб с ... вогнутой шляпкой», хотя такие валуи в природе не встречаются, а *брюки* — только как «мужская одежда», что в наше время выглядит анахронизмом), так и с субъективностью, относительностью употребляемых составителями оценок (например, *лента* у Ушакова определяется как «широкая тесьма», а у Ожегова — как «узкая полоса ткани», *чемодан* — как «твердая коробка», хотя это вовсе не обязательно для чемодана, и т. п.). Число таких ошибок в целом невелико, и они легко поддаются выявлению (ср. выше комментарий в графе 5 табл. 3), поскольку возле каждого слова в словарной статье предусмотрено указание на тот множитель, который послужил основанием для включения этого слова в данное семантическое поле.

## Глава IV

# НЕКОТОРЫЕ ВОПРОСЫ СООТНОШЕНИЯ СЕМАНТИКИ И СИНТАКСИСА С ПОЗИЦИЙ ЛИНГВИСТИЧЕСКОГО КОНСТРУИРОВАНИЯ

---

### § 8. СТАТУС И СМЫСЛ ДЕФИНИТИВНЫХ ПРЕДЛОЖЕНИЙ

Grammatici certant  
Et ad huc sub judice lis est.  
*Horatius. Ars poetica\**

#### Общие замечания

Проблематика взаимоотношений семантики и синтаксиса, возникновение которой в европейской традиции связывают с эллинистической мыслью, к настоящему времени стала едва ли не центральной для науки о языке: к ней обращаются и генеративисты, и представители прикладных областей языкознания, специалисты по искусственному интеллекту, логики, философы и грамматисты. В зависимости от решаемых в каждом случае конкретных задач, от исповедуемых авторами тех или иных методологических принципов (что считать значением вообще, что такое значение предложения и какова его референциальная сущность) ответы на вопрос о соотношении семантики и синтаксиса могут быть различными. В рамках одной только порождающей теории развиваются два крайних, прямо противоположных друг другу направления: одно из них — традиционное, связанное с идеями трансформационизма исходит из примата синтаксиса, полагая, что никаких семантических представлений структуры предложения отдельно от синтаксических не существует; другое, называемое порождающей семантикой, ставит во главу угла семантическое представление смысла предложения, отводя синтаксису роль технического средства, поверхностного уровня воплощения этого глубинного смысла.

---

\*... О том грамматики в споре, Но и доныне их тяжба осталась еще нерешенной (*Гораций. Наука поэзии*).

Каждое из них существует в нескольких модификациях, но наибольшее распространение все же получили иные точки зрения, а именно концепции, в которых в том или ином виде делается попытка найти разумный компромисс между указанными крайностями, подойти к решению вопроса о синтаксических и семантических слагаемых смысла предложения с учетом специфической роли каждого из них. При этом сама возможность сопоставимости, совместимости синтаксических и семантических свойств и единиц их носителей может быть достигнута разными путями.

В одном случае исследователи отправляются от постулированного изоморфизма семантики слова и семантики предложения, рассматривая лексическое значение как иерархически упорядоченное множество элементарных смыслов, аналогичное дереву синтаксических зависимостей<sup>1</sup> в рекурсивной семантике, и здесь образец для «структурного подражания» задает синтаксис. Работы другого плана, в которых смысл отдельного предложения «выводится» из смысла всего текста, причем предложение «доопределяется», достраивается за счет отыскания антецедентов местоимений, снятия синонимии и устранения эллиптизмов, мы склонны трактовать как относящиеся к противоположному направлению, где эталоном изоморфного представления синтаксического и семантического смысла в конечном счете служит лексическое значение<sup>2</sup>. Дело в том, что анализ содержания текстов с применением семантических сетей, в которых смысл каждого предложения отображается самостоятельным графом (фреймом), дает основания не только усматривать принципиальное сходство, но и находить фактическое совпадение между узлами семантической сети (соответственно — вершинами графов) и концептами в семантическом пространстве тезауруса, т. е. дескрипторами, иерархия которых, как известно, задает семантику отдельного слова.

Наконец, совместимость синтаксических и семантических составляющих смысла предложения может быть получена в результате принципиально новой категоризации данной области и введения непрерывного спектра единиц, воплощающих лексическую и синтаксическую семантику. Так, У. Пюршель, рассматривая противоположение синтаксиса и семантики, относит к области первого три разряда явлений: категориальные правила (введение фразовых маркеров); субкатегориальные правила (т. е. правила развертывания сложных символов); правила, связанные

с синтаксическими свойствами лексики. Соответственно ведению семантики подлежат правила интерпретации (или синтезирования целого из частей) и лексикон как инвентарь формативов, конституирующих семантические характеристики. Эти явления в свою очередь представлены следующим спектром единиц: дистингема ( $R_0$ ) — единица нулевого уровня, обеспечивающая минимальное отличие в плане выражения от других знаков; плерема ( $R_1$ ) — языковой знак с минимальным значением; дериватема ( $R_2$ ) — комбинация единиц ранга  $R_1$  по правилам деривации; композитема ( $R_3$ ) — комбинация единиц рангов  $R_1$  и  $R_2$  по правилам композиции; флекстема ( $R_4$ ) — комбинация единиц рангов  $R_1$ ,  $R_2$ ,  $R_3$  с флекстивными граммемами ранга  $R_1$  по правилам образования флексивных форм; конституема ( $R_5$ ) — комбинация единиц рангов  $R_1$ — $R_4$  по правилам образования конституем; симплексема ( $R_6$ ) — комбинация единиц уровня  $R_5$  по правилам образования предложений; паратаксема ( $R_7$ ) — комбинация единиц уровня  $R_6$  по правилам координации; гипотаксема ( $R_8$ ) — комбинация единиц рангов  $R_6$  и  $R_7$  по правилам субординации; текстема ( $R_{8+n}$ ) — комбинация единиц рангов  $R_1$ — $R_8$  по контекстно связанным правилам. Предложенный спектр единиц позволяет осуществлять целостное описание смысла высказывания с учетом объективного содержания знаков и приемов их комбинирования<sup>3</sup>.

Развивая идеи компонентного структурирования смысла отдельных слов и распространяя затем правила комбинирования компонентов за пределы индивидуальных лексем — на их сочетания и в целом на предложение, М. В. Никитин приходит к необходимости выделения трех уровней или трех типов синтаксических отношений, которые позволяли бы в полной мере учесть синтаксическую составляющую в семантике высказывания. «Синтаксические отношения (vs. связи) — это отношения трех уровней: онто-синтаксического, логико-синтаксического

---

<sup>1</sup> Weinreich U. Explorations in semantic theory. — In: Current trends in linguistics, v. III. The Hague—Paris, 1966, p. 442—448.

<sup>2</sup> Падучева Е. В. О семантике синтаксиса. Материалы к трансформационной грамматике русского языка. М., 1974; Скороходъко Э. Ф. Семантические связи в лексике и текстах. — Вопросы информационной теории и практики, сб. № 23. М., 1974.

<sup>3</sup> Pürschel U. Semantisch-syntaktische Relationen. Untersuchungen zur Kompatibilität lexikalischer Einheiten im Deutschen. Tübingen, 1975, S. 36—71.

и коммуникативно-синтаксического. Отношения на онто-синтаксическом уровне — это отношения между десигнатами имен вещей в выражениях с подчинительной связью. В конечном счете — это синтаксическое отражение отношения вещей. Обобщенные категории вещей, квалифицированных по синтаксическим отношениям этого рода, описываются понятиями вроде агенс (агент, иногда также субъект), объект, инструмент, посредство, адресат, актант (более общая, чем предшествующие, категория участника отношения-действия), место, причина, результат, принадлежность и т. д. и т. п. Все это категории вещей-аргументов опто-синтаксических отношений... Логико-синтаксический уровень — это уровень метаотношений между десигнатами имен вещей и их признаков (свойств или отношений) в выражениях с подчинительной связью. Общее содержание этих метаотношений, состоящее в присыпывании вещам признаков, назовем экспликацией, а аргументы метаотношения экспликации назовем... экспликандумом и экспликантом... Наконец отношения на коммуникативно-синтаксическом уровне — это отношения между единицами коммуникативной структуры предложений. Они описываются в таких терминах, как субъект—предикат, предикация, тема—рема, данное—новое, коммуникативный центр высказывания, категория выделения и т. п.»<sup>4</sup>.

Таким образом, современный уровень представлений о соотношении семантики и синтаксиса — и самой постановкой проблем, и используемой аргументацией — решительно отличается от тех попыток рассмотрения этой проблематики, которые имели место 20 лет назад: теперь никто уже не оперирует всерьез разного рода *nonsense ketten* с целью иллюстрировать так называемую независимость синтаксиса от семантики или наоборот — показать полностью служебную роль синтаксиса в выведении к выражению некоторого глубинного смысла. Исследователи для подтверждения развиваемых ими положений обращаются к целому корпусу предложений того или иного языка, способному презентировать все многообразие допустимых в этом языке структурных, логико-грамматических, пропозитивных, интонационных типов конструкций<sup>5</sup>. Что касается самого существа взаимоотношений семантики и синтаксиса, то в усредненном виде оно хорошо, на наш взгляд, иллюстрируется пересечением вертикальной (семантика) и горизонтальной (синтаксис) осей

в трехмерном пространстве, где третья ось — ось глубины — воплощает «прагматику», т. е., согласно автору, отражает отношение элементов и их последовательностей к экстралингвистической информации<sup>6</sup>.

На этом очень бегло обрисованном, но достаточно широком фоне проблематики, связанной с семантикой и синтаксисом, нас будет интересовать относительно узкая сфера типов предложений — так называемые предложения идентификации, которые представлены дефинициями в толковых словарях. Аспект, в котором мы будем рассматривать эти предложения, касается ответа на вопрос, какая доля содержащейся в них информации передается лексическими средствами, т. е. семантикой лексем, а какая принадлежит непосредственно синтаксису. Решение этой задачи позволит в свою очередь вынести суждение об адекватности приемов кодирования информации в словарных определениях, использованных при автоматическом построении русского тезауруса, выявить пути улучшения алгоритма кодирования, с одной стороны, и пути оптимизации самих словарных дефиниций, с другой.

Характеризуя разновидности предложений дефинитивного типа, Н. Д. Арутюнова пишет: «Отношения и о м и и а ц и и соединяют объект и его имя, т. е. элемент предметного мира и элемент языкового кода. Ср.: *Этого мальчика зовут Коля. Это Коля. Коля. Это дерево — сосна.* Отношения именования могут соединять не только предмет и его имя, но также понятие (свойство, качество, событие и пр.) и способ его обозначения в языке (слово). Однако для синтаксической организации предложения существенна лишь наиболее простая форма этих отношений, соединяющая предмет (лицо) и его собственное имя. . .

---

<sup>4</sup> Никитин М. В. Лексическое значение в слове и словосочетании (на материале английского языка). Автореф. докт. дис. Л., 1974, с. 29.

<sup>5</sup> Так, применительно к материалу русского языка, Н. Д. Арутюнова выделяет и анализирует четыре логико-синтаксические модели, достаточные для описания синтаксиса: «1) отношения экзистенции, или бытийности, 2) отношения идентификации, или тождества, 3) отношения номинации, или именования, 4) отношения характеризации, или предикации в узком смысле этого термина» (Арутюнова Н. Д. Предложение и его смысл. М., 1976, с. 18).

<sup>6</sup> Oiler J. W., Jr. On the relation between syntax, semantics and pragmatics. — In: Linguistics at the crossroads. Padova — Lake Bluff, 1977, p. 48.

Инверсия отношений именования, при которой мысль двигалась бы от слова (означающего) к его значению (сигнификату), создает отношения интерпретативного типа, которые очень близки к отношению тождества. Сообщая о том, что значит слово, мы не только соединяем звук со смыслом, но как бы приравниваем значение слова к его толкованию при помощи других слов. Ср.: *Печаль это скорбно-озабоченное, нерадостное, невеселое настроение, чувство*. Такие предложения используются тогда, когда адресату неизвестен первый элемент равенства — означаемое интерпретируемого слова. Они могут быть определены как *экспликативные*. Интерпретативные отношения не соотносятся с особым типом предложений. Они, поэтому, не входят в число логико-синтаксических „начал“<sup>7</sup>. Таким образом, автор фактически исключает указанные предложения из своего анализа, хотя по ходу изложения посвящает их особенностям отдельные замечания<sup>8</sup>.

При рассмотрении словарных дефиниций, взятых в узком смысле этого слова, в качестве особого типа предложений мы сталкиваемся сразу с рядом таких их свойств, которые как будто выводят их за рамки единиц, традиционно называемых предложениями, ср.: *Кушанье из сваренной крупы. Род зубатых китов. Повар в воинской части или в рабочей артели. Судорожный выдох с хрюком (при болезни дыхательных путей). Страдать кашлем, издавать звуки кашля*. Взятые, так сказать, в чистом виде, словарные толкования, т. е. высказывания в правой части толкового словаря, представляют собой номинативные или номинализованные словосочетания, предположительно эквивалентные в семантическом отношении номинативному элементу (слову) в левой части, т. е. на входе словаря. Предложениями они становятся только в соединении с этой левой частью, т. е. при включении самого идентифицируемого слова в состав последовательности лексических единиц, образующих дефиницию, и при реальном или мысленном восстановлении глагола-связки: *Каша (—) • кушанье из сваренной крупы. Кашалоты (представляют собой) род зубатых китов. Кашевар — это повар в воинской части или в рабочей артели. Кашель (есть) судорожный выдох с хрюком. Кашлять значит страдать кашлем, издавать звуки кашля*. В таком виде эти предложения можно трактовать как предложения тождества, в которых приравниваемые друг другу элементы всегда

стоят в исходной форме — существительные в именительном падеже, а глаголы — в инфинитиве: *каша — кущанье, кашалоты — род, кашевар — повар, кашель — выдох, кашлять — страдать*.

Такая трактовка в свою очередь может вызвать дополнительное возражение, поскольку в полноправном предложении обязательно как будто должен быть предикат действия. Но это возражение при ближайшем рассмотрении снимается, коль скоро мы можем показать, что номинализованная форма дефиниции всегда допускает трансформацию ее в конструкцию с предикатом действия. *Линейка — прямая черта на бумаге или доске (помогающая писать прямыми ровными строками) → Линейка, представляя собой прямую черту на бумаге или доске, помогает писать прямыми ровными строками.* Более того, возможен такой угол зрения, при котором, наоборот, любое предложение предстает в форме определения. К такому выводу, в частности, приходят исследователи, разрабатывающие пути алгоритмизации анализа связного текста и установившие, что «описание значения языкового выражения в явном или неявном виде имеется в каждом высказывании»<sup>9</sup>, причем каждый текст характеризуется уникальной комбинацией определений для каждого языкового элемента.

Естественно, что эти взаимопереходы дефиниции в пропозицию и обратно не совершаются бесследно и не безразличны для ее синтаксического статуса. Затрагивают они прежде всего коммуникативный аспект. В самом деле, предложения с двумя номинативами или двумя инфинитивами всегда стояли особняком в классификации русских синтаксических конструкций и вызывали разноречивую интерпретацию у исследователей<sup>10</sup>. С одной стороны, возникает очень сильный соблазн видеть в подобной конструкции идеальное воплощение темы (левая часть равенства) и ремы (правая часть), имени и классифицирующего предиката. Н. Д. Арутюнова так и характеризует их: «В строгом и собственном смысле предложения денотативного тождества, т. е. предложения идентификации, лишены

---

<sup>i</sup> Арутюнова Н. Д. Указ. соч., с. 18.

<sup>8</sup> Там же, с. 304, 319, 325, 328, 329, 334, 337.

<sup>9</sup> Ратцева И. И., Строганов В. А. Семантика текста и алгоритмизация. — В кн.: Сборник научных трудов МГПИИ им. М. Тореза, 1976, вып. 103, с. 119.

<sup>10</sup> См. об этом, напр.: Арутюнова Н. Д. Указ. соч., с. 304—324.

подлежащего и сказуемого, хотя они членятся на тему и рему»<sup>11</sup>. Но с другой стороны, при внешнем сходстве словарных дефиниций с предложениями тождества, о которых говорит Н. Д. Арутюнова, первые лишены важнейшего качества единиц, к которым применимы принципы актуального членения, — в них не выражена коммуникативная направленность.

Правда, об этом свойстве дефиниций тоже существуют разноречивые суждения. Если, констатируя его, мы держим перед мысленным взором какой-то толковый словарь, в котором каждая дефиниция существует как бы сама по себе, в отрыве от других и вне всякого контекста, то, вероятно, наше утверждение не вызовет сомнений, и большинство читателей будет готово с ним согласиться. Однако румынский логик К. Попа, например, строит свою теорию определений как раз на противоположной концепции, исходя прежде всего из коммуникативного характера, адресности дефиниции<sup>12</sup>. Истоки его позиции проследить нетрудно: они восходят к Аристотелю и вообще к традициям древнегреческих ученых, сочинения которых создавались зачастую в форме диалога, как запись определенной беседы, в контексте которой предложения тождества, естественно, приобретали псевдокоммуникативный характер. Упоминание собеседника или собеседников, к которому апеллирует К. Попа для подтверждения своей концепции при раскрытии сущности определения в сочинениях средневековых логиков, а затем и у Локка, Декарта, Паскаля, тоже, на наш взгляд, ничего не доказывает, так как в этом отчетливо проявляется прямое воздействие эллинской традиции. Но уже в логике Пор-Рояля и далее у ряда математиков и философов начиная с XVIII в. мы не встретим упоминания о коммуникативных свойствах дефиниции<sup>13</sup>, как не разделяют полностью коммуникативную точку зрения на определения и авторы послесловия к книге К. Попы — Б. В. Бирюков и Д. П. Горский<sup>14</sup>.

Интересно в этой связи отметить, что у самого Аристотеля, помимо включенности дефиниции в ситуацию диалога, ничто, как кажется, не свидетельствует о ее коммуникативном характере, а многие из сформулированных им условий правильности определения даже противоречат ему. «Эти условия изложены во 2—14-й главах книги VI «Топики». Они представляют собой «общие места», касающиеся того, как избежать неясности в определениях.

В последних не должны использоваться омонимы, метафоры, бессмысленные формулировки, определения не должны содержать лишних слов; предшествующее не должно определяться в них через последующее; определяемый объект нужно относить к его собственному роду; нельзя применять в качестве видового отличия термины, которые не могут быть таковыми, например, категории; запрещается также использовать в этой роли апелляцию к эмоциям, переживаниям и т. д.»<sup>15</sup>.

Итак, перефразируя мысль К. Леви-Строса, можно подвести итог следующим образом: сказать, что язык служит целям коммуникации — значит сказать банальность, но сказать, что все в языке служит целям коммуникации — значит утверждать нечто абсурдное.

### Парадоксы дефиниции

Отсутствие отчетливо выраженной коммуникативной направленности делает излишним в словарных определениях выражение модальности, включение эмоционального момента, разнообразие интонации. Всех этих свойств лишиены предложения-дефиниции, вернее, эти свойства у данных предложений не подлежат варьированию, они фиксированы раз и навсегда в одном из своих значений: интонация — только повествовательная, утвердительная либо отрицательная, при невозможности вопросительной и восклицательной; отсутствие эмоциональности практически исключает возможность употребления в этих предложениях метафор и идиом, а это ведет, в свою очередь, к почти полному снятию проблемы семантической совместимости лексем, составляющих дефиницию, т. е. проблемы несвободной сочетаемости. В словарных толкованиях мы не встретим сочетаний, в которых *совесть*, например, *говорит, требует, допекает, пробуждается, запрещает; страсти волнуются, кипят, захлестывают; радость брызжет, переливается через край*; здесь невозможно гореть желанием, пылать ненавистью, разжигать злобу, вылечиться от любви, мучиться завистью, так же как недопу-

<sup>11</sup> Там же, с. 325.

<sup>12</sup> Попа К. Теория определения. М., 1976.

<sup>13</sup> Попа К. Указ. соч., с. 40—78.

<sup>14</sup> Там же, с. 236.

<sup>15</sup> Там же, с. 23.

стимы лихорадка любви, любовная зарaza, накал страсти, капля жалости или луч надежды<sup>16</sup>. Слова в лексикографических дефинициях употребляются только в своей прямой денотативной или сигнификативной отнесенности, и вопрос их сочетаемости есть вопрос классификационной совместимости соответствующих групп и классов в мыслимой или вещной реальности.

Вместе с тем, констатируя отсутствие эмоционального момента, мы не можем сказать, что предложения-дефиниции начисто лишены эмфазы. Однако эта эмфаза носит не образный, как в приведенных выше примерах, а логический характер и передается средствами семантической гиперхарактеризации, с помощью прилагательных и наречий (*крайне, исключительный, совершенно, чрезмерный, полностью* и т. п.), выступающих в функции усилительных частиц. Ср.: *Преобразовать — совершенно переделать, изменить к лучшему*, тогда как глагол *переделать — сделать заново, по-иному или иным* в общем не нуждается в том усилении, которым он сопровождается в дефиниции *преобразовать; Помешанный — имеющий исключительное пристрастие к кому-чему-н.*, где опять-таки эмфатическое прилагательное *исключительное* ничего не добавляет к семантике слова *причастие*, которое в свою очередь определяется как *сильная (N3) склонность; Полюс — нечто прямо противоположное чему-н. другому*, где эмфатическое *прямо* является гиперхарактеристическим, поскольку *противоположный* уже есть *совершенно (N3) несходный, противоречащий другому*. Аналогично положение эмфатических средств в дефинициях: *Победа — . . . полное (N3) поражение противника; Погореть — потерпеть полную (N3) неудачу. . . ; Плевать — совершенно (N3) не считаясь ни с кем-н., выражать презрительное безразличие; Презрение — глубоко (N3) пренебрежительное отношение к кому-чему-н.; полное (N3) безразличие к чему-н.; Популярный — общедоступный, вполне (N3) понятный по простоте, ясности изложения и т. д.* Эти примеры, с одной стороны, показывают независимость эмфатической гиперхарактеризации от наличия—отсутствия эмоционально-стилевой окраски у самого толкуемого значения (например, *погореть — прост., плевать — разг.*, а слова *победа* и *презрение* нейтральны, лишены каких бы то ни было помет); с другой стороны, они достаточно убедительно свидетельствуют о том, что сами используемые средства такой гиперхарактеризации семан-

тически пусты, играют роль аргументов *ad hominem*, сообщая дефиниции как носителю некоторого объективного знания субъективную выразительность живого речевого акта.

Далее, предложения-дефиниции характеризуются также отсутствием временной прикрепленности: форма настоящего времени глагола-связки придает этим предложениям всевременной, панхронический характер, выражает значение «всегда». Ср.: *Кремль — городская крепость в старых русских городах; Крепить — скрепляя, делать прочным, крепким; Крепление — совокупность устройств для предохранения горных выработок от обвала; Крепостничество — общественный строй, основанный на крепостном праве и эксплуатации крепостных крестьян*. Если же формы прошедшего времени глагола (причастия) изредка и встречаются в дефинициях, то они никогда не противопоставляются настоящему или будущему, а используются либо для характеристики ушедших в прошлое исторических явлений (и в этой функции с равным успехом могут быть заменены «настоящим историческим»), либо в чисто перфектном значении, т. е. «начиная с некоторого момента в прошлом и всегда». Ср.: *Крепостной — относящийся к общественному строю, при котором помещик имел право на принудительный труд, имущество и личность принадлежащих ему крестьян... (находящийся в крепости [стар.], т. е. в крепостной зависимости)*, где глагол-связка оформлен прошедшим временем, а обе причастные формы — *принадлежащих и находящийся* — употреблены в настоящем (историческом). Перфектное значение представлено, например, в случаях: *Нарушитель — тот, кто нарушил какие-нибудь правила, законы, обычаи; Крещеный — подвергшийся крещению*.

Это свойство фиксированности времени в дефиниции дает как будто дополнительное основание для того, чтобы исключить такие предложения из разряда пропозиций, тем более что в них практически не бывает глаголов пропозиционального отношения. Однако противоречивость статуса словарных дефиниций как особого типа синтаксических конструкций в том и состоит, что они допускают сочетание взаимоисключающих свойств. Так, не имея четких формальных показателей коммуникативной направленности и не являясь препозитивным высказыванием, они

---

<sup>16</sup> Ср., напр.: Арутюнова И. Д. Указ. соч., с. 93—111.

тем не менее широко используют, во-первых, разного рода актуализаторы — *этот, такой, том*, во-вторых же, часто опираются на изъяснительные конструкции, которые и служат как раз одной из примет пропозиции<sup>17</sup>. Ср.: *Крепкий — прочный, такой, что трудно разбить, сломать, порвать; то же, что стойкий; Перспектива — искусство изображать на плоскости трехмерное пространство в соответствии с тем кажущимся изменением величины, очертаний, четкости предметов, которое обусловлено степенью отдаленности их от точки наблюдения; Наручники — металлические кольца, соединенные цепочкой, надеваемые на руки преступникам, заключенным; Наркотики — сильно действующие вещества, преимущественно растительного происхождения, парализующие деятельность центральной нервной системы и вызывающие искусственный сон и болезненность; Нарост — то, что наросло, наслолось на чем-нибудь.*

Таким образом, мы приходим к осознанию первого парадокса в статусе дефиниции, который заключается в следующем. Будем понимать под пропозицией в широком смысле «„объективированное“ содержание мысли, отделенное от субъективной модальности и непосредственно соотнесенное с обозначаемым „положением дел“, как если бы постоянным спутником пропозиции был модус утверждения, а единственным ее назначением — „портиривание“, вопроизведение фрагментов действительности»<sup>18</sup>, что позволяет пропозиции сохранить связь с условиями истинности предложения. Тогда словарная дефиниция целиком подходит под это определение и может рассматриваться как пропозиция, а следовательно, семантика входного слова равнозначна пропозитивному значению. Но с другой стороны, пропозитивная семантика с необходимостью «соотносима с осью времени» и «принаследжит к разряду речевых значений»<sup>19</sup>, в то время как в дефиниции временной параметр фиксирован, а дефиниция-предложение не является фрагментом речевого развертывания более крупной темы, не является фрагментом единого текста, развивающегося во времени. Но как же быть в этом случае с условиями истинности предложений, в чем они должны заключаться для дефиниции?

Второй парадокс дефиниции выявляется в связи с возможностью членения ее — как предложения идентификации — на тему и рему. Формально, казалось бы, все довольно просто: тема в этом случае задается входным

словом, а рема представлена толкованием в правой части словаря. Однако при этом возникает противоречие с самыми принципами актуального членения, согласно которым под темой всегда понимается уже известное, данное (*datum* или *topic*), а рема представляет собой новое, еще неизвестное (*новум* или *focus*). В словаре же мы имеем обратную картину: предполагается, что пользователь обращается к нему с целью узнать нечто о *н е и з в е с т н о м* ему слове, которое дано на входе словаря, в то время как правая часть подразумевается понятной и известной пользователю. Иными словами, тема и рема в реальной жизни словаря как бы меняются местами, потому что *новум* оказывается входным словом, которое объясняется через *datum*.

Такое перевертывание для предложений тождества, вообще говоря, представляется допустимым, поскольку в обычном общении тема и рема циклически меняются местами и такая мена в ряде случаев не приводит к нарушению равенства между левой и правой частями предложения. Ср.: *гусары — вид легкой кавалерии=вид легкой кавалерии — это гусары; гусеница — личинка бабочки, обычно червеобразная, с несколькими парами ног=личинка бабочки, обычно червеобразная, с несколькими парами ног — это гусеница.* Но для большинства словарных definicij подобные перестановки не проходят бесследно и приводят к появлению фундаментальной неопределенности. Ср.: *густой — состоящий из многих, близко друг к другу расположенных однородных предметов, но обратное — состоящий из многих, близко друг к другу расположенных предметов* может означать уже не только *густой*, но и *частый*; *струг — старинное речное деревянное судно, но обратное — старинное речное деревянное судно — это и струг, и шлюп, и дощаник; шафран — сорт яблок, тогда как обратное — сорт яблок — это и шафран, и антоновка, и титовка, и штирифель, и джаннатан, и анис; больница — это лечебное учреждение, но лечебное учреждение — это и больница, и клиника,*

---

<sup>17</sup> «... Имя *высота* в definicijii *Высота* есть вертикальное изменение предмета не выражает препозитивного значения, а в предложении *Высота горы, нас поразила* значение сочетания *высота горы* принадлежит к препозитивному типу» {Арутюнова Н. Д. Указ. соч., с. 80).

<sup>18</sup> Арутюнова Н. Д. Указ. соч., с. 34.

<sup>19</sup> Там же, с. 80.

и здравпункт, и амбулатория, и лазарет, и диспансер, и санаторий, и поликлиника, и госпиталь, и лечебница, медпункт, санчасть. Подобная неопределенность является «неопределенностью из-за незнания»<sup>20</sup>, из-за неполноты правой части, и обусловлена она вхождением каждого отдельного значения в свое семантическое поле, фрагменты которого приводятся к выражению в случае актуализации этого значения. Снимается эта неопределенность в дефиниции не путем внесения в формулировки всякого рода мелких уточнений типа: *Клиника — лечебное учреждение, при котором ведется научная и учебная работа; Санаторий — стационарное лечебное учреждение, оборудованное для лечения и отдыха*, которые могут быть бесконечными (потому что в приведенных примерах, скажем, *клиника* — тоже *стационар* и тоже *оборудована для лечения* и т. д.). Однозначность и определенность достигаются принципиально иным образом — включением в дефиницию самого идентифицируемого слова.

В другом месте для обоснования этого приема (а именно, включения слова в число его семантических множителей) кодирования информации при автоматическом построении тезауруса из толкового словаря<sup>21</sup> мы ссылались на семантическую теорию истины А. Тарского, согласно которой всякое высказывание эквивалентно утверждению о его собственной истинности. Теперь к этому логико-философскому аргументу мы можем добавить еще два, уже чисто лингвистических: во-первых, такое включение необходимо для снятия неопределенности, порождаемой законом семантического поля; во-вторых же, как было показано выше, дефиницию можно приравнять по структуре к номинализованному предложению, считать ее предложением тождества и применять (как для анализа отдельной дефиниции, так и для системного осмысления всей их совокупности в словаре) методику графов (фреймов) только при условии внесения самого определяемого слова в состав его собственной дефиниции (см. ниже, § 11).

Наконец, третий парадокс дефиниции обнаруживается при взгляде на нее в синтагматической перспективе. Трактуя ее как предложение, мы вправе ожидать включенности ее в некоторый контекст, потому что никакое предложение не существует изолированно, само по себе, но всегда так или иначе ориентировано на коммуникативную ситуацию, воплощением которой выступает текст, понимаемый в широком смысле этого термина. Формаль-

ными показателями такой ориентированности являются обычно синонимические и местоименные повторы (анагорического и катафорического плана), а логико-содержательным основанием — наличие пресуппозиции. С помощью этих критериев мы легко можем восстановить фрагменты соответствующей коммуникативной ситуации даже по одному предложению. Например, по фразе *В той роще растут белые грибы* можно с большой степенью достоверности установить наличие говорящего и слушающего (или слушающих) в ситуации; наличие поблизости как минимум двух небольших лесных массивов (участков), о чем свидетельствует употребление указателя *той*, предполагающего возможность выбора; далее, сделать вывод о большей удаленности участка (*тот — этот*), характеризуемого указанным свойством, по отношению к другому (другим); наконец, представить себе, что участок, о котором идет речь, засажен одношородным лесом (особенность, бессознательно воспринимаемая носителями языка, но не отмечаемая толковыми словарями для слова *роща*). Что касается прочих фрагментов пресуппозиции рассматриваемого предложения, — является ли рощей другой участок (или другие участки и сколько их), какой породы деревья представлены в данной роще, насколько она удалена, каково количество слушающих, в какое время года происходит разговор, — то они не поддаются восстановлению на основе имеющейся информации. Конечно, предпосылкой понимания данной фразы адресатом (а ориентированность на понимание входит в намерение говорящего) является одинаковое представление у того и у другого о том, что такое *белые грибы*. Однако это представление не входит в пресуппозицию рассмотренной фразы, а составляет, если можно так выразиться, пресуппозицию самой коммуникативной ситуации, т. е. пресуппозицию существования всего текста, условием возникновения которого служит знание языка и мира обеими сторонами.

Таким образом, мы приходим к пониманию того, что пресуппозиция бывает двух рангов, двух уровней: пресуппозиция-*I*, которая действует в пределах отдельного текста, ограничивается его рамками, и пресуппозиция-*II*,

---

<sup>20</sup> Она противостоит «неопределенности из-за сложности», о которой будет сказано ниже.

<sup>21</sup> Караплов Ю. Я. Частотный словарь семантических множителей русского языка. М., 1980, с. 10.

дотекстовая, или предтекстовая, представляющая собой как бы предзнание определенного текста и составляющая часть наших знаний о мире. Составными частями последних являются и так называемые «фоновые знания» в словарях, акцентирующие национальную специфику картины мира и относящиеся к компетенции страноведения, и то, что можно назвать «фондовыми» знаниями, т. е. универсальные, вненациональные, находящиеся на уровне «само собой разумеемости» представления о вещах, свойствах и отношениях в мире, имплицитно присущие человеку как представителю вида, и потому не включаемые в обычные филологические словари.

Если теперь обратиться к дефиниции, то парадоксальность ее статуса проявится в том, что, будучи нормальным предложением, она, тем не менее, не ориентирована с необходимостью на какой-то особый текст. Иными словами, можно, конечно, придумать некоторый специальный текст, куда в качестве контекстно связанного предложения входила бы, например, дефиниция *Народовластие* — это *власть народа, демократия* или *Народоведение есть прежнее название наук, изучающих культуру и быт народов, этнография*<sup>22</sup>, однако такой текст вовсе не обязателен, и при составлении словарных толкований его существование и не подразумевается. Вместе с тем в дефинициях сплошь и рядом употребляются синонимические (ср. в приведенных примерах — *демократия* и *этнография*) или местоименные показатели связности, которые должны как будто бы свидетельствовать о включенности дефиниции-предложения в какой-то более крупный текст. Ср.: *Нарожать* — то же, что *народить*; *Примесь* — то, что *прибавлено, примешано к чему-нибудь*; *Сват* — тот, кто *сватает кого-нибудь кому-нибудь*; *Сохранный* — такой, где можно надежно сохранить что-нибудь; *Сцепной* — такой, который *спепляется, который можно сцепить*; *Швейный* — относящийся к шитью; *Обеспечение* — то, чем обеспечивают кого-, что-нибудь; *Извести* — то же, что *истратить*; то же, что *истребить*; то же, что *измучить*.

Следовательно, для объяснения употребления этих анафорических и катафорических (местоименных и синонимических) показателей связывания нужно либо предположить существование некоторого глобального супертекста, составленного только из словарных дефиниций, либо, наоборот, видеть в этих показателях формальное выражение связей между множеством отдельных само-

стоятельный текстов, а не предложений. Первое предположение не выдерживает критики, поскольку гипотетический супретекст из словарных толкований не отвечает главному текстовому критерию — он лишен единой сквозной темы, так или иначе совместимой с разными частями текста. Принятие второго решения заставляет нас трактовать отдельную дефиницию как самостоятельный текст, что не противоречит представлениям о размерах текста вообще, так как верхняя и нижняя его границы подвижны, и текст вполне может состоять из одного предложения. Проблемы же взаимоотношений между текстами относятся уже к тезаурусостроению, причем, как известно, тезаурус может быть построен только для совокупности связанных друг с другом текстов. Взаимосвязанность текстов-дефиниций обосновывается не только формально — наличием в них анафорических и катафорических коннекторов, но и содержательно — действием на всем объеме словаря «правила шести шагов»<sup>22</sup>, согласно которому максимальное расстояние между двумя произвольно выбранными словами (и их дефинициями) не превышает шести «переходов», или «шагов», до общего элемента в двух сравниваемых дефинициях. Результаты этих рассуждений согласуются и с развиваемыми выше (§ 5) представлениями о принципиальной ориентированности словарных толкований на тезаурус, диктуемой действием системных закономерностей в лексике.

Итак, рассмотрение статуса словарных дефиниций дает основания трактовать их как особый тип синтаксических конструкций, которые допускают сочетание ряда противоречивых свойств, приводящих к определенным парадоксам. Все три парадокса проистекают из двусмысленности положения дефиниций как коммуникативных единиц, и поэтому их анализ должен вестись с учетом следующих особенностей: во-первых, дефиниция становится полным номипализованным предложением только с включением в ее состав самого идентифицируемого слова, причем такое включение служит одновременно снятию неопределенности в ее семантике, порождаемой законами семантического поля. Во-вторых, не обладая препозитивным значением, предложения-дефиниции в качестве условий своей истин-

---

<sup>22</sup> Здесь, вероятно, точнее было бы сказать «этнология», а не «этнография».

<sup>23</sup> Караулов Ю. Н. Общая и русская идеография. М., 1976, с. 77—79.

ности имеют ориентированность на тезаурус. Это их свойство, так же как наличие в них показателей формальной взаимосвязанности, превращает номинализованные предложения-дефиниции в особого рода тексты объяснительного характера, которые включаются в систему отношений тезауруса и подчиняются действующим в нем семантическим закономерностям. Последние проявляются в том, что глубина тезауруса соответствует семантической связности дефиниций по «правилу шести шагов», а сама структура и состав дефиниции определяются в свою очередь максимум шестью признаками, достаточными для выражения любого значения.

### § 9. СЛАГАЕМЫЕ СМЫСЛА И ЛЕКСИЧЕСКИЙ КОМПОНЕНТ СЛОВАРНЫХ ОПРЕДЕЛЕНИЙ (ДЕНОТАТИВНЫЙ АСПЕКТ)

The types of ideas built up into idea stocks could be considered under three headings:

1. Precise named-ideas: horse, democracy, justice, etc.
  2. Vague named-ideas (porridge words): mechanism, device, arrangement, requiron, somewhat, somehow, etc.
  3. Interaction named-ideas: oppose, combine, modify, cause, develop into, etc.
- . . . One only acts upon named-ideas.

*E. de Bono. Practical thinking\**

#### Конъюнкция и дизъюнкция множителей

В семантике дефиниции можно видеть три слагаемых: лексическое, т. е. значения самих слов, составляющих толкование; логико-семантическое, отражающее отношения между предметами реального мира, обозначенными этими словами, и формально не всегда выраженное; синтаксическое, или в широком смысле грамматическое, эксплицитно представленное в виде отношений между словами на синтагматической оси. Всякие другие источники добавочного смысла (коннотативный, ситуативный, пропозитивный) в дефиниции отсутствуют в силу особенностей ее статуса, как было показано выше.

Подобное представление трех родов компонентов смысла в дефиниции не совпадает в точности с составляющими тернарного описания объектов как систем, которые охватывают «вещи», «свойства» и «отношения»<sup>24</sup>. Так, лексический компонент способен выражать все три показателя системного подхода: *Многоугольник — геометрическая фигура*.

*екая фигура на плоскости, ограниченная замкнутой ломаной линией, где многоугольник и геометрическая фигура — «вещи» или «предметы», сочетание на плоскости передает «отношение» местонахождения или же атрибутивное отношение, подчеркивающее плоскостной характер фигуры, а причастным оборотом выражено «свойство» ее ограниченности замкнутой ломаной линией.* Собственно «отношение» может выражаться либо лексически, либо синтаксически. Например, в определении *Жизнь — совокупность явлений, происходящих в организмах, особая форма существования и движения материи, возникшая на определенной ступени ее развития* между двумя «вещами» жизнь и совокупность явлений отношение «иметь элементами множества» фиксировано синтаксически как отношение между подлежащим и именной частью сказуемого, но также лексически — указанием того, что это самое «множество» имеет место (совокупность), тогда как в дефиниции *Жир — нерастворяющееся в воде маслянистое вещество, содержащееся в животных и растительных тканях* отношение между «вещами» жир и вещество — «быть элементом класса» — передано лишь синтаксически, а «свойство» нерастворяющееся в воде — и лексически, и синтаксически.

Опираясь на возможность логически определить свойство<sup>25</sup> через отношение и наоборот<sup>26</sup>, мы имеем возможность свести тернарные формулы системного описания объектов к бинарным, в которых будут фигурировать только «вещи» и «отношения». Опыт такого рода осуществляв-

---

\* Во всем комплексе идей можно выделить три их типа:

1. Конкретные идеи-наименования: лошадь, демократия, справедливость и т. п.

2. Неопределенны выраженные идеи-наименования (слова-виагнегреты): механизм, приспособление, устройство, средство, нечто, как-то и т. п.

3. Идеи-связки: противопоставлять, соединять, изменять, вызывать, развиваться и т. п.

... Действовать человек может только на основе идей, выраженных словами (*E. de Боно. Практическое мышление*).<sup>24</sup>

<sup>24</sup> Уемов А. И. Системный подход и общая теория систем. М., 1978,

с. 11—12.

<sup>25</sup> «Свойство есть то, что, характеризуя вещи, не образует новых вещей. Что же касается отношения, то оно есть то, что, будучи установлено между вещами, образует новые вещи» (Уемов А. И. Указ. соч., с. 86).

<sup>26</sup> «Свойство вещи — это отношение, существующее в той вещи, в которую включается данная вещь. . . Отношение вещи — это свойство, характеризующее ту вещь, в которую включается данная» (Уемов А. И. Указ. соч., с. 88).

лен в информационно-поисковой системе, использующей семантический язык RX-кодов<sup>27</sup> (R — отношение, X — предмет), и в основе его лежит мысль о том, что любое свойство или признак «представляет собой сложное образование, состоящее из элементов двух видов — обозначающих отношение и обозначающих предмет»<sup>28</sup>.

С учетом этих предварительных замечаний рассмотрим последовательно все три названных в начале раздела компонента, три слагаемых смысла дефиниции. Что касается первого из них — лексического, то оно характеризуется как сумма смыслов знаменательных слов, входящих в дефиницию, и такая его трактовка не вызывает возражений до тех пор, пока не встает вопрос о том, что считать знаменательным словом в данном случае. В самом деле, понимание лексического слагаемого как суммы смыслов лексем возникло в результате автоматического переноса на слова естественного языка и их дефиниции правила из логики, согласно которому признаки в определении того или иного понятия находятся в отношении конъюнкции друг с другом. Это правило сохраняет свою силу для терминов, поскольку в общем случае они по содержанию равны понятию, но уже в области терминологии оно применяется с оговоркой, и при семантической записи терминологического текста отношение конъюнкции фактически не учитывается: «Информация о конъюнкции признаков используется в том смысле, что мы не можем игнорировать ни один признак, указанный в определении значения термина»<sup>29</sup>. При семантической же записи определений слов естественного языка у нас тем более нет никаких оснований говорить о конъюнкции, во-первых, в силу того, что значение лексического слагаемого представлено иногда не суммой, а разностью лексем, во-вторых, по причине того, что часть признаков при такой записи словарной дефиниции оказывается, что называется, «непертinentной», порождающей «шум» при тезаурусном распределении слов, т. е. избыточной и потому подлежащей исключению. Поясним обе ситуации на примерах, одновременно напомнив читателю, что предпосылкой семантической записи, или кодирования дефинитивной информации, и соответственно — комплектования дескрипторных статей в тезаурусе русского литературного языка, с самого начала был принцип *унитермности* единиц: для упрощения кодирования и машинного исполнения слово-сочетания на этапе автоматической обработки не рассмат-

ривались и в расчет принимались только отдельные слова. Естественно, что достигаемое при этом упрощение в одном отношении оборачивалось рядом трудностей в «лексикалистском» представлении признаков, оформленных несколькими словами. Рассмотрению этих трудностей и посвящена значительная часть данной главы.

В словарных толкованиях в качестве элементов дефиниции широко распространены так называемые описательные обороты — сочетания существительных с глаголом, в которых глагол выполняет полусвязочную функцию, а вся семантика передается существительным: *обратить внимание* (заметить), *прийти к соглашению*, *установить взаимопонимание* (поладить), *отдать приказ* (приказать), *принять участие* (примазаться), *подвергнуть критике* (протащить), *сделать легче*, *сделать проще* (облегчить), *подвести итог* (подсчитать), *держаться мнения* (подумать), *сделать богаче* (обогатить), *давать возможность* (позволять), *иметь намерение* (подумать), *выработать навык* (приучать), *пустить корни* (приняться), *служить натурай* (позировать), *произвести ликвидацию* (ликвидировать), *вступать в знакомство* (знакомиться),  *положить конец* (подавить), *сохранить*, *удерживать в памяти* (помнить), *сделать заключение* (признать), *прийти в изумление* (поразиться), *издавать треск* (трещать), *совершить нападение* (напасть), *получить привычку* (привыкнуть), *оказать помощь* (помочь), *сделать ошибку* (ошибиться), *нанести оскорбление* (оскорбить), *давать обещание* (обещать), *вынести постановление* (решить, постановить), *находиться в собственности* (принадлежать). Причем не всегда глагол в таких сочетаниях стоит в инфинитиве: ср. *влияние, уважение, которым пользуется кто-либо* (престиж), где полусвязочный глагол, свободно допускающий замену связками *иметь, обладать*, употреблен в личной форме. Рассматривать при семантической записи смысл подобных сочетаний как сумму смыслов составляющих их элементов было бы ошибкой, поскольку, например, *обратить внимание* означает не факт «обращения чего-то» плюс факт «внимания», а просто наличие «вни-

<sup>27</sup> Грязнухина Т. А., Пшеничная Л. д., Скороходъко Э. Ф. Система информационного поиска. Киев, 1964; Информационно-поисковая система «БИТ». Киев, 1968.

<sup>28</sup> Пшеничная Л. Э. Тезаурус в документальной ИПС. Киев, 1977, с. 8.

<sup>29</sup> Пшеничная Л. д. Указ. соч., с. 50.

мания», аналогично *прийти к соглашению* не складывается из факта «прихода, прибытия куда-то или к чему-то» и факта «соглашения», а лишь фиксирует наличие «согласия». Поэтому во всех приведенных случаях следует говорить не о сумме, а о разности компонентов дефиниции, исходя из того, что семантика указанных сочетаний без всякого ущерба, без каких-либо потерь передается самим существительным, которое, кстати, чаще всего оказывается одного корня с глаголом, синонимичным всему сочетанию (ср. *давать совет — советовать, оказать помощь — помочь* и т. п.).

Похожие сочетания глаголы могут иногда образовывать также с прилагательными или причастиями — *застать быть неподвижным* (пригвоздить), *оказаться уличенным* (попасться), их также могут давать два существительных — *отрезок времени* (поздно), *для несения службы* (прислить), *на протяжении длительного времени* (продлиться), и смысл их результируется в последних элементах — *неподвижный, уличенный, время, служба, длительное время*. Такие фрагменты дефиниции приближаются по структуре к устойчивым сочетаниям, которые в правой части словаря встречаются, вообще говоря, не часто и которые в отличие от рассмотренных выше требуют при своей семантической записи либо полной фиксации их составляющих (пусть в условной форме, в виде одного «слова», ср. *обмен веществ — • «обмев», высшее учебное заведение — «вуз», образ действий — «образдей* и т. п.), либо передачи их приблизительного, неполного смысла с помощью одного из составляющих слов (ср. *точка опоры* — опора, *линия фронта* — фронт, *холодное оружие* — оружие, *средняя школа* — школа, *образ мыслей* — мысли). Если разность компонентов для сочетаний, не имеющих характера устойчивых (*оказаться уличенным*), выражается их последним элементом и при исключении других заметной утраты общего смысла как будто не происходит, то в случае устойчивых сочетаний нельзя вывести формального правила о том, какой элемент целесообразнее исключить для компрессии и унитермного представления семантической записи, с тем чтобы потеря в семантике была минимальной. Иногда опускаются первые элементы, независимо от их категориального статуса (существительные, прилагательные и т. д.), как в приведенных примерах, в иных случаях это могут быть вторые элементы: *под открытым небом* (*рынок*) — открытый,

*дыхательных путей* (кашель) — дыхательный, *слизистая оболочка* (полип) — слизистый, *населенный пункт* (селиТЬ) — населенный, *безвыходное положение* (резать) — безвыходный<sup>30</sup>.

Таким образом, не суммой, а разностью смыслов образующих его слов характеризуется в ряде случаев лексический компонент в семантике дефиниции, и общее определение ситуации разности составляющих может быть дано только как эмпирическое обобщение, с указанием полного списка полу связочных глаголов, типов сочетающихся с ними существительных, а также подлежащих исключению имен расширенной семантики в парах существительных (типа *путь, отрезок, пункт, точка, положение*). Значит, эта ситуация представляет собой уже неопределенность иного рода, чем рассмотренная выше, а именно, неопределенность из-за сложности. Число слов, как подлежащих толкованию, так и используемых в дефинициях, хотя и очень велико, но все же конечно; однако выявить исчерпывающим образом все подобные ситуации еще до кодирования, до семантической записи, т. е. по определению, представляется крайне затруднительным, и не только из-за количества единиц, а главным образом из-за противоречивости их употребления. Нельзя, например, сказать, что слова *путь* или *отрезок* во всех случаях должны быть вычитаемыми, а не слагаемыми, тем более, что они не входят в число «нулевых», подлежащих обязательному исключению семантических множителей и их частоты на массиве дескрипторов равны соответственно 3 и 0, т. е. находятся в пределах релевантности частотного селективного параметра<sup>31</sup>. Аналогично обстоит дело и с полу связочными

<sup>30</sup> Здесь стебит, вероятно, отметить аналогию используемого нами методического приема дизъюнкции с объективным процессом семантического стяжения, имеющим место на разных этапах развития языка. Так, специальный анализ показал, например, что социальная значимость распространившегося в недавнее время словосочетания *разрядка международной напряженности*, частота его использования и общепонятность привели к тому, что ради краткости взамен трехчленного стали употреблять двучленное сочетание *разрядка напряженности*, а затем и просто — *разрядка*. Таким образом, ключевое слово взяло на себя всю смысловую нагрузку выражения (см.: Поповцева Т. Н. О формировании новых значений слов путем семантического стяжения (на примере существительного *разряда*). — В кн.: Новые слова и словари новых слов. Л., 1978).

<sup>31</sup> Караполов Ю. Н. Частотный словарь семантических множителей, , ,

глаголами. Так, *оказаться* (част. 1) в толкованиях разных значений одного и того же слова *попасться* выступает в разных качествах: 1. *Оказаться где-н., в каких-н. условиях, обстоятельствах* (здесь — как полнозначное слово). 2. *Оказаться уличенным в чем-н.* (как полусвязочный глагол). Эти трудности и вызывают необходимость включения эвристического момента в алгоритм кодирования (см. выше, § 6).

### **Гиперсемантизация и разновидности избыточных признаков**

Непертинентная информация в дефиниции возникает не только из-за включения в последнюю некоторых индивидуальных характеристик и особенностей предмета или явления — вместо его видовых признаков. Гиперсемантизация и избыточность (подчеркнем еще раз, что речь идет не об избыточности вообще, а об избыточности для представления в тезаурусе) являются в ряде случаев следствием использования в словарной статье энциклопедических данных, относящихся к «технологии» изготовления или применения чего-то, описанию последовательности действий при функционировании или осуществлении чего-то, иногда из-за спорадического включения в толкования элементов научной или технической терминологии, которые, будучи лишены системных связей в метаязыке филологического словаря, повисают в воздухе, не имеют точки соотнесенности в мыслимом или реальном тезаурусе литературного языка, на который должны быть ориентированы дефиниции толкового словаря. Причем, как это ни странно, сказанное касается толкований самых простых слов. Рассмотрим соответствующие примеры.

*Пирог — печеное изделие из тонко раскатанного теста с начинкой.* Взятое в отдельности, само по себе, это определение едва ли может вызвать какие-либо возражения или недоумения: оно является достаточно компактным и информативным. При его семантической записи мы, казалось бы, должны только опустить одно знаменательное слово — *изделие*, указывающее на родовую принадлежность («элемент класса изделий») и несущее минимум информации (ср. возможность его замены словами *продукт, вещь, предмет, нечто*). Однако при попытке соотнести каждое из оставшихся слов с мысленным тезаурусом мы сразу же сталкиваемся с невозможностью убе-

дителью и однозначно сделать это для элементов *тонко* и *раскатанный*. Понятие *тонкий* является нечетким, расплывчатым, даже по отношению к «тесту для пирогов» (ср. пирог-кулебяку, нижний слой теста для которого делается в несколько раз толще верхней корочки, и обычный, слепленный пирожок, у которого верхняя и нижняя части примерно одинаковы), а кроме того, оно ничему не противостоит, так как трудно представить себе пирог из «толсто раскатанного» теста. Что касается глагола *раскатать*, то он, вероятно, войдет в воображаемую или реальную дескрипторную статью *тесто* в числе действий, имеющих объектом предмет, обозначенный заглавным словом, а поскольку понятие *тесто* есть презумпция понятия пирог, то относить глагол *раскатать* и к последнему нет уже никакой необходимости. Таким образом, элемент данной дефиниции *тонко раскатанное* несет избыточную, не нужную для тезаурусной идентификации слова *пирог* информацию, учет которой в семантической записи приведет к «шуму», к неправильному дополнительному распределению рассматриваемого слова также и по понятиям, обозначающим нечто *тонкое* или связанным с *катанием*. Уже само осмысление этого непертипентного элемента дефиниции заставило нас входить в такие детали технологии, которые не могут относиться к компетенции толкового филологического словаря, а должны иметь место в специальном кулинарном или энциклопедическом справочнике, а значит, в лексическое слагаемое семантики дефиниции их включать нельзя.

Другой пример: *Полоскать — промывать после стирки, погружая в чистую воду*, — где после исключения предлогов в семантической записи неясным остается смысл компонента *погружать*: если он относится к описанию последовательности действий при полоскании, то с ним должен был бы соседствовать еще один глагол типа *поднимать, вынимать*. Но даже при такой логической последовательности операций включение подобных элементов в дефиницию столь простого действия, как *полоскать*, нельзя признать обоснованным, поскольку тест этой дефиниции на тезаурус сразу же показывает их несостоятельность, относя входной глагол *полоскать*, помимо его естественных дескрипторов *стирка* или *чистый*, также к дескрипторам *погружение* (или *опускание*) и *глубина*. Аналогичным образом обстоит дело с толкованием глагола *жевать — растирать зубами и движсениями языка*.

Во-первых, едва ли языком можно *растирать* твердую пищу, а во-вторых, более обобщенное понятие *движение* — это настолько отдаленный род по отношению к жеванию (ср. жевание — как элемент из класса движений), который уже не может нести никакой полезной информации. К подобным просчетам приводит стремление описать в филологическом словаре механизм того или иного действия, тогда как в обычном случае пользователь хотел бы найти там прежде всего сведения, необходимые для идентификации самого действия и, возможно, для характеристики его последствий. Ср.: *Взрыв — сопровождающееся сильным звуком воспламенение чего-н. вследствие химического разложения вещества* (в других случаях — химической реакции. — Ю. К.) и *образования сильно нагретых газов*, где слова, указывающие причину взрыва (*химическое разложение вещества*) явно избыточны. То же можно сказать о сочетании *канал ствола* в дефиниции — и без того обстоятельной — слова *выстрел*: *Выстрел — взрыв заряда в канале ствола огнестрельного оружия, выбрасывающий пулю, снаряд на определенную дальность, а также звук этого взрыва. Ствол*, вообще говоря, имеет семантическую связь со словом *выстрел* (как деталь или часть оружия, из которого выстрел производится), хотя эта связь не непосредственная, а значит, довольно слабая и для семантического поля, возможно, нерелевантная. С формальной точки зрения идеальным решением казалось бы сохранение в семантической записи всего терминологического сочетания *канал ствола*, но это противоречило бы нашему принципу унитермности кодирования и распределения, а кроме того, вносило бы в набор семантических множителей и фактическую ошибку, поскольку взрыв заряда все-таки происходит в гильзе патрона, а уж потом сильно нагретые газы выбрасывают пулю по каналу ствола. Значит, в содержательно-семантическом плане последнее сочетание оказывается излишним и его следует опустить.

Наибольшие трудности составителям словарей доставляет построение толкований общенаучных терминов и понятий, ставших постоянным достоянием общелiterатурного языка, именно потому, что в таких дефинициях нельзя обойтись без использования специальной терминологии, а последняя, как было сказано, не имеет опоры в стоящем за системой дефиниций филологическом по природе тезаурусе. Отсюда естественное стремление сократить эту

специальную терминологию в правой части словаря, что порождает такие усложненные определения: *Вычитание — обратное сложению математическое действие, посредством которого из двух чисел (или величин) получается третье, которое, будучи сложенным со вторым, дает первое* или *Вычитание — одно из четырех арифметических действий, посредством которого по двум числам отыскивается третье, дающее в сумме со вторым первое*. Здесь попытка обойтись минимумом терминов (используется лишь антоним определяемого — *сумма* или *сложение*) приводит к нагромождению последовательностей из первого, второго, третьего и даже четвертого (одно из четырех арифметических действий), пробраться сквозь которые пользователь словарем оказывается в состоянии лишь благодаря своим школьным познаниям о смысле арифметических действий, а соотнесение которых с тезаурусом в качестве семантических множителей совершенно невозможно. Думается, что естественнее в таком определении было бы опереться на понятие не *сумма*, а *разность*, тем более, что терминологическое значение этого слова отмечено в том же словаре. Ту же картину можно наблюдать в «процедурном» определении значения слова *мягкий* по отношению к звуку речи (*Мягкий — о согласных звуках: произносимый с приближением средней части языка к твердому нёбу*), которое при отсутствии разъяснений в словаре по поводу таких понятий, как *твердое нёбо* или *средняя часть языка*, может быть расшифровано только лингвистом, а для своей тезаурусной классификации нуждается в дополнительной информации. Но чаще все же бывают ситуации, когда дефиниции страдают не от недостатка, а от избытка признаков и когда включение термина или терминологического сочетания в толкование становится излишеством, затрудняющим системную квалификацию и тезаурусную привязку определяемого слова. Ср.: *Плоскогорье — местность, лежащая высокою над уровнем моря, с равнинной или холмистой поверхностью*, где компонент *над уровнем моря*, не поддающийся универсальному представлению в семантической записи, должен быть исключен, тем более, что он нигде больше не встречается в дефинициях избранного словарника и реально ничего не добавляет в данном случае к семантике хотя и нечеткого, но содержательного понятия *высоко*.

Смысл наших рассуждений по поводу лексического компонента в семантике словарной дефиниции сводится,

таким образом, к тому, что его нельзя интерпретировать как простую сумму (конъюнкцию) составляющих знаменательных лексем, а необходимо вводить поправку, учитывающую условия истинности дефинитивного предложения-текста в виде его ориентированности на тезаурус. При внесении такой поправки лексическое слагаемое и предстает иногда как разность «семантических множителей», т. е. требует опущения некоторых знаменательных слов при получении семантической записи текста дефиниции.

Те лексические элементы дефиниции, которые не обладают дифференцирующей функцией, т. е. неспособны однозначно соотносить определяемое слово с тем или иным дескриптором или семантическим полем, будем в дальнейшем называть избыточными семантическими множителями. Такое название призвано подчеркнуть ту особенность их положения в дефинитивном предложении-тексте, что, будучи с одной стороны необходимыми элементами самого создания, построения взятого в отдельности этого текста и его восприятия как единичного произведения, они являются вместе с тем семантически внесистемными, их включение в дефиницию продиктовано либо формальными синтаксико-семантическими закономерностями сочетаемости (полусвязочные глаголы), либо субъективными моментами сиюминутного семантического наполнения, кажущегося оживления, псевдоактуализации данного речевого произведения. В этом смысле их семантическая значимость не выходит за рамки отдельной дефиниции, и одни и то же, казалось бы, избыточные семантические множители оказываются несопоставимыми друг с другом в разных дефинициях. Сказанное относится прежде всего к многочисленной группе качественных (по преимуществу) прилагательных субъективной оценки (*крепкий*, *высокий*, *прямой*, *тонкий*), которые почти всегда выступают соотносительными парами (т. е., наряду с названными, также *слабый*, *низкий*, *кривой*, *толстый*). Ср. употребление в реальных дефинициях пары *узкий—широкий*: *Половик* — плетеный или сшитый из разноцветных обрезков ткани *узкий длинный коврик, постилаемый на пол*; *Прилавок* — род узкого, закрытого спереди стола в магазине. . .; *Полоса* — длинный, ровный след. . . лента, длинная узкая часть какого-н. пространства; *Гряда* — узкая полоса вспаханной земли в огороде или цветнике; *Лента* — узкая полоса ткани; *Ленты* — широкая тесьма; *Лопата* —

*орудие для копания земли с рукояткой и широким плоским отточенным концом; Рубанок — столярный инструмент для гладкого строгания дерева в виде деревянной колодки с вставленным в нее наклонно широким лезвием (железком); Барабан — ударный музыкальный инструмент в виде широкого цилиндра, с обеих сторон обтянутого кожей. Полотенце — в одном случае узкая и длинная полоса ткани, преимущ. льняной или бумажной, для вытирания частей тела или посуды, а в другом случае — неширокий и удлиненный кусок ткани для вытирания чего-л. и т. д. Условность употребления подобного рода оценок становится очевидной при их сопоставлении друг с другом: гряда — узкая, а лопата — широкая, лента может быть представлена либо как узкая, либо как широкая, в зависимости от выбранного эталона (тесьма либо полоса), полотенце — либо узкое, либо неширокое, прилавок — узкий, а лезвие рубанка — широкое и т. п.*

Таким образом, имея в виду предпосылку тезаурусной (Ориентированности дефиниции, мы констатируем семантическую нейтральность избыточных множителей в ней, которые практически ни на что за пределами дефиниции не ориентированы и являются в пей средством гиперсемантизации. Конечно, можно предположить наличие в каждом случае некоторого интуитивного эталона, некоторого масштаба сравнения вне дефиниции, и специальный анализ может иногда этот масштаб выявить<sup>32</sup>. Однако решающим моментом, выводящим избыточные лексические элементы за скобки дифференцирующей семантики определяемого слова, является, на наш взгляд, наличие скрытого, замыкающегося в рамках единичной дефиниции противопоставления при употреблении прилагательного

---

<sup>32</sup> Так, по наблюдениям Р. В. Мадояна, подтвержденным па материале словарей разных индоевропейских языков, эталоном для отнесения домашней птицы к крупной разновидности является курица: то, что равно ей или превосходит ее по величине, расщепляется как крупное, большое. Для диких же птиц такой моделью является воробей: то, что по величине соизмеримо с ним, — мелкое, небольшое (например, скворец); то, что значительно превосходит, — большое, крупное (например, ворона). Аналогичные соотношения для квалификации на большое и маленькое устанавливаются для диких животных, с одной стороны, и для домашних — с другой. См.: *Мадоян Р.* Теория идеографии и методические приемы идеографической классификации лексики. — Изв. Ереванского ун-та. Сер. обществ. наук, 1979, № 30.

субъективной оценки. Второй член этого противопоставления иногда присутствует в явном виде (*Прут — тонкий металлический стержень, кусок толстой проволоки*), чаще же бывает не назван, но обязательно подразумевается, и поэтому употребление в дефиниции прилагательного субъективной оценки есть всегда сигнал передачи им некоторого оттенения. В самом деле, конец лопаты рассматривается как широкий по отношению к рукоятке, прилавок узок по отношению к его длине, а лезвие рубанка является широким по сравнению с деревянной колодкой, почти совпадая с ней по ширине. Ср. еще несколько определений, в которых прилагательное длинный характеризует относительные размеры частей, деталей описываемого предмета, т. е. выступает как специфический способ передачи некоторых отношений локального масштаба: *Половик* — ... длинный коврик; *Посох* — длинная трость; *Гитара* — род струнного щипкового музыкального инструмента... с длинным грифом; *Лягушка* — бесхвостое земноводное с длинными задними ногами...; *Бороздка* — длинное мелкое углубление, складка; *Зонтик* — приспособление для защиты от дождя или солнца... с длинной ручкой; *Спица* — длинная тупая игла для вязания и т. д.

Рассмотренные примеры подтверждают важность и необходимость использования такого рода семантических множителей при определении и характеристике индивидуальных явлений, по в то же время показывают и условность подобных характеристик, относительную значимость оценок, не позволяющую придавать их носителям дифференцирующую семантическую роль в масштабе всей лексической системы, всего словаря. Если мы определяем мыло как твердый кусок вещества, то при этом отдаем себе отчет, что твердость в данном случае совсем не та, какой обладает, например, древесина (плотная твердая часть дерева...) или сталь (твердый серебристый металл...). Все дело в отсутствии универсальных критериев, общезначимых эталонов употребления прилагательных субъективной оценки. Эти эталоны не только варьируются для разных групп явлений, но зачастую оказываются одноразовыми, эталонами, так сказать, на случай: для посоха это трость, для задних ног лягушки это, очевидно, передние, для спицы — игла, а для кинжала (... с коротким лезвием), вероятно, шпага. Это, конечно, придает выразительность и нестандартность отдельной дефиниции, на-

полняет ее живой плотью, вносит творческий момент в ее восприятие<sup>33</sup>, но одновременно препятствует системной идентификации признаков и свойств определяемого явления. Эти особенности и позволяют нам говорить об избыточности указанных лексических компонентов дефиниции, а значит, необязательности их включения в число кодов при получении ее семантической записи.

Сделанный вывод оформился у автора сначала интуитивно, в ходе семантического кодирования исходного для тезауруса словаря, но затем он получил и формальное подтверждение, когда оказалось, что все интуитивно относимые к избыточным семантические множители в частотном списке попали в зону высокой частотности (встречаемость на массиве дескрипторов  $\geqslant 6$ ), т. е. зону нерелевантности их как дифференцирующих элементов. Таким образом, выявилась возможность формализовать в алгоритме кодирования процедуру исключения из семантической записи большой группы избыточных множителей, их «нулевизации», задав их списком.

Приводимый ниже список содержит около 120 слов такого рода, расположенных по убывающей частоте парами, составленными из единиц с противоположным значением, причем первым в каждой паре названо слово с большей частотностью. Поскольку подсчитывались слова в их орфографической форме (точнее сказать, множители), а не значения, в которых эти слова фигурировали в том или ином контексте, т. е. многозначность не учитывалась, одно и то же слово в нашем списке может входить в несколько антонимических пар. Однако противопоставление, однажды уже упомянутое в списке (напр., *твёрдый — мягкий*), при перевертывании его составляющих не повторяется, так как частотность его второго члена меньше частотности первого, а принятый порядок предполагает уменьшение частотности в парах слева направо. Поэтому мы встречаем сначала *твёрдый — мягкий*, *хороший — плохой*, а в соответствующем месте, ниже по списку: *мягкий — резкий*, *плохой — благоприятный* и т. п.

---

<sup>33</sup> С некоторой долей условности можно даже сказать, что прилагательные субъективной оценки приближаются здесь по своей функции к дейктическим словам, придавая тем самым использующим их дефинициям оценочный или полуоценочный характер.

жидкий (44)	— густой (10)	легкий (13)	— трудный (7)
твёрдый (36)	— мягкий (18)	плотный (13)	— упругий (8)
	— нетвёрдый (8)	свой (13)	— чужой (8)
	— упругий (8)	сладкий (13)	— соленый (7) <sup>35</sup>
плоский (33)	— выпуклый (7)	слоистый (13)	— однородный (8)
высокий (28)	— низкий (11)	хороший (13)	— плохой (10)
	— маленький (13)	прямой (12)	— боковой (3)
	— небольшой (13)		— кривой (7)
	— незначитель- ный (11)	быстрый (11)	— медленный (8)
большой (27)	— маленький (13)		— тихий (7)
тонкий (27)	— толстый (7)	искусствен- ный (11)	— настоящий (7)
	— грубый (13)		
тяжёлый (23,	— лёгкий (13)	красный (11)	= ? зелёный (8)
сильный (22)	— слабый (14)	теплый (11)	= холодный (7)
белый (21)	— чёрный (11)		— умеренный (7)
	— красный (11)	узкий (11)	= широкий (9)
светлый (19)	— темный (3)	машинный (10)	= ручной (7)
мягкий (18)	— резкий (11)	нравственный (10)	— безнравствен- ный (7)
	— упругий (8)	плохой (10)	= благоприят- ный (7)
	— строгий (7)		
целый (18)	— полый (7)	свободный (10)	= тесный (8)
ценный (18)	— ? обычный (8)		= несвободный (6)
короткий (17)	— длинный (11)	отрицатель- ный (8)	= положитель- ный (7)
верхний (16)	— нижний (9)	простой (9)	= сложный (8)
мелкий (15)	— глубокий (9)	обычный (8)	= особенный (7)
	— ? крупный (3) <sup>34</sup>	основной (8)	= ? боковой (8)
специаль- ный (15)	— обычный (8)	полный (8)	= • пустой (7)
необычный (14)		ровный (8)	= кривой (7)
слабый (14)	— крепкий (10)	яркий (8)	= серый (8)
близкий (13)	— далекий (17)	вертикальный (7)	= горизонталь- ный (7)
грубый (13)	— гладкий (8)	равный (7)	= разный (7)
	— точный (7)		

Кроме того, несколько высокочастотных избыточных множителей не обнаружили с отчетливостью своей пары, либо такая пара относится в частотном списке к зоне релевантности семантических множителей:

больной (31)	— здоровый (6)	главный (9)	— второстепен- ный (1)
определенный (16)	— неопределён- ный (2)	гладкий (8)	— шероховатый (1)
чистый (14)	— грязный (5)	зрелый (8)	— незрелый (1)
нужный (13)	— ненужный (3)	веселый (8)	— невеселый (2)
высший (11)	— низший (5)	настоящий (7)	— прошлый (5)
новый (11)	— старый (6)	ручной (7)	— дикий (4)
частый (10)	— редкий (2)	смешной (7)	— грустный (1)
важный (9)	— неважный (1)	забавный (7)	— скучный (2)

Думается, что множители, не имеющие антонимичной пары, целесообразнее было бы считать неизбыточными, релевантными, однако **формализовать эти случаи и про-**

цедурно задать их в программе кажется весьма затруднительным. Поэтому мы сознательно идем на некоторое упрощение, огрубление приемов установления семантической связи, целиком исключая все высокочастотные избыточные множители независимо от наличия—отсутствия у них антонимической пары.

### § 10. ОТРАЖЕНИЕ В ДЕФИНИЦИИ ЛОГИКО-СЕМАНТИЧЕСКИХ ОТНОШЕНИЙ РЕАЛЬНОСТИ (РЕЛЯЦИОННОЕ ЗНАЧЕНИЕ)

Proof is often no more than lack of imagination — in providing an alternative explanation (De Bono's 2nd Law.)

*E. de Bono. Practical thinking\**

#### Типы отношений и их исчисление

Теперь можно перейти к рассмотрению второго слагаемого — скрытых в дефиниции логико-семантических отношений между предметами реального мира. В противоположность локальным отношениям, передаваемым избыточными множителями и носящим, как мы старались показать выше, окказиональный характер, релевантный лишь для данной дефиниции, логико-семантические отношения универсальны и системны. Они довольно разнообразны по своему качеству и составу и в принципе, вероятно, полностью исчислимы, если заранее условиться

---

34 Столь малая частотность (3) множителя *крупный* — на деле одного из самых распространенных — на массиве дескрипторов как раз отражает тот факт, что он с самого начала подвергался на этом массиве «нулевизации» на основе иптуитивного отнесения его к избыточным. Ср. его частоту на массиве словника (31), где при кодировании он не исключался из семантической записи. А частоты на двух массивах, как известно (см. «Частотный словарь семантических множителей. . .»), коррелируют. То же замечание касается ряда других избыточных множителей, которые, хотя и попали в зону нерелевантности, фактически обладают более высокой частотностью, так как подвергались частичной «нулевизации» (*хороший — плохой, быстрый — медленный, сильный — слабый, специальный* и др.).

35 Статистически выделяется именно эта пара, тогда как частотность более ожидаемых антонимов к слову *сладкий* оказывается ниже: *горький* (3), *кислый* (3).

\* Доказательство есть зачастую не более чем недостаток воображения в поисках альтернативного объяснения (2-й закон де Боно) {*E. de Bono. Практическое мышление*}

о пределе их членения вглубь/ В самом деле, как только мы ограничимся отношениями самого общего характера и глобального масштаба, они представляются нам вполне обозримыми. Это прежде всего субъектные отношения, характеризующиеся двунаправленностью, т. е. амбивалентностью и конверсивостью: «быть субъектом» (*Позорный — являющийся позором*) и «иметь субъектом» (*Поиск — действия ищущего...*). Далее — объектные отношения, распадающиеся на те же разновидности — «быть объектом», например, пищи (*Порей — род луковичного растения, употребляемый в пищу*) и «иметь объектом» (*Покражка — украденная вещь*). Аналогична по структуре атрибутивная связь (*Покаяние — добровольное признание в совершенном проступке, в ошибке*), где добровольность есть атрибут признания, называемого покаянием, а проступок или ошибка — его объекты; *Радиус — в математике: отрезок прямой линии, соединяющий центр с любой точкой окружности или поверхности шара*, где линия имеет атрибут *прямая*, *шар* имеет атрибут *поверхность*, а *отрезок* является субъектом соединения *центра с любой точкой...*. Ту же картину мы обнаруживаем при анализе причинно-следственной связи (*Гипноз — состояние, похожее на сон или полусон, вызываемое внушением и сопровождающееся подчинением воли спящего воле усыпляющего, а также самый способ такого внушения*), где *внушение* является причиной данного состояния, а *подчинение воли спящего воле усыпляющего* — его следствием. На два типа делятся и отношения состава, содержания — «быть составной частью» (*Полк — воинская часть, обычно входящая в состав дивизии или бригады*) и «иметь в своем составе» (*Нож — инструмент для резания, состоящий из лезвия и ручки...*; *Вода — прозрачная бесцветная жидкость, представляющая собой в чистом виде химическое соединение водорода и кислорода*). Близким к последнему, но не тождественным ему является отношение «быть элементом множества, класса» (*Водосток — сооружение для отвода, стока воды...*, т. е. элемент класса сооружений; *Веревка — изделие из крученых или витых в несколько рядов длинных прядей пеньки...*, т. е. элемент множества изделий; *Покой — состояние тишины, отдыха, бездеятельности, отсутствие беспокойства*, т. е. элемент класса состояний) и отношение «иметь элементом» (*Физиология — совокупность жизненных процессов, происходящих в организме и его органах*; *Репертуар — совокупность пьес*,

*музыкальных и иных произведений, идущих в театре. . .).* Далее следует назвать функциональное отношение (*Показной — рассчитанный на внешнее впечатление, притворный; Показательный — устроенный для всеобщегоознания, сведения; Понтона — плоскодонное судно, служащее опорой временного моста, каких-н. временных сооружений*), пространственное, т. е. «иметь местом совершения, действия» (*Полевой — находящийся или производимый в поле; Поломка — поломанное место, повреждение; Физиология — . . ., происходящих в организме и его органах*), отношение эквивалентности, или тождества, равенства, выражаемое обычно в дефиниции с помощью синонимов (*Прелесть — очарование, обаяние, привлекательность; Прель — то же, что гниль*), отношение зависимости (*Политика — деятельность. . . отражающая общественный строй и экономическую структуру страны. . . определяемая их интересами и целями*). Можно, наконец, выделить отношение как таковое, без дифференциации его природы, в самом общем виде охватывающее многие разновидности логико-семантических связей — «быть связанным с, относиться» (*Порнография — крайняя непристойность, циничность в изображении чего-н., связанного с половыми отношениями; Полярный — относящийся к полюсу, связанный с деятельностью на полюсе, около полюса*), и на этом закончить их перечисление, считая, что более частные отношения всегда могут быть подведены под один из рассмотренных их типов. Однако материал словарных дефиниций показывает, что такое подведение не всегда получается последовательным, и объективность анализа требует по мере роста материала введения все новых и новых отношений. Так, наряду с эквивалентностью появляется отношение подобия (*Полоумный — сумасшедший, слабоумный, а также подобный сумасшедшему, безрассудный; Гипноз — состояние, похожее на сон. . .*) и затем — отдаленного подобия (*Пила-рыба — хищная рыба с отростками на морде, напоминающими зубцы пилы*); наряду с функциональным — отношение «иметь целью» (*Забота — мысль или деятельность, направленная к благополучию кого-чего-н.*); наряду с атрибутивным — «иметь форму, вид» (*Гитара — струнный щипковый музыкальный инструмент с деревянным корпусом-резонатором в форме восемерки; Крест — предмет в виде стержня с перекладиной под прямым углом как символ христианского культа*). Дальнейшая детализация отно-

шений приводит к выводу о необходимости выделения также следующих их разновидностей:

иметь склонность	<i>Полемичный — склонный к полемике, свойственный полемике; Порочный — подверженный пороку, безнравственный, свидетельствующий о склонности к пороку</i>
быть чреватым, иметь потенцию	<i>Вероломный — способный нарушить свое слово, обязательство, доверие</i>
иметь отличие, отличаться	<i>Православие — христианское вероучение, сложившееся в Византии как восточнохристианская церковь, в отличие от возникшего на Западе католицизма</i>
быть противоположным	<i>Правый — находящийся, расположенный на стороне, которая противоположна левой; Полюс — нечто прямо противоположное чему-н. другому</i>
быть лишенным	<i>Прилизанный — о стиле, произведении; искусственно гладкий, лишенный яркости; Улица — среда, лишенная культурного воздействия общества, семьи</i>
быть заслуживающим	<i>Презренный — заслуживающий презрения, вызывающий его; Пресловутый — широко известный, наушивший, но сомнительный, или заслуживающий отрицательной оценки</i>
выражать и выражаться в ...	<i>Честный — проникнутый искренностью и прямотой, добросовестный; выраждающий такие свойства; Понос — расстройство деятельности кишечника, проявляющееся в частых и жидких испражнениях</i>
изготовленный из . . . и (используемый) для изготовления	<i>Шелк — нити, изготавливаемые из выделений гусеницы шелкопряда; Порох — взрывчатое вещество для изготовления зарядов огнестрельного оружия</i>
иметь условием, предпосылкой	<i>Квалифицированный — требующий специальных знаний</i>
иметь порядковый номер	<i>Деление — обратное умножению математическое действие, посредством которого из двух чисел (или величин) получается третье, которое, будучи умножено на второе, дает первое</i>
быть другим	<i>Вычитаемое — число или выражение, которое вычитается из другого</i>
иметь содержанием	<i>Преступный — содержащий преступление, являющийся преступлением; Полный — содержащий в себе что-н. до своих пределов, замещенный целиком</i>

Здесь в общей сложности перечислено 32 разновидности наиболее показательных, как нам представляется, отношений, извлеченных из дефинитивных текстов филологического словаря. Они, естественно, не исчерпывают

всего их многообразия: даже в рамках того ограниченного 10 тысячами единиц словника, который послужил основой русского тезауруса (см. главу III), их оказывается гораздо больше<sup>36</sup>, и по мере продвижения по материалу в процессе семантического кодирования появляются отношения «быть сопутствующим» и «иметь в качестве сопутствующего», «быть подверженным», далее — целый комплекс временных отношений, набор отношений степени (менее, не очень, более, слишком. . .) и т. п. Отношения начинают дробиться, множиться и создается впечатление, будто число их необозримо и они имеют тенденцию к регрессу в бесконечность, что, вообще говоря, противоречит здравому смыслу и нашим представлениям о конечности познанного и фиксированного в языке мира, т. е. о конечности наших знаний. Иными словами, мы исходим из убежденности, что отношения должны быть исчислимы.

На пути исчисления логико-семантических отношений в дефиниции мы встречаемся с двумя трудностями. Во-первых, это возможная неоднозначность идентификации, когда с равными основаниями одно и то же отношение может быть квалифицировано как принадлежащее к разным типам. Так, в определении *Полярный — относящийся к полюсу, связанный с деятельностью на полюсе, около полюса*, приведенном в качестве иллюстрации отношения «связанный с», последнее можно трактовать и как пространственное; отношение «иметь в своем составе» в определении *Нож — инструмент для резания, состоящий из лезвия и ручки.* . . допускает и атрибутивную его интерпретацию, и т. д. Во-вторых, это специфика отражения самих отношений в дефинициях как особого типа текстах, связанная с известной редуцированностью последних, стремлением к максимальной компактности при их формулировании. Отсюда — продиктованное этим стремлением исключение в ряде случаев второго члена отношения (ведь отношение — это всегда отношение между чем-то и чем-то, между двумя «вещами» или двумя «свойствами», представленными как «вещи») из дефиниции или замена его некоторыми обобщенными символами, не имеющими реальной дифференцирующей силы, а потому и противоречащими самой сути дефиниции. Например, в дефиниции

---

<sup>36</sup> Для сравнения отметим, что в языке RX-кодов, использованном для семантической записи в документальной ИПС, фиксировано всего 50 отношений (*Пшеничная Л. Э. Указ. соч., с. 44*).

*Полезный* — . . . пригодный для определенной цели, идущий в дело мы можем констатировать функциональное отношение только благодаря наличию формальных его показателей — предлога для и причастия с предлогом *идущий в*, но ведь формальные показатели отношений не всегда присутствуют в дефиниции в явном виде. Что же касается вторых членов данного функционального отношения, выраженных существительными *цель* и *дело*, они представляют собой типичные «релятемы», т. е. слова, используемые в дефинициях филологических словарей, так же как причастия и предлоги, для обозначения «отношений», но не «вещей»<sup>37</sup>. Получается, что выражение отношений дублируется, они передаются дважды — и синтаксическими средствами (например, предлогами), и лексически, тогда как «вещи», между которыми устанавливаются данные отношения, в определении не фигурируют. Аналогичным образом утрачивается второй член отношения и в определении *Поле — большая ровная площадка, пространство, специально оборудованное, предназначенное для чего-н.*, где на его месте выступает отсылочный анафорический показатель, не допускающий какой-либо конкретизации на основе наличной информации. В самом деле, включение как бы с целью конкретизации слов *специально оборудованное* дает минимум сведений, лишь относя идентифицируемый предмет *поле* к классу «вещей», подвергшихся предварительной целенаправленной обработке. И однако оба этих как бы конкретизирующих слова при семантической записи данной дефиниции должны быть подвергнуты «нулевизации», не только по соображениям их высокой частотности<sup>38</sup>, но и исходя из реального употребления слова *поле*, которое показывает, что пресуппозиция *специально оборудованное* вовсе не является для него необходимой. Так, мы не погрешим против «условий истинности» использования этого слова, назвав *полем* лужайку в лесу, где мы играли в футбол во время воскресного пикника и где «воротами» нам служили положенные на землю рубашки. Здесь важен, таким образом, только функциональный аспект.

Понятно, что обе рассмотренные нами трудности в квалификации логико-семантических отношений в дефиниции так или иначе преодолимы, коль скоро мы отдаем в них себе отчет, и в конце концов можно представить, что все типы отношений, зафиксированных на ограниченном массиве слов, исчислены и классифицированы. Сам характер

и содержание всех выделенных отношений<sup>37</sup>, как свидетельствуют приведенные выше примеры, показывают, что каждое из них, будучи неотъемлемой частью определения того или иного слова, т. е. характеризуя его видовой или родовой признак, являются вместе с тем своеобразным «маркером» семантической связи данного слова с именами соответствующих семантических полей, к которым оно относится. Например, функциональное отношение в определении *Поплавок — устройство, служащее для придания плавучести чему-н.* указывает, что исходное слово обладает функциональной связью с понятиями (дескрипторами) *плавать* или *плавучий* и входит в эти семантические поля в качестве их составного элемента. То же самое следует сказать об отношении «выражаться в» в definicции *Помяться — проявить некоторую нерешительность, колебание в решении чего-н.*, которое свидетельствует, что определяемый глагол входит в группу слов («парцеллу») семантического поля **НЕРЕШИТЕЛЬНОСТЬ**, передающих свойство «способы выражения нерешительности», и т. д.

Словом, если бы мы имели полную матрицу всех отношений, мы могли бы на ее основе автоматически строить семантическое поле для любого слова. Задача состояла бы в том, чтобы по указанному отношению найти «вещь» или «вещи», которые реализовали бы его с данной, исходной

---

<sup>37</sup> Обе эти релятемы входят, естественно, в список нулевых множителей, т. е. слов, подлежащих исключению из definicции при ее семантической записи. Причем показателен такой факт: если релятена *дело*, с самого начала намеченная к безусловной нулевизации, характеризуется нулевой частотой на *обоих* массивах — и на массиве множителей дескрипторов, и на массиве множителей слов, — то релятена *цель*, нулевизация которой вначале была частичной, обнаруживает на указанных массивах соответственно частоты 15 и 62, что резко превышает эмпирический порог релевантности множителей ( $> 6$ ), т. е. подтверждает необходимость полной нулевизации и этой релятемы. См.: Караулов Ю. Н. Частотный словарь семантических множителей. . .

<sup>38</sup> Семантический множитель *спец* имеет частотность 15 и 47 на соответствующих массивах, т. е. является практически опустошенным в смысловом отношении, нерелевантным как дифференцирующий признак. Иначе обстоит дело с множителем *обору*, который подвергался лишь частичной «нулевизации» и потому был включен в список нулевых множителей, но в ряде случаев выступал в definicциях и как содержательный дифференцирующий элемент, что нашло отражение в частотном словнике, где его частота зафиксирована в пределах значимости селективного параметра — 3 и 11.

«вещью». Вероятно, ни один Дескриптор не может реализовать всего набора отношений, как и не существует такого отношения, которое было бы характерно для всех без исключения дескрипторов. Вместе с тем уже сама возможность заполнить для того или иного слова при построении его семантического поля некоторый обязательный минимум (пусть даже оцениваемый чисто количественно) отношений будет в этом случае служить показателем его потенциально дескрипторного масштаба, основанием для его использования в качестве дескриптора. В самом деле, если взять тот же глагол *помяться*, то набор отношений, который мог бы быть представлен в его поле, окажется весьма незначительным. Иное дело — существительное *нерешительность*, которое позволяет развернуть целый спектр отношений и которое вследствие этого фигурирует в роли самостоятельного концепта во всех идеографических словарях.

Подобный взгляд на способ представления логико-семантических отношений, выявляемых в словарных определениях, и их дальнейшее возможное применение к конструированию семантических полей выдвигает альтернативу широко распространенным приемам нахождения семантических связей между понятиями, между «словами» и «концептами» — по «вещам», т. е. на основе лексических единиц, являющихся носителями и выражителями некоторого «вещного», материального содержания. Возникает как будто возможность устанавливать такие связи по «отношениям». Если в гипотетическую матрицу отношений, упоминавшуюся выше, ввести информацию о способах выражения каждого из них, мы получим некоторый набор формальных правил для наполнения семантического поля относящимися к нему словами. Тогда процедура построения полей могла бы состоять в том, чтобы к некоторому заданному слову последовательно «примеривались» все отношения из матрицы и для релевантных отношений указывались бы «вещи», т. е. слова, реализующие соответствующие семантические связи. Например, для потенциального дескриптора *нерешительность* нехарактерны были бы отношения «состоять из» или «быть предназначеным (служить) для», но зато показательны отношения «быть элементом класса», которое фиксирует принадлежность его к множеству «черт человеческого характера» и ассоциирует с *мягкостью, робостью, неуверенностью, слабостью*; отношение «быть лишенным», связывающее его

**с** понятиями *твёрдость, решительность, настойчивость, упорство*; отношение «выражаться в», которое требует включения в поле слов *размыщение, колебание, сомнение* с соответствующими глаголами, в том числе глаголом *помянуться* и т. д.

На таком пути можно предвидеть одну серьезную трудность: вводя в гипотетической матрице специальную графу, где указаны способы выражения отношений, мы не всегда будем иметь возможность эту графу заполнить. Ведь отношения, как было отмечено выше, передаются не только лексическими средствами, но и синтаксическими, в числе которых, помимо предлогов, также и порядок слов, и морфологические показатели синтаксической зависимости, которые не поддаются однозначной фиксации. Тем самым мы лишаемся критерия надежности матрицы, поскольку она не дает формальной опоры при оценке релевантности того или иного отношения для испытуемого концепта, не обеспечивает в каждом случае исчерпывающего перебора способов выражения отношений. Таким образом, рассматривая альтернативу «от вещей к отношениям» при построении семантического поля или «от отношений — к вещам», мы должны без колебаний отбросить вторую возможность и выбрать первую. Тем не менее с точки зрения «вклада» отраженных в дефиниции логико-семантических отношений в ее общую семантику способы их выражения следует рассмотреть более подробно.

### **Способы выражения отношений и их словарная фиксация**

Как показывает анализ словарных определений, большая часть отношений в них передается лексическими средствами, и распределение этих средств по частям речи следующее: из 160 семантических множителей, признанных нулевыми по определению<sup>39</sup>, 64 составили имена существительные, 62 — причастия, 26 — глаголы и 8 —

<sup>39</sup> Полный их список приведен в нашем «Частотном словаре семантических множителей русского языка», но при обращении к нему нужно иметь в виду, что в список не вошли предлоги и наречия степени, которых мы частично коснулись выше. Статистический анализ, подтвердив релятивный характер, так сказать, исконно нулевых семантических множителей (см. выше, § 7), вместе с тем значительно расширил их список, особенно в части, касающейся существительных и прилагательных, и образцы этих дополнений будут приводиться по ходу раскрытия состава каждой из групп.

прилагательные и наречия. Среди существительных можно выделить довольно большое число разнообразных групп, одну из которых образуют слова, различающиеся не характером выражаемого ими отношения (поскольку отношение здесь одно и то же — «быть элементом класса»), а типом самого класса или множества. Прежде всего это множество «вещей», под которое в силу специфики словарных определений подводятся те факты, явления и предметы реального мира, которые трудно поддаются дифференциации на небольшом пространстве дефинитивного текста. Лучшим обозначением этого класса было бы, вероятно, «чего», которое в реальных определениях передается словами *вещь, предмет, тело* и их контекстуальными синонимами *начало* (*Зараза — болезнестворное начало, распространяющее микробами*), *образование* (*Волос — роговое нитевидное образование.* . .) и др. Далее следует указать два в известной мере противостоящих друг другу класса «вещей» — абстрактных, таких как дело (*Искусство — умение, мастерство, знание дела*), вопрос (*Позиция — точка зрения, мнение в какой-н. вопросе*), область (*Познания — совокупность знаний в какой-н. области*), обстоятельство, и т. п., а с другой стороны, конкретных, а именно класса, подчеркивающего принадлежность входящих в него «вещей» к артефактам: изделие (напр., *медальон*), устройство (напр., *поплавок*), приспособление (*щетка*), сооружение (*трубопровод*), механизм, прибор, продукт, товар и некот. др. Параллельная линия конкретизации относит некоторые из «вещей» к классу веществ с выделением подклассов: металл, вода, жидкость, минерал, воздух, газ, масло и т. п., причем все перечисленные множители оказываются нулевыми, и их релятивный характер отражается в показателях их частотности, которые всегда превышают пороговую, критическую для дифференциального элемента частоту. Далее, необходимо отметить набор множителей, используемых в дефинициях для называния различных подмножеств живых существ: *существо, животное, человек, лицо, организм, насекомое, млекопитающее*. Наконец, последний способ передачи рассматриваемого отношения «быть элементом или представителем класса» реализуется в дефинициях с помощью высокочастотных, т. е. нулевых, множителей *вид, род, сорт, тип, семейство, разновидность, представитель, элемент*.

Выражение противоположного отношения «быть множеством элементов» осуществляется в словарях нулевыми

множителями — совокупность, система, цикл, группа, собрание, ряд, тогда как примыкающее, казалось бы, к этому перечню существительное серия уже имеет частоты в пределах релевантности селективного параметра (1 и 3). Ср.: Религия — одна из форм общественного сознания — совокупность мистических представлений. . . Гигиена — раздел медицины, изучающий условия сохранения здоровья, а также система действий, мероприятий, направленных на поддержание чистоты, здоровья. Математика — цикл наук, изучающих величины, количественные отношения, а также пространственные формы. Широко представленное в дефинициях отношение «быть частью» обозначается главным образом одним из самых частотных множителей часть (Рукав — часть одежды. . ., Стена — часть здания. . .), а также его менее частотными, но все же находящимися за пределами зоны дифференцирующей релевантности множителей, синонимами: принадлежность (чулок), кусок (полотенце), отдел (хирургия), раздел (геометрия), отрасль. Самостоятельную и многочисленную группу составляют существительные — носители отношения «местонахождение». Сюда помимо наиболее частотных место и местность входят также нулевые (высокочастотные) пространство, помещение, расположение, вместеище, участок, край, граница, предел, центр, поверхность, сторона, верх, низ, которые дополняются набором прилагательных — носителей того же отношения: внутренний, внешний, пространственный, наружный, промежуточный, верхний, нижний, центральный, боковой. Особо следует выделить группу множителей, квалифицирующих общие, в некотором смысле неопределяемые, и абстрактные понятия: материя, энергия, действительность; категория; качество с синонимами свойство, признак, характеристика, черта; отношение; число и количество; уровень, степень; время с синонимами момент, период; явление, событие факт, процесс; форма, вид, фигура, знак, изображение.

И заканчивая рассмотрение существительных в роли носителей логико-семантических отношений в дефинициях, назовем несколько отдельных групп более конкретных отношений, составленных нами главным образом в результате анализа и предварительной классификации верхней части частотного списка семантических множителей<sup>40</sup>, хотя все они так или иначе представлены и в списке

---

<sup>40</sup> См.: Караполов Ю. Н. Частотный словарь семантических множителей. . .

« нулевых по определению». С учетом охарактеризованных выше, приводимый перечень можно считать указателем основных типов отношений в дефинициях толковых филологических словарей. Множитель, стоящий первым в каждой группе, выступает как ее символ, имя группы. В силу немногочисленности и, так сказать, вторичности в этой роли прилагательных и глаголов представляется целесообразным включить в данное перечисление носителей отношений эти части речи в скобках, под соответствующими названиями групп. Что касается причастий, то они подверглись полной нулевизации в ходе семантической записи (ср. список «нулевых по определению») и потому практически не попали в частотный словарник. О них как о носителях отношений речь пойдет ниже. Итак, список отношений таков:

зависимость, влияние, воздействие, правило, принцип, закон, ритм, порядок, право, соответствие;

создание, созидание (создать), творение, творчество (творить), развитие, подъем, строить, построить, возникать, объединять, соединять;

движение, двигаться, ход, перемещение (перемещаться), течение, изменение (изменяться), колебание, передвижение (передвигаться), распространение (распространяться);

сохранение (сохранять), предохранение (предохранять), охрана (охранять), хранение (хранить), соблюдать (соблюдать), защита (защищать), поддерживать, удовлетворять, постоянный, норма, нормальный;

деятельность, действие (действовать), поведение, организация (организовывать), труд (трудиться), работа (работать), делать;

изготовление (изготовить), производство (производить), обработка (обрабатывать), воспроизводить, приготовить, мерить, измерение (измерять), готовить, готовый;

отрицание (отрицать, отрицательный), разложение, разрушение, нарушение (нарушать), уничтожение (уничтожать), лишаться, избавиться, затруднение, недостаток; несоответствие, ненормальный, недействительный, неправильный, беспорядочный, не имеющий (не иметь), не прямой, несущественный;

сознание (сознавать), способность (способный), мысль (мыслить), ощущение (ощущать), восприятие (воспринимать), сомнение (сомневаться), смысл, понять, понимание;

условие, требование (требовать), потребность, необхо-

димость (необходимый), возможность (возможный), опора, основа;

средство, способ, путь, инструмент, орган;

сила, власть, управление (управлять), государство (государственный), предписывать;

направление (направлять, направляться), расстояние, путь, протяженность (протяженный), линия, величина, размер;

другой, отличаться (отличный), разный.

При знакомстве с этим списком обращает на себя внимание его соответствие, а местами буквальное совпадение с наполнением структуры фасетно-блочного тезауруса. Как правило, тематическая область в таком тезаурусе делится на две большие зоны — предметные фасеты и процессуальные фасеты. В число первых включаются группа «вещей» и группа «идей» (абстракций), которым в нашей классификации отвечают, с одной стороны, имена классов, воплощающих отношения «быть элементом класса, множества» (ср. вещь, предмет, дело, вопрос, обстоятельство, изделие, приспособление, прибор и т. п.) и «быть множеством элементов» (ср. совокупность, система, цикл и т. д.), а с другой, абстракции — категория, качество, отношение, количество, уровень, форма. Эти названия множеств, классов и их элементов выступают, таким образом, в роли блокообразующих понятий. Зона же процессуальных фасетов делится па группу технологических процессов и группу явлений, различия между которыми определяются наличием в первой активного субъекта действия. Приведем для сравнения структуру двух последних групп, где А обозначает принадлежность к первой из них, Б — ко второй, а цифра соответствует номеру блока и блокообразующего понятия <sup>41</sup>:

А — активный субъект

А-0 — деятельность

А-1 — созидание

А-2 — отрицание (уничтожить, лишить, т. е. созидание «наоборот»)

А-3 — перемещение

А-4 — сохранение (защита, предохранение от изменений)

А-5 — познание

---

<sup>41</sup> Соколов А. В. Методические материалы по разработке информационно-поисковых тезаурусов. Л., 1975, с. 45.

Ё — безличный фактор

Б-0 — созидание (разбухание, старение, цветение)

Б-1 — отрицание (деформация, умирание, распад)

Б-2 — перемещение (падение, дифракция, колебание)

Б-3 — зависимость (влияние, тяготение, сходство)

Б-4 — состояние (быть неподвижным. . .).

Каждая из этих групп предполагает в структуре тезауруса дальнейшую конкретизацию: первая — с помощью творительного фасета, т. е. набора инструментальных блоков, и фасета характеристик, т. е. набора блоков, раскрывающих условия и обстоятельства технологического процесса, вторая же — только с помощью характеризующего фасета, поскольку явление не предполагает наличия инструмента.

Итак, сравнение выделенных в словарных дефинициях типов отношений с блокообразующими понятиями и их структурой, т. е. фабулой фасетно-блочного тезауруса, дает нам еще раз возможность убедиться, что принцип ориентированности на тезаурис и объективные предпосылки такой ориентированности заложены в самой природе всякого толкового словаря. Подтверждается и высказанная нами ранее мысль о том, что смысл выраженных в дефиниции логико-семантических отношений не принадлежит непосредственно ее семантике, не является реальной составляющей ее собственного смысла, а есть показатель системных связей более высокого уровня обобщения. На основе этого показателя осуществляется структурирование семантических полей («блоков») и устанавливаются базовые отношения в тезаурисе. Наблюденные нами и охарактеризованные выше отношения, так же как способы их обозначения, складывались и на всем массиве толкового словаря стихийно, что лишний раз, конечно, говорит об их объективности. Одновременно эта стихийность, непрограммированность их появления и называния (т. е. использования слов естественного языка для их именования в словарных определениях) привела к известной незавершенности их системы в толковых филологических словарях, к возникновению некоторых отклонений и непоследовательностей (с одной стороны, довольно широкая и свободная синонимия в назывании отношений, когда трудно провести грань между равнозначностью и обозначением близким к «блокообразующему» словом (нулевым множителем) уже совсем иного отношения, а с другой стороны, возможность использования од-

ного и того же нулевого множителя для именования разных отношений, например вид, путь и т. п.

Из сказанного становится понятным, почему имена отношений не покрывают все 691 множитель, которые можно отнести к рангу нулевых на основе статистического критерия (частотность на массиве дескрипторов более 6) <sup>42</sup>. На деле в их числе оказываются не только так называемые избыточные множители, т. е. слова (прилагательные) субъективной оценки, не обладающие дифференцирующей силой, но составляющие как бы «плоть» дефиниции и имеющие в ней тот же статус, что интегральные фонетические признаки в составе фонем (твердый—мягкий, толстый—тонкий, большой—маленький, теплый—холодный <sup>43</sup> и т. д. — о них подробно было сказано выше), но и целый ряд имен отношений, не поддающихся без специального анализа идентификации и однозначному, непротиворечивому отнесению к той или иной их группе.

Таким образом, прямое использование эмпириически выделенных из дефиниций отношений в качестве материала для построения схемы базовых, парадигматических связей между дескрипторами тезауруса было бы, на наш взгляд, неоправданным, поскольку сами эти отношения нуждаются в логическом упорядочении, а связи между ними (т. е. схема, в которую они выстраиваются) — в достижении некоторой завершенности. Предметом особого рассмотрения должно стать и сопоставление этих отношений с нулевыми множителями разных типов в зоне высокой частотности. Но для наших целей достаточно было показать, что заключенные в дефиниции логико-семантические отношения между «вещами» реального мира, во всяком случае, как стало ясно, те из них, которые передаются именами существительными, не добавляют

---

<sup>42</sup> См.: Караплов Ю. Н. Частотный словарь семантических множителей...

<sup>43</sup> Справедливости ради следует заметить, что эти, называемые избыточными с точки зрения тезаурусной ориентированности, множители в иной системе анализа и отсчета, например, по осгудовской методике, тоже могут стать основой структурирования базовых отношений между понятиями. Ср.: Pálh Cs., Czigler J. Kerérel politikai kifejezések megítélezésére alkalmas szemantikus differenciál skálák kidolgozására. Budapest, 1976, где такое их использование для лексикометрической обработки результатов психолингвистического эксперимента вполне отвечает их субъективному характеру.

к семантике дефиниции никакой новой информации по сравнению с той, что уже содержится в ней благодаря указаниям на сами «вещи». Коль скоро «вещи» названы, то «отношения» между ними становятся само собой разумеющимися, и это следует из особого статуса именно дефинитивных текстов: их ориентированность на тезаурус с неизбежностью эксплицирует эти сконцентрированные в нем, но рассредоточенные по дефинициям базовые отношения. Такое «предсуществование» отношений позволяет в некоторых случаях в конструктивных целях исключать их из самостоятельного рассмотрения, поскольку они автоматически учитываются при простом сополагании имен «вещей». В самом деле, назовем ли мы *щетку* «приспособлением для чистки одежды» или предметом, предназначенным для того, чтобы чистить», назовем ли мы *чулок* «вязанным (либо трикотажным) изделием» или «принадлежностью одежды», для индивидуального восприятия дефиниции и идентификации соответствующей «вещи» использование различных способов обозначения одного отношения или даже констатация отличающихся друг от друга отношений при определении одной и той же «вещи» оказываются несущественными./

Эта предварительная оценка роли отношений в дефинитивных текстах проясняет обоснование использованного нами при семантической записи методического приема: для выявления тезаурусной распределенности некоторого слова первым и необходимым условием является фиксация «вещей», участвующих в его дефиниции, т. е. конъюнкция задающих их лексических компонентов. Это условие становится также и достаточным (как было в нашем случае), если логико-семантические отношения между вещами заданы априори, т. е. структура тезауруса считается заранее известной. Такая ситуация и позволила нам с самого начала исключать из семантической записи множители — нулевые по определению, т. е. носители отношений. В числе их рассмотренные нами существительные составляют значительную долю, но далеко не исчерпывают всех. Для полноты картины коснемся и других лексических способов передачи отношений между предметами в дефинициях.

Значительная роль здесь принадлежит причастиям. В функции носителя отношения причастие, как правило, вводит в предложение-дефиницию ту или иную изъяснительную конструкцию: *Положение — состояние кого-чего-*

*н., обстоятельства, сложившиеся для кого-чего-н.; Помощь — содействие кому-н. в чем-н., участие в чем-н., приносящее облегчение кому-н.; Покровительственный — оказывающий покровительство, поощрение кому-н., создающий благоприятные условия для какой-н. деятельности, и тем самым причастная форма есть уже сигнал к нулевизации множителя. Это правило работает почти без исключений, так как довольно редкие случаи употребления причастия в значении прилагательного, когда оно несет существенную смысловую информацию, легко опознаются при семантическом кодировании. См.: Пример — действие, служащее образцом для подражания, а также вообще выдающийся образец чего-н., где первое причастие оформляет функциональное отношение, тогда как второе выдающийся является содержательным компонентом семантики определяемого слова; или Выхлоп — выход отработанных <sup>44</sup> газов из цилиндра двигателя внутреннего сгорания после рабочего хода поршня, где причастием передан момент участия газов в рабочем цикле двигателя: они «работали» и «отработали».*

Причастия-релятемы, как отмечалось выше, целиком вошли в список нулевых множителей по определению, поэтому частота их может быть оценена лишь интуитивно, и эта интуитивная оценка, сделанная нами по опыту кодирования 1500 дескрипторов, подтверждает их относительность в частотном списке к зоне нерелевантности семантических множителей. Наш вывод можно подкрепить также и результатами других исследований. Так, для выяснения роли причастий в текстах описания терминов электроники был произведен подсчет их «валентности», под которой в данном случае понимается число различных, не совпадающих друг с другом членов, с которыми сочетается причастие (как слева, так и справа от него) на всем массиве анализируемых текстов. Например, в приведенном ниже фрагменте списка для причастия *характеризующий, -емый* оно имеет валентность 5, поскольку встречается в сочетаниях с пятью разными сло-

---

<sup>44</sup> Форма страдательного причастия продиктована здесь наивным человеческим восприятием, с точки зрения которого мы не делаем ошибки, когда говорим *Солнце опускается (садится)*, и поэтому она могла бы и не считаться здесь за ошибку, если бы вся дефиниция не была построена с претензией на терминологическую точность. В этих условиях необходимо было употребить форму *отработавших*.

вами — разряд, наличие, испускание, падение, интенсивность:

разряд, характеризуемый наличием. . . ,  
разряд, характеризуемый испусканием. . . ,  
разряд, характеризуемый падением. . . ,  
разряд, характеризующийся интенсивностью. . . , и т. д.

Таким путем был получен список из пятидесяти с лишним причастий (заметим в скобках, что их число коррелирует с количеством причастий — носителей отношений (62), выделенных нами на большем по объему материале дефиниций толкового филологического словаря), обладающих валентностью  $>4$ , т. е. встречающихся не менее двух раз. Эти и только эти причастия признаны релятами в языке RX-кодов, остальные же причастия одноразового использования передают некоторый содержательный признак<sup>45</sup>. В качестве комментария к данному результату хочется подчеркнуть, что фактически речь здесь идет не о валентности, как считает автор названной работы, а о частоте встречаемости причастий, взятой, правда не в ее абсолютном выражении, а с некоторой поправкой, предусматривающей исключение повторяющихся в сочетаниях слов. Таким образом, мы имеем дополнительный случай убедиться в правильности нашего решения об использовании статистического критерия релевантности семантических множителей при построении асимптотической модели русского тезауруса (§ 7).

По сравнению с существительными — носителями отношений причастия отличаются прикрепленностью к иным типам фасетов, т. е. отражают в целом и иные отношения. Если характер отношений, передаваемых нулевыми множителями-существительными, связан по преимуществу с организацией фасетов I (универсальные) и II (предметы, идеи, явления, процессы) уровней в фасетно-блочном тезаурусе, то причастия относят характеризуемые ими ситуации почти исключительно к фасетам более низких, более конкретных рангов — творительному и условий и обстоятельств. Здесь мы встречаемся с отношениями «инструмента» (производимый), «материала и изделия» (изготавляемый, получаемый), «иметь причиной» (покоящийся на, вызываемый) и «быть причиной» (определяющий,зывающий), «иметь следствием» (приносящий, полученный) и «быть следствием» (создающий, отражающий, вытекающий). Среди отношений, образующих фасет характеристик явлений и процессов, — «иметь условием» (тре-

бующий), «иметь обстоятельством» (характеризующий, сопровождающийся), «выражаться в» (свидетельствующий, проявляющийся), «содержать в себе, заключать» (содержащий, заключающий, выражющий, замещенный), «быть объектом» (подверженный), отношения «направления» (идущий, направленный), «местонахождения» (простирающийся, находящийся, лежащий), функциональное (идущий на, устроенный, употребляющийся, служащий, рассчитанный, предназначенный, посвященный, оборудованный, занятый, ведущий) и другие. Единственная, как представляется, хотя и довольно многочисленная группа, причастий способна, наряду с существительными в этой роли, передавать отношения более высокого ранга — экзистенции, наличия, факта (существующий, совершенный, совершающийся, сложившийся, происходящий, принятый, пользующийся, покрытый, обладающий, охваченный, исполненный, проникнутый, отмечаемый, осуществляющий, основанный на, находящийся, имеющий, имеющийся, получивший, возникший, оказываемый, оказывающий, устанавливаемый).

Осознание семантической «нулевости» причастий отразилось в попытке сокращенного написания в дефинициях, правда, только одного из них — *употр.* вместо *употребляющийся, употребляемый*, призванной подчеркнуть как бы внешний для семантики дефиниции характер передаваемого отношения, необязательность его полного обозначения. Ср.: *Дрова — пиленные и расколотые* (№ здесь содержательный признак) *деревья, употр. как топливо; Просвир — белый круглый хлебец, употр. в обрядах православного богослужения; Ковер — изделие из тяжелой ворсистой узорчатой ткани, употр. для покрытия пола, украшения стен.* Не менее распространенные в словарных определениях и столь же стандартные по своей функции *предназначенный, изготовленный, оборудованный, связанный с, свидетельствующий, обладающий, служащий* и под. такой операции не подверглись. Аналогичная непоследовательность наблюдается и в передаче квантивативных наречий: *часто, редко, обычно, всегда* вообще входят в дефиницию целиком, без сокращений, а *преимущественно* — в сокращенной записи. Ср.: *Забор — ограда, преимущ. деревянная; Прототип — первоначальный образец, прообраз, преимущ. действительное лицо как источник для*

---

<sup>45</sup> Пшеничная Л. 9. Указ. соч., с. 44—46.

*создания литературного образа, героя.* То же самое следует сказать об избыточных множителях *некоторый, какой-нибудь и определенный, основной, особый, различные, специальный*: первые два выступают всегда в сокращенном виде, прочие же — нет. Ср: *Дукат — старинная монета в нек-рых западноевропейских странах; Дуля — название нек-рых сортов небольших груш; Ряд — совокупность каких-н. явлений, следующих или расположенных в определенной последовательности; Полотно — гладкая льняная ткань особого переплетения; Полка — приделанная к стене, вделанная в стену или шкаф горизонтальная доска для различных предметов; Сцена — специальная площадка, на к-рой происходит театральное представление, а также вообще театр, театральная деятельность;* и т. п.

В попытках составителей прибегнуть к условному, почти символическому обозначению хотя и важных, по стандартных и предсказуемых сведений об отношениях между предметами реального мира, из слов-имен которых составляется дефиниция, можно видеть стремление вынести информацию об этих отношениях, так сказать, за скобки семантического наполнения дефиниции, приправить обозначения их по статусу к грамматическим, стилистическим и другим пометам, которые тоже не входят непосредственно в семантику определяемого слова (ср.: *прил., перен., разг.* и т. п.). Эта тенденция, не проведенная в филологических словарях последовательно ни для каких разрядов нулевых множителей, т. е. слов релятивного содержания, коснулась тем не менее каждого из них, в том числе и релятив-существительных: *Пикша — северная морская промысловая рыба из сем. тресковых; Помидор — огородное растение сем. пасленовых...* Думается, коль скоро нерелевантность логико-семантических отношений в смысловой структуре дефиниции можно считать доказанной, способы их обозначения в принципе тоже следовало бы стандартизовать и формализовать, хотя бы по образцу синтаксических отношений, передаваемых, как, например, управление, неопределенными местоимениями *кому-н., чем-н., кого-чего-н.* и т. п., и распространить соответственно на все их типы, встречающиеся в филологическом словаре.

Итак, подводя итог рассмотрению реляционных слагаемых смысла предложения-дефиниции, мы констатируем наличие в ее составе двух больших групп отношений: 1) локальные отношения субъективной оценки, выступаю-

щие как средство передачи специфической информации, выражаемые главным образом прилагательными (*узкий*—*широкий*, *большой*—*маленький*, *сильный*—*слабый* и т. п.) и играющие своего рода эмфатическую, гиперхарактеристическую роль в составе словарного определения; 2) системные, логико-семантические отношения, отражающие структуру базовых связей в семантическом пространстве лексики и языка в целом и воплощенные в организации тезауруса. Что касается способов их передачи, то они могут выражаться и лексически, и грамматически, и синтаксически. С другой стороны, сами синтаксические средства обладают возможностью фиксировать как отношения двух первых родов, так и собственно синтаксические отношения. Если отношения первой группы играют оценочно-прагматическую функцию в дефиниции, как бы приспособливая ее к индивидуальному восприятию путем апеллирования к само собой разумеющемуся и общезвестному (ведь всем понятно без лишних разъяснений, что такое *широкий* или *толстый*, например), то отношения второй группы выполняют классификационную задачу, поскольку с их помощью устанавливается место определяемого слова в системе. При этом классификационная привязка исходного слова только имплицитно является тезаурусно-ориентированной, тогда как внешним ее выражением служит указание на любой один или несколько «системных промеров» лексической семантики (т. е. внешне — неполная тезаурусная ориентированность), будь то синонимия, антонимия, семантическое поле, тематическая группа или какое-то иное измерение семантического пространства. Что же остается на долю синтаксиса в том суммарном смысле, который передается словарным определением?

## § 11. СИНТАКСИЧЕСКИЙ КОМПОНЕНТ В СЕМАНТИКЕ ДЕФИНИЦИИ

Они заняты были не столько тем, чтобы установить какую-то достоверность, сколько стремлением показать, как далеко они зашли в поисках истины: *quam docti fingunt, magis quam norunt.*

*M. Монтень. Опыты*

Функциональный синтаксис — глубинные структуры — семантическое представление

Обращаясь к собственно синтаксическому аспекту смысла дефиниции, мы словно утрачиваем общую перспективу, которая позволяла нам ранее довольно четко соот-

носить каждое из рассматриваемых слагаемых (денотативное, реляционное) с семантикой всего предложения. В самом деле, если попытаться сопоставить с последней, например, структурную схему предложения, то мы не найдем здесь точек соприкосновения: «Более очевидно, что многообразие явлений действительности отражается в семантике лексической. Принято думать, что фильтр синтаксической структуры ломает, деформирует эти отражения. Если разграничивать в синтаксисе уровень абстрактной, голой схемы предложения и уровень ее конкретного лексического наполнения, то лексическая семантика, отражающая действительность, уходит сквозь ячейки синтаксической сети, как мелкая рыбешка, а сама сеть не помогает нам удержать семантическое содержание синтаксиса»<sup>46</sup>. Если же обратиться к более частным средствам выражения синтаксических значений внутри предложения, то и в этом случае не удается обнаружить сопоставимые с суммарным смыслом синтаксические слагаемые: предлогами выражаются отношения (т. е. опять же реляционный момент) зависимости и следования, флексии передают отношения связи и подчинения, аналогична роль порядка слов и т. д. Синтаксические средства выражения значения как будто размываются. Создается впечатление, что при достаточно полном учете денотативного и реляционного слагаемых мы практически исчерпаем смысл дефинитивного предложения, т. е. что синтаксис в нем несет минимум семантической информации. В общем-то это впечатление закономерно, поскольку свидетельствует о том, что однозначным образом провести разграничительную линию между семантикой и синтаксисом не удается. Два слова являются синтаксически зависимыми, если форма одного из них зависит от грамматической природы другого; два слова являются семантически зависимыми, если значение одного из них дополняет, поясняет, развивает значение второго. Но коль скоро грамматическая природа слова неразрывно связана с его семантикой, то можно говорить о существовании в предложении непрерывного спектра перехода от семантики к синтаксису и от синтаксиса к семантике. Этот вывод, понятно, не решает нашу задачу, но заставляет ставить вопрос о поиске таких способов представления синтаксических значений, которые делали бы их соизмеримыми со способами описания лексической семантики, т. е. вопрос о поисках общего знаменателя в представлении

семантики и синтаксиса. Таким образом, мы попадаем в русло исследований так называемой синтаксической семантики<sup>47</sup>, которая оперирует понятиями грамматической структуры предложения, семантической структуры предложения и актуализованного высказывания.

Воспользоваться результатами, полученными в рамках синтаксической семантики, для наших целей не удается, главным образом потому, что данное направление не выработало пока отчетливого операционного аппарата. Однако сопоставление развивающегося им трехуровневого подхода с категориями и приемами анализа структуры и семантики предложения в иных синтаксических теориях позволяет сделать заключение, что такой подход не является исключительной прерогативой синтаксической семантики: в том или ином виде эти три плана, три уровня предложения учитываются многими синтаксическими теориями, причем ряд из них предлагает и свои способы универсализации грамматической (синтаксической) и семантической структур предложения. Рассмотрим некоторые из этих способов.

Прежде всего, содержательную трактовку всех трех уровней мы находим при традиционном изучении семантики предложения в функциональном синтаксисе, анализирующем конструкцию по частям речи, членам предложения, грамматическим категориям (рода, числа, падежа, времени, наклонения) и функциональным категориям (субъекта, предиката, объекта). Коммуникативный аспект предложения при функциональном анализе приобретает довольно сильный логико-грамматический акцент, но, будучи нерелевантным для предложений-дефиниций, нас он интересовать не будет. В итоге такого рассмотрения каждой лексеме, входящей в состав предложения, приписывается некоторая самым общим образом сформулированная значимость. Последовательность этих зависимостей определяет семантическую структуру предложения, а добавление их к семантике соответствующих лексических единиц конкретизирует и уточняет последнюю. Например, словарное определение *Перекат — мелководный участок в русле реки* выглядит в таком анализе следующим

<sup>46</sup> Золотова Г. А. К дискуссии о соотношении синтаксиса и семантики. — В кн.: Сборник научных трудов МГПИЯ им. М. Тореза, 1977, вып. 112, с. 62.

<sup>47</sup> См., напр.: Москальская О. И. О предмете синтаксической семантики. — Там же, с. 12.

образом: первый член этого предложения подлежащее *перекат* выражено именем существительным — частью речи, обозначающей предметы и явления реального или мыслимого мира, мужского рода, в форме именительного падежа, выступающего в данном случае в чисто назывной, номинативной функции; второй член предложения — сказуемое *мелководный участок в русле реки* выражено именной группой с пулевой связкой, группой, составленной из существительного *участок* (составленно предикат, семантически эквивалентный субъекту-подлежащему *перекат*) с двумя определениями, одно из которых передано прилагательным, а другое — предложно-падежным сочетанием двух существительных... и т. д. Добавления к лексической семантике составляющих данное предложение единиц, которые мы можем сделать в результате этого анализа, таковы: во-первых, это идея «места», заключенная в семантике локатива *в русле*, но эта идея не приносит новой информации, поскольку она уже содержится в самой лексеме *участок*, относящей определяемое слово *перекат* к «множеству мест» (ср. выше, предыдущий параграф); во-вторых, это могла бы быть идея «принадлежности к реке», передаваемая родительным падежом, но эта информация представляется избыточной, поскольку само соположение лексем *руслу* и *река* уже говорит достаточно об их тесной семантической связи; наконец, это указание на то, что свойство *мелководный* относится именно к существительному *участок*, а не к какому-то другому, что представляется важным для правильного восприятия семантики всего высказывания и что определяется устанавливаемыми в таком анализе отношениями между членами предложения. Синтаксическая форма (а именно, флексия *-ый* и позиция слова) однозначно привязывает прилагательное к указанному существительному и делает невозможными независимые сочетания *\*мелководный в русле* или *\*мелководный реки*. Если опустить флексию и предположить, что прилагательное в данной definicции может занимать любую позицию, т. е. что порядок слов нерелевантен для смысла всего выражения, то мы получим такие варианты: 1) *Перекат — участок мелководного русла реки*, 2) *Перекат — участок русла мелководной реки*. Из них первый представляется еще допустимым с точки зрения близости к истинному смыслу определяемой лексемы, тогда как второй уже искачет истину. Рассмотрение возможностей функционального синтаксиса в представле-

ния семантики высказывания можно было бы углубить, Но, думается, это уже не даст новых наблюдений, поскольку подобный анализ не выявляет, за исключением порядка слов, таких добавок к семантике лексических единиц, которые не были бы учтены в наборе отношений, передающих реляционное значение (см. предыдущий раздел). Это легко показать, использовав представление структуры дефинитивного предложения в виде графа, в вершине которого помещено определяемое слово, в связанных с ней узлах — «предметы», через которые дается определение, а дуги соответствуют отношениям<sup>48</sup>.

Тогда для нашего примера граф ом. на рис. 5, где 1 — *перекат*, 2 — участок, 3 — мелководный, 4 — русло реки, а — быть элементом множества, б — иметь атрибутом, в — быть составной частью.

Подобный график напоминает ячейку «лексического конуса» М. Молинер<sup>49</sup>

или фрагмент «семантической сети», с помощью которой может быть представлена как семантическая структура отдельного текста, так и схема базовых отношений в тезаурусе. При этом все реляционные значения, которые мы могли бы установить в данной фразе методами функционального синтаксиса, отражены в построенном графике называниями соответствующих дуг. Естественно, один пример еще не дает оснований для заключения обо всех дефинициях, но отсутствие в лингвистической литературе исследований, в которых была бы сделана попытка соотнести семантику индивидуальных лексем со структурой синтаксической конструкции именно с помощью указанных методов, укрепляет нашу позицию. Что же касается порядка слов, несущего в дефиниции релятивную семантическую информацию, его следует учитывать при формализованной записи смысла, приписывая определенные веса единицам, составляющим предложение-дефиницию. В нашем случае эти веса приписываются на основе статистического селективного параметра (см. выше, § 7).

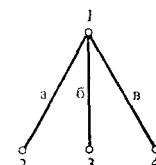


Рис. 5

<sup>48</sup> Ср.: Пшеничная Л. д. Указ. соч., с. 9—15.

<sup>49</sup> См., напр.: Караполов Ю. Н. Общая и русская идеография, с. 252—253.

Обратимся к другим способам изучения семантики синтаксиса. В так называемой грамматике зависимостей структурная схема предложения организуется вокруг ядра-глагола и предстает либо в виде своеобразного конуса, либо в скобочной записи, так что число исходящих ветвей (или односторонних скобок) характеризует валентность соответствующей лексемы. Валентностная структура и составляет инструмент и главную идею синтаксиса зависимостей. Однако воспользоваться его возможностями для нашей задачи оказывается нереальным, поскольку специфика анализируемых нами дефинитивных предложений как раз и заключается в эксплицитной непредставленности глагола-сказуемого, или предиката-действия, набор валентностей которого образует семантический каркас предложения. Не может ничего добавить к нашему рассмотрению и классический структуральный синтаксис, опирающийся на анализ по непосредственным составляющим, применение которого к нашему случаю позволяет лишь выявить попарные связи конституентов предложения в соответствующих бинарных сегментах, что не дает ничего нового по сравнению с приемами функционального синтаксиса. Правда, структурирование дефиниций по непосредственным составляющим подводит к установлению субSTITУционных классов, которые задают допустимые границы перефразирования для того или иного сегмента. Но чтобы определить реальное наполнение каждого такого класса лексическими единицами, мы должны уже до анализа обладать тезаурусной информацией, которую еще только надеемся извлечь из словарных дефиниций.

Некоторую перспективу в разрешении этих трудностей открывают, как будто, трансформационный синтаксис с его попытками найти точки соприкосновения лексической и синтаксической семантики на уровне глубинных структур, а также порождающая семантика, идущая в этом отношении дальше и стремящаяся эксплицитно выявить синтаксические и лексические средства, которые на поверхностном уровне отражали бы различия в глубинных структурах. При отсутствии единообразной трактовки последних сама идея использования глубинных структур как адекватного уровня сопоставления, соизмерения лексической и синтаксической семантики разделяется многими исследователями. Ее высказывает, в частности, Л. С. Бархударов, понимая под глубинной структурой предложения его референциальное синтаксическое

значение, т. е. обозначение системы отношений между референтами лексических компонентов предложения<sup>50</sup> (ср. нашу трактовку этих отношений в предыдущем параграфе). Иное представление о глубинной структуре развивает, например, Р. Ружичка, который видит неизбежность обращения к ней для анализа трех аспектов соотношения лексической и синтаксической семантики: «1. Различие в восприимчивости синтаксических структур к семантическим классам лексем, что ведет к сопротивлению синтаксических структур в зависимости от определенных лексем. 2. Необходимость постулирования более абстрактных синтаксических структур с целью опосредствовать явные лексические различия и скрытые синтаксические. 3. Значимость расширенного понятия семантики, которое включает и системы познаний и предположений при исследовании лексической и синтаксической семантики»<sup>51</sup>. При этом автор не считает, что уровень глубинных структур достаточен для прямого соотнесения семантики лексики и синтаксиса: глубинно-синтаксические репрезентации специфичны для каждого языка, а общим, единым является семантическое представление, из которого развертываются различные глубинно-синтаксические структуры. Так, одному и тому же семантическому представлению соответствуют в русском, согласно автору, три следующие глубинные структуры:

- *резкий . . . тон был непозволительным  
грубое обращение с иностранцем было недопустимым*
- *резкий . . . тон не мог быть позволен  
грубое обращение не могло быть допущено*
- *нельзя было { позволить } Г резкого тона  
{ допустить } \ грубого обращения*<sup>52</sup>.

Аналогичное понимание и использование глубинных структур мы находим у Е. В. Падучевой: «Глубинная структура — это своего рода промежуточный уровень между синтаксическим и семантическим представлением

<sup>50</sup> Бархударов Л. С. Глубинная структура предложения в семиотическом аспекте. — В кн.: Сборник научных трудов МГПИИ им. М. Тореза, 1977, вып. 112.

<sup>51</sup> Růžička JR. О трех аспектах взаимопроникновения синтаксиса и семантики. — В кн.: Linguistische Arbeitsberichte, N. 16. Leipzig, 1977, S. 30.

<sup>52</sup> Růžička R. Синтаксис и значение предложения. — Zeitschrift für Slawistik, Bd. XX, N 5—6. Berlin, 1975, S. 750—751.

предложения, а именно уровень, на котором, произведены отождествления смыслов синтаксических конструкций, а те отождествления смыслов предложений, которые требуют обращения к смыслам индивидуальных лексем, будут сделаны на следующем этапе анализа»<sup>53</sup>. Соответственно синонимичные либо синтаксически (*большое ветвистое дерево* ↔ *большое и ветвистое дерево*), либо лексически (*древа загорелись* ↔ *древа начали гореть*) предложения вовсе не обязательно должны иметь единую глубинную структуру<sup>54</sup>.

Таким образом, мы приходим к необходимости обращения от глубинных структур к семантическому представлению смысла предложения для адекватного изучения соотношения его лексических и синтаксических составляющих. В рамках интерпретативной семантики такое представление опирается на компонентное описание смысла лексем, составляющих синтаксическую конструкцию (работы Катца и Фодора)<sup>55</sup>, либо оперирует понятием «глубинного падежа» (Филлмор). Генеративная семантика использует с той же целью предикатно-аргументную нотацию символической логики, единицами которой являются атомарные предикаты (V) типа «каузировать», «испытать», «отрицание», «привести к состоянию», в том числе выраженные конкретными лексемами данного языка — «радостный», «мертвый», «усталый» и т. п., и референциальные переменные (NP), соответствующие единицам номинации. В этом случае селективные признаки инкорпорированы в единицы словаря, т. е. в предикаты и аргументы, и выступают как сочетательные ограничения, тем самым семантика и синтаксис не разделяются (Лакофф, Макколи).

Попытаемся сопоставить глубинное семантическое представление предложения-дефиниции, сделанное по правилам генеративной семантики, с семантическим кодом, использованным нами при построении тезауруса для записи смысла словарных определений с помощью компонентов-квазиоснов (см. выше, § 6). Возьмем дефиницию из филологического словаря: *Опока* — рама (ящик без дна) с формовочной землей для заливки металлом. Для установления правильных отношений между составляющими ее элементами, т. е. для определения предикатов (в качестве консультационной базы) будем обращаться к энциклопедическому определению того же слова: *Опока* — в литеином производстве приспособление в виде жесткой рамы (открытого ящика), служащее для удержания в нем фор-

моловчной земли в процессе изготовления разовой земляной литейной формы, транспортировки ее и заливки металлом. О. изготавляются из стали, чугуна, алюминиевых сплавов и реже из дерева. В зависимости от конфигурации отливок О. делаются прямоугольные, круглые... Предикатно-аргументное дерево, соответствующее филологическому (т. е. первому из приведенных) определению, можно представить двумя способами:

1) вне связи с конкретным лексиконом, с обозначением объекта (Y) и возможного субъекта (X) действия:



Рис. 6

2) с ориентировкой на замещение предикатных мест в поверхностной структуре соответствующими именными группами (переменными, или аргументами) (см. рис. 7). Предикат «действовать таким образом, чтобы» предполагает наличие потенциального субъекта действия — X, который легко восстанавливается при обращении к более развернутой энциклопедической definicции. Из указания на область, к которой относится определяемое слово, а именно литейное производство, следует, что потенциальный субъект X — *литейщик*. Вообще использование информации из энциклопедического определения расширяет семантическое представление предложения — в данном случае

<sup>53</sup> Падучева Е. В. О семантике синтаксиса, с. 22.

<sup>54</sup> Там же, с. 23.

<sup>55</sup> Ср. также развитие этих идей применительно к задачам синтаксической семантики в работе: Селиверстова О. П. К вопросу о задачах и методах синтаксической семантики. — В кн.: Сборник научных трудов МГПИИ им. М. Тореза, 1977, вып. 112.

главным образом по линии увеличения набора предикатов: с учетом этой информации дерево предикатов должно включать дополнительные ветви: 1 — «придавать форму», с потенциальным субъектом действия  $X_1$  (*формовщик*); 2 — «сохранять (удерживать форму (и землю)» и 3 — «перемещать в пространстве» (*двигать, переносить, перевозить, транспортировать*). При этом в число аргументов

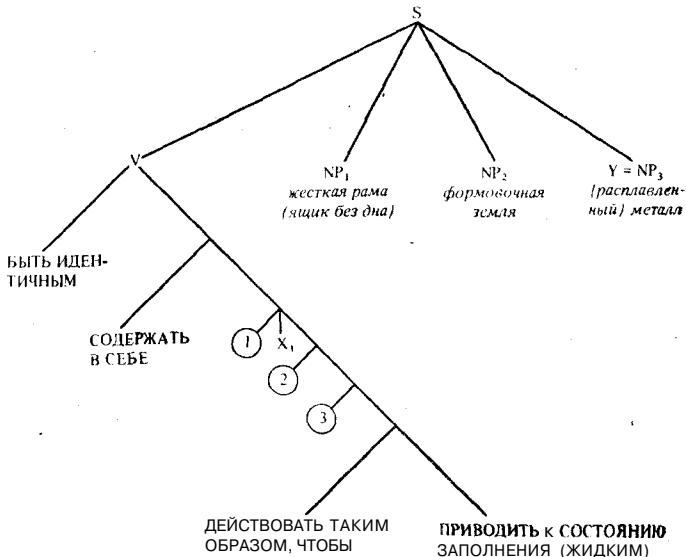


Рис. 7

добавится лишь один — между  $NP_2$  и  $NP_3$  — «разовая земляная литейная форма»<sup>56</sup>. Но что показательно в данном случае — изменение семантического представления предложения совершенно не затрагивает его глубинной синтаксической структуры, в обоих вариантах — филологическом и энциклопедическом — она остается одной и той же, а обогащение семантики происходит за счет смысла вводимых индивидуальных лексем. На поверхностном же уровне способы выражения предикатов, за исключением первого из них — «быть идентичным» и последнего — «приводить к состоянию заполнения», в двух сравниваемых предложениях оказываются различными: в филологическом определении предикат «содержать» пе-

редается синтаксическим средством — предлогом *с*, в энциклопедическом сочетание синтаксических слов *служащее для удержания в нем* заключает два предиката — «содержать» и «сохранять»; в филологическом предикат «действовать таким образом, чтобы» обозначен предлогом *для*, в энциклопедическом он передается вспомогательным сочетанием *в процессе*. Что касается способов выражения аргументов, то для тех из них, которые представлены в обоих определениях, различия могут быть связаны с расширением именных групп — не просто *рама*, а *жесткая рама*, и далее *из стали, из чугуна и т. п., прямоугольные, круглые* и т. п., не просто *форма*<sup>57</sup>, а *разовая земляная литейная форма* и т. п.

Если сопоставить теперь наш семантический код рассматриваемой дефиниции (опок рам ящ бездн формов зем зали метал) <sup>58</sup> с ее глубинным семантическим представле-

---

<sup>56</sup> В этих двух употреблениях лексемы *форма* — в предикате и в аргументе — отчетливо проступает метонимическая связь двух ее значений: «внешнее очертание» — в предикате и «приспособление для придания такого очертания» — в аргументе, что хорошо, па наш взгляд, согласуется с абстрактной природой первого и конкретной вещественностью второго.

<sup>57</sup> Отсутствие аргумента «форма» в филологическом определении, строго говоря, искажает семантическое представление, поскольку в цепочке выпадает одно звено: опока служит для помещения в ней земли, земля — для изготовления форм, а уже форма — для заливки ее металлом. Однако для филологического определения такое опущение представляется допустимым, и именно в силу его стремления к максимальной компактности: идея «формы» инкорпорирована здесь (в расчете на ассоциативное восприятие носителя языка) в прилагательное *формовочная* (земля). Впрочем, более адекватной была бы, вероятно, другая концовка в филологической дефиниции — для отливки металлических *форм*.

<sup>58</sup> Способы фиксации квазиоснов изложены выше, в соответствующем месте § 6, но приведенный пример, при всей его прозрачности по отношению к исходной дефиниции, требует некоторых пояснений. Последовательностью букв «бездн» кодируется именно сочетание *без дна*, которое не является таким уж редким, помимо данного случая, дважды встретившись в дефинициях дескрипторов и один раз на массиве словарника (см. наш «Частотный словарь семантических множителей. . .»). Слово же *бездна*, которое может, на первый взгляд, ассоциироваться с этим буквенным сочетанием, очевидно, в силу его стилистической отмеченности не встретилось в проанализированном массиве слов и, можно предполагать, вообще не употребляется в правой части толковых словарей. Однако, если его все же потребуется кодировать, оно получит другое буквенное обозначение — «бездна». Далее, квазиоснова «зали» развертывается в *залить, заливать, заливка*, тогда

гнием, то мы должны констатировать, что в нашей записи нашли отражение практически лишь аргументы (правая часть второй схемы), тогда как почти все предикаты, образующие каркас для переменных и создающие предпосылку для понимания всего предложения, опущены. Предложение расформировано, превратилось в конгломерат как будто случайных гиперлексем, его глубинная синтаксическая структура разрушена и адекватное его понимание без предикатов оказывается невозможным. Предложенная семантическая запись в квазисновах допускает только различные истолкования, более или менее приближающиеся к исходному смыслу, и неограниченное ассоциирование вокруг каждой отдельно взятой квазисновы, но не дает условий для однозначного восстановления исходного текста. Однако понимание было бы еще менее достижимо, если бы из семантического представления мы взяли только его левую часть — одни предикаты, или если бы ограничились первой схемой: «быть идентичным», «содержать в себе», «действовать таким образом, чтобы», «приводить к состоянию заполнения». В поверхностной структуре способы выражения этих предикатов могут быть различными, как показало сравнение филологического и энциклопедического определений, даже для одной и той же глубинной синтаксической структуры. Поэтому адекватное восстановление текста, а значит, и его понимание на основе одной только предикатной структуры тоже оказывается невозможным. Более того, последняя не допускает и свободного ассоциирования, поскольку в лучшем случае предполагает строгое перечисление набора языковых средств для выражения соответствующих предикатов. В этом отношении аргументная часть представляется более содержательной и конкретной: при свободном ассоциировании вокруг каждого аргумента выстраиваются определенные смысловые ряды, образующие «семантическую сферу» индивидуальной лексемы, которые при взаимном наложении и пересечении выявляют общую, единую для всех аргументов зону, куда входит и соответствующее семантическое представление. Эта зона значительно шире семантического представления данного предложения, но она полезна уже тем, что ограничивает область выбора.

Кодовая семантическая запись, не ставя задачей достижение понимания текста в обычном смысле слова, имеет некоторые преимущества перед аргументной частью глубинного семантического представления. Прежде всего это

касается большей легкости развертывания и ассоциирования квазиоснов, чем целых слов, во-вторых, связано с большей четкостью итоговой зоны пересечения смысловых рядов, четкостью, которая обусловлена в этом случае дополнительными ограничениями. Последние проистекают из того, что в семантическую запись согласно алгоритму кодирования включаются все знаменательные слова, в числе которых оказываются не только аргументы, т. е. переменные величины в предложении, но и часть предикатов, в тех случаях, когда они совпадают с реальными лексемами данного языка. Так, в рассматриваемой нами дефиниции, помимо квазиоснов всех аргументов анализируемого предложения, фигурирует и квазиоснова одного из предикатов, которому в поверхностной структуре соответствует слово *заливка*. Таким образом, используемая нами кодовая семантическая запись, обладая свойствами простоты и однозначности получения, позволяет в то же время приблизиться к уровню глубинного представления смысла предложения, на котором слагаемые синтаксической и лексической семантики вступают в эквивалентные отношения и представимы в сравниваемых друг с другом формах. Предпринятое с этой точки зрения рассмотрение некоторых синтаксических теорий позволяет сформулировать гипотезу, согласно которой при кодировании содержания дефиниции с помощью квазиоснов мы фиксируем, кроме аргументов, также и те предикаты глубинного семантического представления, которые выходят за рамки выражения чисто реляционных значений или логико-семантических отношений системно-классификационного типа (ср. предыдущий параграф), и непосредственно передают элементы Смысла определяемого слова. Иными словами, предполагается, что фиксация знаменательных слов в кодовой записи не только необходима, но и достаточна для адекватного представления как лексических, так и синтаксических слагаемых смысла. Тем самым решается вопрос о способе передачи релевантных для итогового смысла синтаксических значений в такой записи: они как бы лексикализуются и выступают в виде самостоя-

---

(окончание сноски)

как для слова *залив* (морской) используется обозначение «залив»; «формов» представляет группу *формовать*, *формовка*, *формовочный*, *формовщик*, а для слов *форма*, *формальный*, *формировать* используется квазиоснова «форм».

тельных семантических множителей. Следовательно, цель применяемой нами процедуры кодирования не в том, чтобы добиться понимания получаемой семантической записи, а в «аргументизации» части (логических, глубинно-семантических) предикатов, несущих, как мы пытались это показать, в том числе и синтаксическую информацию. Выскажем некоторые соображения, конкретизирующие, на наш взгляд, характер этой информации.

В ряде работ последнего времени, посвященных комплексному анализу (т. е. системно-семантическому, синтаксическому, стилистическому) отдельных групп лексики, или фрагментов лексико-семантической системы, среди прочих результатов описания выделяются попытки установить определенное соответствие между компонентным составом, или семенным наполнением лексемы, и реализуемой этой лексемой структурной схемой предложения. При этом оказывается, что, коль скоро такую соотнесенность удается зафиксировать, сами семы, обнаруживающие эквивалентность целой синтаксической конструкции, на глубинном уровне выступают в качестве предикатов. Сошлемся на две работы такого рода.

В кандидатской диссертации В. П. Абрамова <sup>59</sup> на материале большого числа разных подгрупп глаголов, содержащих сему «дать» (типа *вручить, предоставить, навязать, дарить, наградить, продать, платить, одолжить, завещать, оставить, послать, доставлять, сообщать, сигнализировать, вратить, доносить, объяснять*) последовательно проводится мысль о существовании устойчивой связи между этой семой и структурной схемой предложения  $N_1VN_4N_3$ , где  $N_1$  — субъект передачи,  $V$  — глагол передачи,  $N_4$  — объект передачи,  $N_3$  — ее адресат (*Управдом вручил ключи жильцу*). Причем у нас есть основания полагать, что отношение между семой «дать» и формулой конструкции  $N_1VN_4N_3$  является рефлексивным, т. е. по данной формуле мы с определенностью можем заключить и о наличии семы «дать» в конституирующем ее глаголе. Если теперь обратиться к осмыслинию статуса самой семы «дать», то мы склоняемся к тому, чтобы видеть в ней полноценный предикат — и не только в поверхностной структуре высказывания, но и на глубинном уровне <sup>60</sup>. Одновременно данные нашего «Частотного словаря семантических множителей» относят этот предикат к зоне нерелевантности (частота на массиве слов 42), т. е. квалифицируют его как нулевой семантический множитель,

С другой стороны, рассматривая свойства «фактообразующей номинализации», Н. Д. Арутюнова устанавливает следующие особенности лексемы *факт*: «Номинализации повествовательного предложения . . . соотносительны с существительным *факт*. Это имя в подавляющем большинстве случаев может их вводить, замещать, а также использоваться в качестве их предиката. . . Имя *факт* представляется некоторым обобщением значения номинализаций, специфику которого иногда стремятся выяснить через употребление этого существительного. Между тем семантика имени *факт*, как и многих других абстрактных имен, расшатана не только небрежным с собой обращением говорящих и пишущих, но и вследствие своего промежуточного положения по отношению к категориям мышления и коммуникации, с одной стороны, и явлениям действительности — с другой. Значение этого имени не является ни единой, ни стабильной величиной. . . Постепенно отрываясь от плана знаний и констатаций, *факт* приближается по значению к таким именам, как *событие*, *происшествие*, *случай*, *обстоятельство*, *явление действительности*. . . однако . . . с ними не идентифицируется . . . Для наших целей важно одно значение имени *факт* — то, которое реализуется в конструкции *тот факт, что*. . . соотносящей это имя с пропозицией»<sup>61</sup>.

Эти недвусмысленно высказываемые идеи о существовании таких индивидуальных лексем, которые оказываются эквивалентными целой пропозиции, укрепляют нас в мысли пересмотреть статус нулевых семантических множителей именно под этим углом зрения. Характерно, что перечисленные в цитате из статьи Н. Д. Арутюновой имена, с которыми сближается *факт*, согласно данным нашего частотного словаря тоже квалифицируются как нулевые семантические множители. Ср. (первая цифра

---

<sup>59</sup> Абрамов В. П. Структурно-семантическое описание глаголов передачи в русском языке. Канд. дис. М., 1980.

<sup>60</sup> Запись ее на этом уровне в виде «каузировать кого-нибудь начать иметь что-либо» вовсе не является канонической и содержит известную долю условности. Точно так же мы можем условиться оперировать этим предикатом на уровне глубинной структуры в форме «дать кому-то что-то».

<sup>61</sup> Арутюнова Н. Д. Сокровенная связка. (К проблеме предикатного отношения). — Изв. АН СССР. Сер. лит. и яз., 1980, т. 39, № 4, с. 348—350.

указывает частотность на массиве слов, вторая — на массиве дескрипторов):

событие 36, 15	происшествие 7, 2
случай 31, 10	(очевидно его частотность в данном случае
обстоятельство 30, 9	следует суммировать с частотностью гла- гола)
явление 45, 14	
	происходит 39, 15

Хотя следует отметить, что по частотности самого слова *факт* оно попадает в зону релевантности семантических множителей, поскольку ее величина не превышает критического порога ( $2 < 6$ ), что можно объяснить, вероятно, спецификой дефинитивных текстов филологического словаря, в меньшей степени использующих эту лексему, по сравнению с текстами ярко выраженной коммуникативной направленности. Задача могла бы состоять в том, чтобы установить определенный набор пропозиций или структурных схем предложения, соотносимых с каждой индивидуальной лексемой, находящейся в ранге нулевого множителя, как и наоборот, — вывести набор правил перехода от ядерного предложения-пропозиции к соответствующей представляющей ее лексеме (или лексемам). Такой результат, без сомнения, даст в руки исследователя эффективный инструмент для изучения соотношения семантики и синтаксиса.

### Семантическая запись по квазиосновам и фрейм-подход

Для полноты картины остановимся еще на двух способах разделения и приведения к общему знаменателю синтаксической и лексической семантики, способах, которые не вмещаются в границы собственно синтаксических теорий, но которые по своим результатам подтверждают, как нам кажется, правомерность самого выдвижения указанной гипотезы.

Речь пойдет прежде всего о так называемой функциональной стратиграфии, как она представлена, например, в статье Р. Лангакера<sup>62</sup>. Автором предлагается такой способ анализа предложения, при котором его семантическое представление имеет вид иерархии функциональных страт. Выделяются следующие четыре страты: перформативная, модальность—время—вид, экзистенциальная, объективное содержание. Так, предложение *Он, вероятно, служащий* «стратиграфируется» на рис. 8.

Обращает на себя внимание тот факт, что выразителями «объективного содержания» при таком подходе являются именно знаменательные лексемы (анафорический показатель в нашем примере потребует, естественно, содержательной замены), точно так же как при семантическом кодировании по квазиосновам.

Приоритет знаменательных лексем (т. е. аргументов — перед предикатами) для семантического представления предложения становится очевидным при попытке анализировать подобным стратиграфическим образом семантику

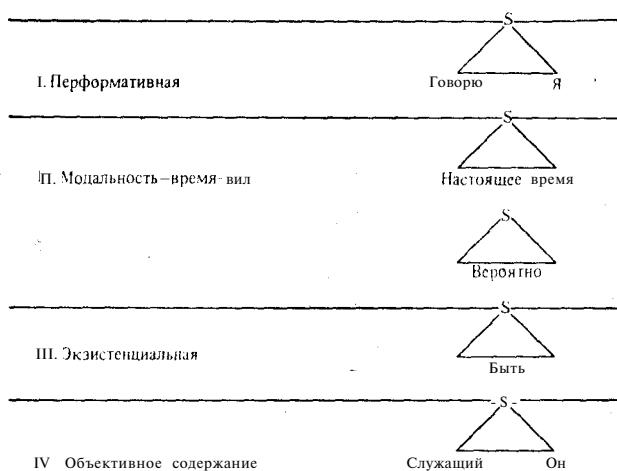


Рис. 8

словарных определений. Рассмотрим дефиницию *Служащий — лицо, работающее по найму в различных учреждениях, в сфере обслуживания (но не на производстве)*. Специфика дефинитивных предложений (см. выше, § 8) позволяет исключить из рассмотрения две первые страты, поскольку их содержание для анализируемых конструкций всегда одно и тоже: перформативная функция, практически не выраженная для каждого отдельного определения, является предпосылкой организации всего толкового словаря, т. е. всех определений, и состоит в приписывании максимальной вероятности тому, что данное высказывание истинно; грамматическое время в дефиниции

<sup>62</sup> Langacker R. W. Functional stratigraphy. — In: Papers from the parasession on Functionalism. Chicago, 1975, p. 352.

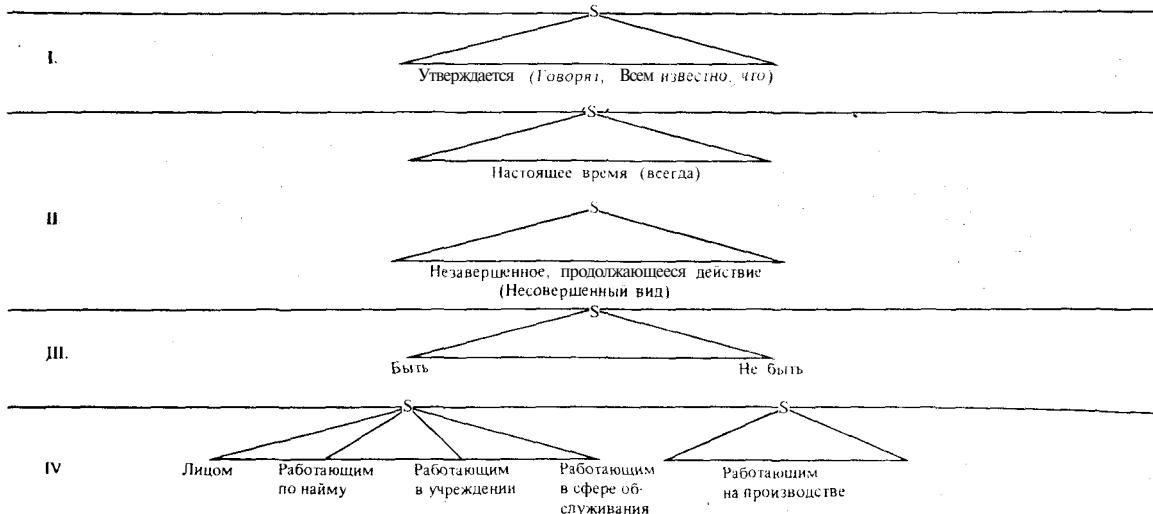


Рис. 9

всегда настоящее, вид — всегда несовершенный, а модальность исключена. Что касается третьей страты, то ее значение тоже может быть зафиксировано как постоянное, имея в виду, что отрицательные определения с равным успехом передаются как через отрицание предиката, так и с помощью отрицательных аргументов. Именно последний способ и избран нами при семантическом кодировании по квазиосновам, а так как число определений через отрицание в словаре довольно значительно, количество отрицательных семантических множителей (отрицательных аргументов) оказалось тоже велико и составило более 12% от общего их числа (см. «Частотный словарь семантических множителей...» — квазиосновы на *без-* и *не-*). Таким образом, сравнение семантической записи данной дефиниции, сделанной с помощью квазиоснов (служ лиц раб наем разл учрежд сфер обслу не-произв) <sup>63</sup>, с наполнением уровня «объективное содержание» при стратиграфическом анализе того же предложения показывает их полное совпадение, что и должно свидетельствовать в пользу выдвинутой гипотезы, поскольку синтаксическими слагаемыми смысла, представленными на первых трех стратах, без ущерба для «объективного содержания» можно в данном случае пренебречь.

Другой способ развивается в русле исследований искусственного интеллекта и связан с построением интеллектуальных банков данных. Он получил название фрейм-подхода, или теории (концепции) фреймов <sup>64</sup>. Фреймы являются фрагментами, блоками семантической сети, с помощью которой могут быть представлены знания о какой-либо предметной области (проблемной среде). По структуре фрейм представляет собой сложный граф, отображающий определенные события, характеристики понятий (предметов и явлений) и логические связи (функции)

---

<sup>63</sup> Для нашей записи затем вступают в силу правила исключения множителей — нулевых по определению (лиц, разл, сфер) и нулевых по статистическому селективному критерию (раб).

<sup>64</sup> Минский М. Фреймы для представления знаний. М., 1979. См. также: Оныккий Б. Н., Кокорева Л. В. Фрейм-подход к представлению знаний в системе управления НИР. — В кн.: Тезисы докладов 2-го семинара «Интеллектуальные банки данных» (Цахкадзор Арм. ССР, 10—18 марта 1980 г.). Ереван, 1980; Гаспарян А. С., Иванченко В. В. Фрейм-представление базовых трансформаций программного комплекса. — Там же, с. 26—27; Ускова Т. В. Фреймовый подход к машинному представлению словарных энциклопедических статей, — Там же, с. 94.

-в моделируемой проблемной среде. «Фреймы являются элементами сети и содержательно представляют большие сложные структуры данных с привязанными процедурами, которые выражают общее знание о стереотипных ситуациях»<sup>65</sup>. При конструировании фреймов оперируют предикатами и аргументами, которые выражают общие понятия, константы (имена собственные) и типовые переменные, и отношения между

которыми осложняются кванторами, указанием иерархии, временной последовательности и некоторыми другими конструктивными элементами.

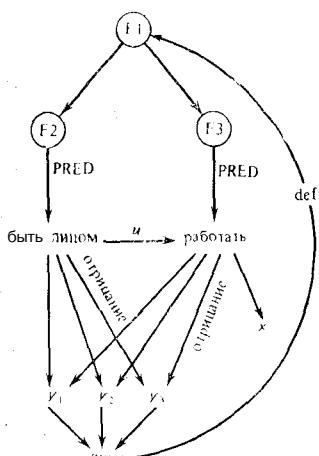


Рис. 10

иерархии и включения в множество между лицом (вообще) и лицом-1, работающим в учреждении, лицом-2, работающим в сфере обслуживания и т. д. Но даже и использованное, огрубленное, или агрегированное, представление структуры позволяет показать, что содержательная ее интерпретация в сильной степени зависит от аргументов, а все эти аргументы находят отражение и в нашей семантической записи этой структуры с помощью квазиоснов.

Не имея возможности развивать здесь далее наши рассуждения в этом направлении, отметим, что обращение к концепции фреймов продиктовано стремлением подчеркнуть плодотворные и интересные, на наш взгляд, перспективы, которые открывает указанный подход в изучении соотношения лексической и синтаксической семантики<sup>66</sup>. Возможности этого подхода требуют серьезной

Если прибегнуть к фреймовой структуре для изображения только что рассмотренной дефиниции, то мы отчетливо увидим, что она складывается из двух фреймов — «быть лицом» и «работать» и представлена на рис. 10 в виде схемы, где аргументы означают:  $x$  — условие (по найму),  $y_1, y_2, y_3$  — место работы.

Для более точного представления этого фрейма необходимо ввести кванторы «различные» и «сфера» (обслуживания), а также установить отношения

проработки, особенно на материале словарных определений. Лингвистической предпосылкой приложения фреймовой методики к анализу дефиниций служит идея о непосредственной соотносимости высказывания с ситуацией<sup>67</sup>, а с другой стороны, сама организация фрейма, его составные части обнаруживают полную аналогию с теми элементами, из которых складывается любая дефиниция. «Фрейм в FRL — это совокупность списков, содержащих до пяти уровней вложенности подструктур типа слот, фасет, данное, комментарий и сообщение. Слоты определяют структуру фрейма, т. е. свойства, которыми обладает описываемый объект. Фасеты задают аспекты рассмотрения этих свойств. . . Подструктура данное определяет конкретное значение слота в соответствии с заданным аспектом, а подструктуры сообщение и комментарий используются для дополнительного описания объектов»<sup>68</sup>. Соотнося указанные подструктуры с устройством дефиниции в словаре, мы констатируем следующие аналогии: все словарное определение можно приравнять к сообщению (в других терминах это описание или единица — ср. ниже), тогда данное представляет собой входное слово, комментарий — правую часть толкового словаря, слотом является каждый релевантный (т. е. не нулевой и не избыточный) семантический множитель, а фасет зафиксирован как раз нулевыми и избыточными множителями (ср. предыдущий параграф, где показано, что все нулевые и избыточные несут информацию об отношениях разного ранга). Принципиальный изоморфизм дефиниции с фрейм-структурой подтверждается и при обращении к другим языкам представления знаний: «Формализм описания знаний

---

<sup>65</sup> Вольфенгаген В., Воскресенская О. В., Горбанев Ю. Г. Система представления знаний с использованием семантических сетей. — В кн.: Вопросы кибернетики, вып. 55. М., 1979, с. 53.

<sup>66</sup> Ср. идеи Ф. Р. Адрадоэа о том, что наиболее эффективным изучение соотношения семантики и синтаксиса высказывания должно быть на уровне ментальных процессов (*Adrados F. R. Estudios de Semántica y Sintaxis*. Barcelona, 1975, p. 148—149), которые и моделируются как раз в рамках проблематики искусственного интеллекта и, в частности, с помощью фреймовой методики.

<sup>67</sup> См., напр.: Гак В. Г. Высказывание и ситуация. — В кн.: Проблемы структурной лингвистики. 1972. М., 1973.

<sup>68</sup> Васильев В. И. Язык представления знаний. — В кн.: Тезисы докладов 2-го семинара «Интеллектуальные банки данных». Ереван, 1980, с. 28—29.

в языке KRL основан на структурировании понятийных объектов предметной области в виде описаний, объединенных в специальные структуры, названные единицами. Каждая единица имеет собственное имя (ср. наше входное слово. — Ю. К.) и один или несколько поименованных слотов (ср. наши семантические множители в виде квази-основ), содержащих описания событий, связанных с понятийными объектами, относящимися к единице в целом... Таким образом, единица является формальной структурой данных в языке KRL, подобно тому как фрейм является специализированной структурой данных языка FRL и т. д.»<sup>69</sup> Углубляя аналогию дефиниции с фреймом, мы можем сослаться на более детальное определение отдельных подструктур последнего<sup>70</sup>, а расширяя ее, указать на то обстоятельство, что фрейм как агрегат семантической сети занимает по отношению к ней такую же позицию, какую словарная дефиниция занимает по отношению к концептуальной схеме общеязыкового тезауруса (ср. выше § 5 и § 8).

Оперирование предикатами и аргументами для конструирования семантического представления предложения как будто сближает фрейм-подход в его истоках с аналогичными построениями порождающей семантики. Однако в силу различной целевой направленности двух этих типов анализа (исследовательской в последнем случае и «разработнической», конструктивистской — в первом) обнаруживается своеобразное «оборачивание метода» в трактовке предикатов при разработке семантических сетей с помощью фрейм-методики. Если предикат в рамках порождающей семантики воспринимается как пропозициональная функция, которой в поверхностной структуре может соответствовать в принципе любое грамматическое средство (носителем отношений может быть предлог, падеж, отдельная полнозначная лексема, порядок слов и т. п.), то в системах с использованием семантических сетей в роли предиката выступает даже не «сказуемое суждения», как в классической логике, а уже только глагол или глагол-сказуемое предложения. Например, в диалоговой информационной системе по экономике<sup>71</sup> «понимание» запроса обеспечивается этапом семантического анализа, в процессе которого входное предложение перерабатывается во внутреннее представление (семантический граф), удобное для поиска информации в конкретной сети и базе данных. Семантический граф, эксплицитно задаю-

ший смысл входного запроса, состоит из некоторого числа вершин-понятий, связанных между собой через вершины-предикаты, и для опознания запроса накладывается на семантическую сеть, совмещаясь с каким-то ее участком. При этом выделяются три группы предикатов: предикаты действия, функциональные и пустые. «Предикаты действия характеризуют состояние объектов, которые реально существуют и которые можно наблюдать в моделируемом мире. Например: *поставлять*, *выпускать*, *передавать*, *перевозить*, *находиться* . . . Предикаты функционального типа представляют операторы над предикатами действия. Они указывают, какие операции необходимо совершить со вторыми для получения ответа на запрос. Например: *выполнить*, *уменьшить*, *перевыполнить*. Функциональные предикаты не имеют представителя в семантической сети. . . Все синтаксические отношения, связанные с функциональными предикатами, перевешиваются на подчиненные им предикаты действия. К пустым предикатам относятся собственно пустые глаголы (*осуществлять*), модальные глаголы, связки. Эти глаголы не несут реальной семантической нагрузки и в глубинном представлении не участвуют. Они выполняют служебную роль, поэтому семантический анализ опускает их, перенося все поверхностные отношения на подчиненный им синтаксический объект. В предложении всегда должен быть предикат действия, а это означает, что, в конце концов, все обстоятельственные отношения, комплетивные и предикативные, будут отнесены к этому предикату»<sup>72</sup>. Такое, может

<sup>69</sup> Рыбина Г. В., Храмов А. А. Язык представления знаний. — Там же, с. 29.

<sup>70</sup> «В настоящее время в ИС (интеллектуальных системах. — Ю. К.) все чаще используется понятие фрейма как для представления понятийных знаний, так и для анализа фраз естественного языка. В докладе дается определение F-фрейма:  $F = \lambda (x_1 : w_1) (x_2 : w_2) \dots (x_m : w_m) \Phi$  ( $\langle \text{comm}, x_1 \rangle$  ( $\langle \text{comm}, x_2 \rangle \dots (\langle \text{comm}, x_m \rangle)$  . . .)), где  $x_1, x_2, \dots, x_m$  — свободные переменные,  $w_1, w_2 \dots w_m$  — области интерпретации соответствующих переменных; *comm* — комментарий, в качестве которого выступают ролевые метки, определяющие место и значение переменной во фрейме. Конструкция  $\langle \text{comm}, x_i \rangle$  называется слотом. . .» (Вольфенга-ген В. Э., Дубинина, В. Г., Начинкин А. Д. Интеллектуальные банки данных и исчисление  $\lambda$ -конверсии. — Там же, с. 17).

<sup>71</sup> Бакланов В. М., Родионов Е. В. Система, воспринимающая запросы на естественном языке: семантика и формирование ответа. — Вопросы кибернетики, вып. 55. М., 1979.

<sup>72</sup> Бакланов В. М., Родионов Е. В. Указ. соч., с. 185—186.

быть несколько прямолинейное, приравнивание логико-семантических предикатов к грамматической категории глагола требует в подобных системах введения ряда дополнительных условий, связанных, в частности, с различной методикой обработки и анализа словосочетаний «с моделями управления» и именных словосочетаний, так что в целом степень адекватности этого приравнивания оказывается достаточной для используемого ограниченного подъязыка.

В нашем семантическом кодировании смысла дефиниции по квазиосновам не предусмотрено специального, методически поддерживаемого противопоставления аргументов и предикатов, за исключением особого положения предикатов, соответствующих нулевым множителям (к которым относятся имена существительные и глаголы специфической семантики и практически все причастия) и избыточным (парные антонимические прилагательные, образующие шкалы субъективных оценок). Основания подобного неразличения строятся на двух предпосылках. Во-первых, поскольку мы оперируем данными не специального подъязыка или ограниченного естественного языка, а считаем, что обрабатываемый массив представляет словарь и язык в целом, то мы можем отказаться, и реально в нашем кодировании отказываемся, от некоторых ограничений, не проводя различий, например, между частями речи. Известно, что чем более узкая область или часть объекта исследуется в том или ином опыте, в том или ином анализе, тем больше условностей надо соблюсти исследователю, тем большее число ограничений надо наложить на материал и приемы анализа. Расширение сферы исследования, понятно, освобождая от одних ограничений, заставляет вводить другие, связанные уже с объемами обрабатываемого материала и временем обработки. Но зато в нашем случае предварительное, заранее до опыта предусмотренное различение предикатов и аргументов в предложениях-дефинициях мы можем не проводить, ссылаясь на то, что словарь анализируется целиком, а это значит, что объективно заложенные в нем свойства, сущностно характеризующие его закономерности, с неизбежностью будут учтены и соответствующим образом проявят себя при таком подходе.

Вторая предпосылка должна выявить, насколько полно использованный нами способ представления смысла предложения-дефиниции по квазиосновам отражает объектив-

ные, существенные свойства языка. Эта предпосылка отчасти является следствием и развитием первой, но имеет и собственные истоки. В свое время, рассматривая вопрос о типах имен, или дескрипторов, в словаре-тезаурусе, мы выделили две их разновидности: имена денотативного (ситуативно-тематического) характера (*город, жилье, дом, мебель, больница, театр, одежда, почта*) и имена десигнаторного, или сигнификативного) характера (*лечить, промышленный, космос, национальный, животное, двигаться, организовать*), противопоставление которых друг другу основано на том, что первые предполагают перечисление подчиненных им и включающихся в них элементов и структурируются на предметном отношении целого и части, тогда как вторые отражают главным образом логические родо-видовые связи и предусматривают равноправно-альтернативные отношения между составляющими их автономными видами<sup>73</sup>. Отправляясь от этого наблюдения и распространяя его на все единицы словаря, а не только на дескрипторы, Ю. С. Степанов показал, что различия между двумя типами имен гораздо глубже и коренятся в самой природе слов, одни из которых как бы предназначены к тому, чтобы быть в языке предикатами по преимуществу, а другие выступают всегда лишь в роли аргументов<sup>74</sup>. Эта мысль об онтологических (морфологических) основах деления имен в естественном языке на предикатные и аргументные поддерживается и данными словообразовательного анализа, по результатам которого можно выявить принципиальные различия в структуре соответствующих лексических гнезд<sup>75</sup>. Следовательно, квазиоснова, будучи своеобразным intermediate impossible, т. е. образованием, которому ничто реально не соответствует в структуре языка, с неизбежностью вбирает в себя существенное свойство предикатной либо аргументной принадлежности той группы слов, которую она представляет и замещает. А это и объясняет, почему семантическое кодирование по квазиосновам позволяет, не проводя явного разделения между предикатами и аргументами в словарной definicции, имплицитно учесть его, определенным образом лексикализовав, т. е. инкорпорировав синтакси-

<sup>73</sup> Караполов Ю. Я. Общая и русская идеография, с. 132–134.

<sup>74</sup> Степанов Ю. С. Имена, предикаты, предложения. (Семиологическая грамматика). М., 1981, с. 46–74.

<sup>75</sup> Гинзбург Е. Л. Словообразование и синтаксис. М., 1979, с. 58–69.

ческую информацию и ее отражение в глубинном семантическом представлении предложения в морфологический облик кодируемого слова.

Избранный нами путь приведения к общему знаменателю синтаксического и лексического значения в предложении, основанный на лексикализации и морфологизации синтаксического компонента, релевантен, вероятно, только для текстов объяснительного характера, каким является словарное определение. Естественно, и по отношению к текстам этого типа данное решение не является единственным, и можно предложить его альтернативу, опирающуюся, например, на компонентное представление суммарного смысла лексических и синтаксических слагаемых дефиниции, т. е. распространить на изучение предложения методику, широко и с успехом применяемую к анализу словосочетаний. Вместе с тем мы отдаем себе отчет в том, что любое подобное решение, давая более или менее адекватный задаче результат, будет оставаться конструктивным паллиативом — весьма полезным, в ряде случаев единственно возможным, однако приоткрывающим лишь часть истины. Какое-то время наука может удовлетворяться такими просто адекватными решениями, но в конце концов наступает момент, когда остро осознается необходимость идти дальше, двигаться по ту сторону адекватности, а это было и остается прерогативой лингвистической теории, которой и принадлежит решающее и пока не сказанное слово в вопросе о соотношении семантики и синтаксиса.

## SUMMARY

The Introduction lays down general principles of aim-oriented approach in linguistics referred to as linguistic constructing which holds that in science the answer to the question «how to make» one or another object is often the very best answer to the question «why». The distinctive features of such approach are as follows: a) a trend to make «a new linguistic object» generating a new information about language, which is connected as a rule with experimenting on the natural language; and as a consequence of this b) coverage of all language structure levels in solving individual, even particular task; c) wide application of non-absolute, i. e. technical solutions, palliatives or ready-made patterns in the field where objective laws have not as yet been found or cannot be found. The heuristic methods of linguistic constructing comprise unusual categorization of the area or object under study, introduction of new description units, application of «reverse in the method» («Umschlag in der Methode») principles and averaging principles. The relations between linguistic constructing and theory are those of mutual supplement, since the principles of the former are functioning in the area where theory has not as yet been developed; linguistic constructing differs from pure empiric approach being more distant (but not as much as linguistic theory) from the object of study (since between the former and the language as an object there lies a «new linguistic object») and having an aim to comprehend theoretically or «to ontologize» successful and resultful technical solutions interpreting them in the light of the language essence properties.

§ 1 is devoted to characteristics of a new feature for linguistics of the last decade which is manifested in the need to make the results of study (of different, and ultimately of all language levels, language units and their relations)

lexicographically arranged and extrapolated into dictionary. Such lexicographic reduction, registration of language phenomena of various classes in the form of dictionaries is exercised in a specific type of lexicographic parameters which can either agree or disagree with traditional categorization of language as an object and traditional description units. There are 2 variants of this tendency in the world lexicography: centrifugal tendency expressed in compilation of one-parameter dictionaries (orthographic, accentological dictionaries, dictionary of synonyms and etc.), and centripetal tendency which supposes, in an extreme case, unification of all language characteristics and all parameters in a single universal dictionary.

There are distinguished 2 types of lexicographic parameters — structuregeneous parameters associated with ontological properties of language (pronunciation, gender, morphological segmentation of a word, word changing parameter and etc.) and linguistic parameters covering pragmatic, denotative, historical-and-cultural characteristics of language units and phenomena, connected not only with language mechanism and its functioning but with conditions of its spreading, existence, study, i. e. with linguistics in particular (language-orientated-cultural-studies, language history, terminological, bibliographic, emotionally-appraising and other parameters). The former parameters are characterized by the meaning (e. g. the gender parameter in the Russian language acquires one of three meanings — masculine, feminine, neuter); the latter parameters are characterized by the depth (compare the possibility of broad variations in detailed presenting of areal or etymological parameters in one or another dictionary). No principle restrictions are imposed upon parameters to form combinations with each other, and therefore in the dictionary the law of freedom of parameters is in effect.

§ 2 contains matrix presentation of 67 lexicographic parameters picked out on the basis of the most complete world dictionaries («a matrix of discovery») which makes it possible to control the purpose and contents of the dictionary produced as well as to compile new, in principle, types of dictionaries (e. g. a reverse dictionary of loanwords, onomasiological phraseological dictionary, the dictionary of semantic components, the dictionary of grammar meanings, and etc.) by implementing the choice of the

number of parameters from those inherent in a given language, the method of their setting and depth.

§ 3 develops the parametric approach to the analysis and the computerized constructing of dictionaries.

In § 4 there are discussed questions of general structure of thesauri, the specific character of this type of dictionaries being seen in their increase of dimensions of the reflected semantic space of the language. This specificity is manifested in the possibility to realize 4 inputs into the dictionary formed by the relations between «sign» and «concept»: concept—concept (notion scheme of thesaurus); concept—sign (descriptor or ideographical part); sign—concept (index of words with descriptors addresses); sign—sign (permutation index). The existing types of terminological and general thesauri (thematic, analogical, associational, ideographic) are within the offered classification scheme differing from each other in presence-absence of some inputs or a combination of any 2 inputs in one.

§ 5 gives the answer to the question «Why is thesaurus necessary» meaning the sense which has become disseminated of late according to which thesaurus is considered to be both a real lexicographic object («thesaurus-dictionary») and an ideal method of semantic space organization («thesaurus-world»). Generalizing the whole set of system relations in lexics which, as a rule, are studied separately (synonymy, associativeness, polysemy, etc.) thesaurus-dictionary serves as an analogue of the lexical system itself, differing from other methods of its embodiment in the complexity first of all. The place occupied by thesaurus in lexicography is determined by its interrelations with the explanatory dictionary, which gives definitions — explicitly or implicitly — taking into account the hierarchy of descriptors, on which the word in question is dependent in a real or implied thesaurus so that the connection of one definition with another and the depth of each of them are subordinated to «the rule of six steps». Fourdimensional thesaurus semantic space turns out to be adequate not only to fix the lexical semantic sphere but to present knowledge of the world reflected in the human language both in special branches and in its universal expression (Encyclopaedia Britannica). The above feature of thesaurus organization brings us to understanding of premises to carry it to the mind of an individual and to identify it with the lexicon organi-

zation of a native speaker of the language. This consideration is argumented by psycholinguistic and heurological data.

The understanding of thesaurus' role in the process of speech generation and speech perception shows that the former proceed within the dimensions in which the final point is «the sign» (i. e. concept—sign and sign—sign) whereas the orientation of perception processes is contrary and the hearer's and reader's position is characterized by dimensions the final point of which is «concept» (i. e. sign—concept and concept—concept). Prognostic functioning of the thesaurus in the speech generation process is illustrated by some peculiarities of creative work in producing texts; the examples of diagnostic functions of the thesaurus are the mechanism of reader's perception and storage in memory of the text from a work of literature and art given in the book as well as similar to mechanisms of «fast reading», indexation («table reading») and content-analysis.

In accordance with the principle «how to make» § 6 determines the necessary amount of initial, axiomatically set parameters of the Russian thesaurus and parameters deduced which are subject to preliminary analysis prior to automatic processing. Descriptors catalogue and vocabulary of the thesaurus under construction are formed by the method of ordinary compilation. The set of semantic connections is established as a result of the analysis of definition content, with an instrument of such analysis being the semantic coding of word definitions' information with the help of quasi-stems which act as semantic components. Determination of three massif composition — descriptors, words and semantic connections (semantic components) and postulation of words' distribution algorythm according to descriptors based on community of semantic components conclude the organization of the thesaurus model called a phenomenological one.

§ 7 is dedicated to comprehension of preliminary results of phenomenological model action and to error correction in automatic word distribution according to descriptors (semantic fields). The analysis of errors leads to the necessity of semantic components ranging on the basis of their frequency and of introduction to distribution algorythm the selective statistic criterion. The final automatic production model of the Russian thesaurus regarding the semantic

component statistic rank which singles out zero and redundant components along with the relevant ones becomes known as asymptotic model.

Chapter IV, being the last one, is an attempt to «ontologize» a number of technical solutions which were found in the process of constructing the Russian thesaurus by computer, and particularly to make things somewhat clearer in the problems of interrelations between semantics and syntax in explanatory texts to which should be referred definitions in explanatory dictionaries; this chapter is also an attempt to find the «common denominator» to present lexicological and syntactical semantics in them.

§ 8 says that when establishing the status of definition sentences the investigator faces paradoxes caused by uncertainty of definition communicative orientation. These paradoxes are accounted for: definition ambiguous status both propositional and non-propositional statements, i. e. with the impossibility to solve the problem simply whether the definition sentences belong to propositions; then, — reversibility of rheme—theme relations in them; and finally — vagueness of interpretation for them of presupposition notion, i. e. the problem of whether they should be included into a broader context or not. Consideration of definitions' paradoxes makes possible to formulate a number of conditions under which correlation of semantics and syntax study should be carried out in sentences of this type.

§ 9 forecasts in the most general outline three types of summary sense constituents of identity sentences; these are: a) the meaning of individual lexemes which reflects denotative aspect of definition semantics; b) logical and semantic relations among referents of individual lexemes in extralinguistic reality (relational significance); and further; c) the syntactical constituent of sense taking account of language units' connections resulting in the definition text on the syntagmatic axis.

The analysis of lexical (denotative) constituent shows that the final sense of the definition is, in a number of cases, not only the result of addition (conjunction) of semantic components but of their difference (disjunction). When coding, there should be excluded those features which cause hypersemantization in definition and are related to characteristics of not specific differences but of individual

peculiarities of the object or phenomenon defined, to the technology of its production or application, to the description of sequence of operations in functioning or implementing something, etc. Those lexical elements of a definition text, which do not display their differentiating function when semantically recorded by quasi-stems, i. e. they are unable to bring strictly the word defined into correlation with one or another descriptor or semantic field thus engendering unpertinent information, are called redundant elements. Apart from the elements reflecting individual features these comprise adjectives of subjective estimation (big—small, hard—soft, narrow—wide and etc.) and some semi-copulative verbs (to use, to get, to come and etc.).

The discussion of relational significance in definition sentences in § 10 leads to revelation reflected in them of various types of logical—and—semantic relations among the objects of the real world: to be a subject and to have as a subject, to be a cause and to have as a cause, to be an element of a class, to incorporate, to be similar to, to be a condition. Contrary to local relations reproduced by redundant components, of occasional character and relevant only for a particular definition, logic—and—semantic relations are universal, reflect the structure of basic connections in semantic space of lexics (in the thesaurus) and may be calculated. As to the ways of their expression they are reproduced by nouns of vague semantics (porridge words), copulative words, the greater part of semi-copulative verbs, participles, prepositions, word order. In the process of semantic recording by quasi-stems the relational significance is conveyed by semantic components, referred to as zero ones.

To determine the share of syntax constituent in definition semantics § 11 analyses their structure by methods of functional syntax, transformational grammar, generative semantics, functional stratigraphy and frame theory. The comparison of the results of these approaches with the possibilities of semantic recording by quasi-stems, which by three types of components — relevant, zero and redundant ones — embraces denotative and relational aspects of the summary meaning of definition, allows to make two conclusions:

- 1) The existing methods of analysis of syntactical semantics do not reveal any new constituents of meaning which have not been taken into account while considering

the relational aspect of semantics and ave not been reflected in redundant and 'zero semantics components when recording them by quasi-stems.

2) So the recording by quasi-tems as a method of linguistic constructing turns out to be an adequate way of presenting the sense of definition sentences, reducing a lexical-and-syntactical semantics in them to the same denominator; the quasi-stems themselves fixing, in the latent form, the predicate—and—argument structure of syntax construction reflect the ontological foundations of lexical units' division in natural language into predicative and argumentative ones.

## УКАЗАТЕЛЬ ИМЕН

- Абдурахманов М. 41  
Абдушелишвили М. Г. 25  
Абрамов В. П. 338, 339  
Агабаян К. Г. 231  
Адлер М. Дж. (Adler M. J.)  
Адрадос Ф. Р. (Adrados F. R.)  
Александров В. А. 227  
Алексеев М. П. 189  
Алексеев П. М. 37  
Алекторова Л. П. 181  
Андреева И. В. 63  
Андрющенко В. М. 125  
Арапов М. В. 35, 223  
Аристотель 280  
Арутюнова Н. Д. 277, 279,  
280, 283, 285, 339  
Ахманова О. С. 39
- Бабкин А. М. 55  
Бабов К. 155, 157, 227  
Бакланов В. М. 347  
Балли III. 178, 179  
Баратынский Е. А. 121  
Бардович А. М. 63  
Барт Р. 202, 203  
Бартлетт Дж. 122  
Бархударов Л. С. 330, 331  
Баско Н. В. 63  
Батюшков К. Н. 121  
Беляева Л. Н. 37, 39, 69, 131  
Верков В. П. 53, 59, 231  
Бертель А. Е. 182, 183  
Бирюков Б. В. 280  
Богданова И. Г. 177  
Бондалетов В. Д. 37  
Боно Е., де (Bono E. de)  
Бонфантэ Дж. (Bonfante G.)  
Борецкий М. И. 213  
Ботвинник М. М. 31, 249  
Брутян Г. А. 190, 191, 227  
Букэ М. 39  
Буш Г. Я. 147
- Ванников Ю. В. 125
- Вариг Г. 58, 59, 68 (см. также  
Wahrig G.)  
Вартбург В. 172, 179 (см.  
также Warthburg W.)  
Васильев В. И. 345  
Васильев Г. М. 39  
Введенский Л. А. 53  
Веревченко А. П. 159  
Верещагин Е. М. 63  
Вернадский В. И. 6  
Вертель В. А. 117  
Вертель Е. В. 117  
Викс Ю. А. 73, 111, 137 (см.  
также Wiks O.)  
Виленская С. К. 35, 37  
Виноград Т. 145, 237  
Вовшин Я. М. 63  
Вольфенгаген В. Э. 345, 347  
Воргулев А. (Въргулев А.)  
Воробьев Г. Г. 191, 215, 217  
Воскресенская О. В. 345  
Вотяк Г. (Wotjak G.)  
Въргулев А. 155, 157, 227
- Гак В. Г. 345  
Гаспарян А. С. 343  
Герасимов В. Н. 41  
Герд А. С. 35, 175, 223  
Гессе Г. 198, 205, 206, 212  
Гете И.—В. 218  
Гинзбург Е. Л. 57, 349  
Гораций К. Ф. (Horatius)  
Горбанев Ю. Г. 345  
Горбачевич К. Г. 25, 41, 55  
Горбис Б. И. 63  
Городецкий Б. Ю. 111  
Горский Д. П. Из. 280  
Горький М. 43, 57  
Готлиб К. Г. М. 39  
Граллерт В. 39  
Граудина Л. К. 39  
Грачева Н. А. 117  
Грдличка М. 63

- Григорьев В. П. 65, 203  
 Грушке В. 39  
 Грязнухина Т. А. 293  
 Даль В. И. 41, 66, 67, 245, 246  
 Декарт Р. 34, 74, 226, 280  
 Дементьева Н. В. 177  
 Денисов П. Н. 43, 185  
 Денисова М. А. 63  
 Дерибас В. М. 41  
 Дорнзайф Ф. 155, 184, 212,  
     227 (см. также Dornseiff F.)  
 Дубинина В. Г. 347  
 Евсеев И. 39  
 Егорова Е. В. 35, 133  
 Ельмслев Л. 175, 176  
 Есенин С. 201  
**Жергонн Ж. Д. ИЗ**  
 Залевская А. А. 191–193  
 Зализнак А. А. 41, 59, 115  
 Зиман Ю. Л. 35, 133  
 Золотова Г. А. 41, 327  
 Зотова Т. Я. 235  
 Иванов Вал. Вас. 39  
 Иванов Вяч. Вс. 193  
 Иванова Т. Ф. 43  
 Иванченко В. В. 343  
 Ингарден Р. 212, 213  
 Ингве В. 231  
 Ицкович В. А. 39  
 Каменская О. Л. 53  
 Капица П. Л. 21, 30, 31  
 Касарес Х. 71, 149–157, 160,  
     167, 169, 182, 197–200, 227,  
     229 (см. также Casares J.)  
 Катлинская Л. П. 39  
 Катц Дж. 332 (см. также  
     Katz J.)  
 Клышка М. К. 39  
 Кокорева Л. В. 343  
 Колесников Н. П. 39, 53  
 Колин А. Ж. Т. 117  
 Коломиец В. Т. 68, 69  
 Костомаров В. Г. 63  
 Котелова Н. З. 45  
 Крисевич В. С. 37, 131  
 Кристева Ю. (Kristeva J.)  
 Кубурлис Д. (Koubourlis D. J.)  
 Кузнецов О. А. 215  
 Кун Т. 6, 7  
 Курбаков К. И. 193  
 Кюппер Г. (Kiipper H.)  
 Лакофф Дж. 332  
 Лангакер Р. (Langacker R. W.)  
 Ларин Б. А. 56, 57  
 Леви-Строс К. 281  
 Левый И. (Levy I.)  
 Лейбниц Г. В. 23, 210, 226  
 Ленин В. И. 43  
 Леонтьев А. А. 39, 163  
 Лепинг А. А. 43  
 Липницкий С. Ф. 37, 131  
 Локк Дж. 224, 280  
 Лурия А. Р. 218  
 Львов М. Р. 53  
 Мадоян Р. 223, 301  
 Макаров М. Г. 9  
 Макколи 332  
 Малкиль Ж. (Malkiel J.)  
 Малов В. Н. 59  
 Маркс К. 21  
 Марчук Ю. Н. 35, 41, 49, 109  
 Маршакова И. В. 37, 223, 235  
 Мариотина А. И. 139  
 Маслов С. Ю. 19, 21, 197  
 Маяковский В. В. 201  
 Милославский И. Г. 39, 57, 63  
 Минский М. 343  
 Мириманова М. С. 223  
 Молинер М. 153, 165, 329 (см.  
     также Moliner M.)  
 Монтень М. 325  
 Морковкин В. В. 37, 148, 149,  
     155, 175, 183, 223, 227  
 Москальская О. И. 327  
 Мукаржовский Я. 202, 205  
 Муравьев В. Л. 65  
 Налимов В. В. 33  
 Начинкин А. Д. 347  
 Недялкова Г. И. 231  
 Нейберт Г. 133  
 Никитин М. В. 275, 277  
 Никитина С. Е. 185, 187, 223  
 Ожегов С. И. 44, 45, 57, 171,  
     179, 244, 245, 253, 271  
 Олдырева Л. П. 69  
 Оныкий Б. Н. 343  
 Павловский О. М. 25  
 Падучева Е. В. 275, 333

- Палагина** В. В. 63  
**Даллас** П. С. (Pallas P. S.)  
**Панькин** В. М. 39  
**Паркер—Роудс** А. Ф. 179  
**Паскаль** Б. 280  
**Пассмор** Дж. (Passmore J.)  
 Петров Ю. Г. 159  
 Пиотровский Р. Г. 37, 131  
 Пирс Ч. 199  
 Планк М. 20, 22  
 Плотников Б. А. 177, 235  
 Поллукс Ю. 182  
 Половинкин А. И. 147  
 Попа К. 113, 280, 281  
**Поповцева** Т. Н. 295  
 Прохоров Ю. Е. 63  
 Прут М. 202  
 Пушкин А. С. 43, 109, 120, 121  
 Пшеничная Л. Э. 177, 201 293, 309, 323, 329  
 Пюршель У. (Fürscheil U.)  
 Рангантан Ш. Р. 184, 185  
**Ратцев** И. 279  
 Робер П. 153, 165 (см. также Robert P.)  
**Рогожникова** Р. П. 117  
 Родионов Е. В. 347  
 Роже П. М. 142, 144, 150, 152, 169, 212, 227 (см. также Roget P. M.)  
 Розанова В. В. 233  
 Розенталь Д. Э. 43  
 Розов В. 162  
**Ружичка** Р. (Růžička R.)  
 Рыбина Г. В. 347  
**Салтыков-Щедрин** М. Е. 166  
 Свифт Дж. 29  
 Седлова Л. П. 177  
 Селезнев И. Ф. 67  
 Селиверстова О. Н. 333  
 Сенека Л. А. 7  
 Сепир Э. 10  
 Сердюкова Н. Н. 53  
 Сидорченко В. Д. 223  
 Симонов П. В. 203  
 Скородяко Э. Ф. 275, 293  
 Скреплина Л. М. 27  
 Соколов А. В. 35, 155, 169, 183, 221, 223, 317  
 Соколовская Ж. П. 172, 173  
 Сороколетов Ф. П. 55  
**Сохор** А. М. 37
- Станчев Н. Т. 37  
**Старинец** В. С. 231  
 Степанов Ю. С. 23, 171, 349  
 Степанова М. Д. 39  
 Страхова Н. П. 43  
 Строганов В. А. 279  
 Тарский А. 286  
**Теленкова** М. А. 43  
 Тимофеев В. П. 38  
 Титова Л. Н. 39, 53, 163  
 Тихомиров Б. Д. 41  
 Тихонов А. Н. 39, 115, 159  
 Тодоров Ц. 203 (см. также Todorov Tz.)  
 Тышлер И. С. 55  
**Убин** И. И. 41  
**Уемов** А. И. 291  
 Улуханов И. С. 57  
 Уорф Б. Л. 10  
 Ускова Т. В. 343  
 Ушаков Д. Н. 57, 59, 77, 78, 149, 245, 246, 253, 271  
 Уэбстер Н. 48, 49, 55, 144  
 Федоров А. И. 63  
 Феррари Дж. (Ferrari G.)  
 Филиппов А. В. 39  
 Филлмор Ч. 131, 144, 332  
 Флобер Г. 202  
 Фодор Дж. А. 332  
 Фрумкина Р. М. 233  
 Хабло Е. И. 41  
 Халлиг Р. 172, 179 (см. также Hallig R.)  
 Храмов А. А. 347  
 Хромов Л. Н. 215  
**Цонев** В. А. 9  
 Цывин А. М. 43, 175  
 Чантуришили Д. С. 65  
 Черняк В. С. 21, 23  
 Шакун Л. М. 63  
 Шаляпина З. М. 37, 133  
 Шанский В. Н. 63  
 Шанский Н. М. 39  
 Шахматов А. А. 66  
 Шведова Н. Ю. 179—182  
 Шекспир В. (Shakespeare W.)  
 Шемакин Ю. И. 151, 159  
 Шоу Дж. Т. (Shaw J. Th.)
- Шрейдер Ю. А. 35, 223  
 Штинглова И. 117  
 Щекина И. А. 177  
 Шерба Л. В. 56, 57, 174, 175, 191  
 Щербинин В. И. 41  
 Эйнштейн А. 20  
 Эткис П. 43  
 Якобсон Р. 202, 203 (см. также Jakobson R.)  
 Якушин В. В. 215  
 Ясперс К. (Jaspers K.)  
 Adjemian Ch. 15  
 Adler M. J. 189  
 Adrados F. R. 345  
 Andersson B. 217  
 Andrews W. 147  
 Aschiero H. R. 55  
 Bailey R. W. 127  
 Baldinger K. 65  
 Baratynskii E. A. 121  
 Barret E. F. 55  
**Batjushkov** K. N. 121  
 Bender A. D. 147  
 Bonfante G. 69  
 Bono E. de 148, 290, 305  
 Boussinot R. 39, 53  
 Bruner Th. F. 127  
 Casares J. 39, 149, 169, 199, 219  
 Caude R. 74  
 Chauzaud H. B. du 53  
 Constantinescu S. 55  
 Czigler J. 319  
 Daum E. 41  
 Deweze A. 133  
 Dijk T. A. van. 205  
 Diller T. 143  
 Dornseiff F. 219  
 Edmundson H. R. 129  
 Epstein M. N. 129  
 Ferguson Ch. A. 33  
**Ferrari** G. 129  
 Filipec J. 55  
 Finkenstaedt Th. 123  
 Flexner S. B. 37
- Fodor J. A. 53  
 Francis W. N. 143  
 Gilbert P. 39  
 Goethe J. W. 43, 218  
 Gove P. B. 49  
 Greimas A. 53  
 Hallig R. 179, 227  
 Hayakawa S. I. 53  
 Heisenberg W. 225  
 Heller L. G. 45  
 Hernádi M. 41  
 Holsti O. R. 63, 217  
 Horatius 273  
 Huntsman J. F. 141  
 Jakobson R. 201  
 Januschek F. 33  
 Jaspers K. 213  
 Ikola O. 73, 123  
 Ingram W. 121  
 Karjalainen Y. 73, 123  
 Katz J. 53  
 Kelly E. F. 133—135, 217, 218  
 Kintsch W. 205  
 Kiss G. 141, 165  
 Koubourlis D. J. 117—119  
 Kristeva J. 205  
 Kučera H. 143  
 Küpper H. 60, 61  
 Laffal J. 141  
 Langacker R. W. 341  
**Leemjng** H. 69  
 Leisi E. 123  
 Leonard R. 127  
 Levy I. 201  
 Lodwig R. R. 55  
 Macris J. 45  
 Malkiel J. 175  
 Mandelstam O. 117  
 Martin J. 37  
 Martinet A. 39  
 Maxwell E. R. 133  
**McPherson** J. 147  
 Mepham M. 61  
 Moles A. 75  
 Moliner M. 153  
 Nentworth H. 37  
 Neuhaus H. J. 123

- Oiler J. W., Jr. 277  
Olney J. 143  
Paikday T. M. 127  
Pallas P. S. 66  
Papp F. 129  
Passmore J. 29  
Petöfi J. 203  
Pietro R. J. di 33  
Pleh Cs. 319  
Piirschel U. 274  
  
Remmel M. 73  
Robert P. 153  
Robertson T. 219  
Robinson J. L. 127  
Rodale J. I. 41  
Røget P. M. 151, 169, 219  
Růžička R. 331  
  
Saareste A. 55  
Scheler M. 213  
Schenk W. 41  
Scherman D. 53, 55, 63, 135  
Schröer M. M. A. 39  
Schumann J. H. 15  
  
Shakespeare W. 122, 123  
Shaw J. Th. 119—121  
Shibayev V. 71  
Smith R. N. 127, 133  
Spevack M. 123  
Stachowitz A. 133  
Stone Ph. J. 133—135, 217, 218  
  
Todorov Tz. 203  
Townsend W. C. 55  
  
Valesiq P. 19  
Viks U. 25, 35, 73, 117  
Wahri? G. 27, 37, 47—49, 53  
Walter H. 39  
Wartburg W. von 179, 227  
Weinreich U. 275  
Wilson A. G. 75  
Wilton M. T. 133  
Wolf D. 123  
Wotjak G. 167  
  
Zettersten A. 127  
Zwický F. 75

## УКАЗАТЕЛЬ КЛЮЧЕВЫХ СЛОВ

- алгоритм кодирования 171, 241—244, 277, 337  
алгоритм распределения (слов по дескрипторам) 236, 256—259, 262  
аргумент (см. также предикат) 143, 332—338, 340, 343, 344, 346, 348, 349  
вход в словарь 71, 75, 78, 112, 149, 152, 154, 158—161, 163—166, 182, 185, 187, 194, 220, 221  
типерсемантизация (см. также избыточные семантические множители) 296, 301  
глубина тезауруса 186—189, 222, 290  
декодирование 21, 241, 252, 254  
дескриптор (ср. ключевые слова) 19, 150—158, 165, 468, 169, 171, 177, 181, 189, 194, 215, 217, 218, 220, 225, 230, 233, 236, 239, 244—246, 249—257, 259—271, 274, 295, 303, 305, 311, 312, 319, 349  
дескрипторные семантические множители 249, 250, 255, 256, 261, 305  
дескрипторов каталог (массив) 228—230, 245, 247, 256, 257, 339  
закон свободы параметров 65—68, 70  
замкнутость семантическая (словарника, поля, системы) (ср. открытость) 173, 181, 182, 232, 251  
зона нерелевантности (семантических множителей) 270, 303—305, 311, 315, 321, 338  
избыточные семантические множители 238, 271, 296, 300—305, 319, 345  
категоризация объекта (предметной области) 23, 31, 40, 274  
квазиоснова 24, 25, 240, 241, 250—254, 335, 337, 343, 348, 349  
ключевые слова (ср. дескриптор) 150—157, 160, 165, 167, 171, 187, 194, 215  
кодирование (см. семантическая запись по квазиосновам)  
компиляция 16, 114, 116, 119, 127, 137, 227  
компонент (см. семантический множитель)  
лемматизация 122, 238, 239  
матрица открытия (=матрица параметрического анализа словарей) 42, 45, 71, 74, 75  
новый лингвистический объект 16—18, 31, 32, 33, 75, 138, 140  
нулевизация множителя 305, 310, 311, 316  
нулевые семантические множители (компоненты) 238, 239, 241, 249, 271, 295, 313—316, 318—324, 338—340, 343, 345  
оборачивание метода (принцип оборачивания метода) 20—22, 31, 172, 176, 346  
открытость (словарника, поля, системы) (ср. замкнутость) 173  
отношения логико-семантические 305, 309, 312, 313, 318, 319, 325, 337  
отношения локальные 302, 324, 325  
параметра глубина 59—61, 68, 70, 112  
параметра значение 58—62, 64, 65, 73, 111  
параметры лексикографические 40, 42, 44—46, 48, 53, 55, 56, 58—77, 109, 111—116, 121—124, 126—137, 140—145, 147, 193, 195, 221

- параметры лингвистические 70,  
112, 124  
параметры структурогенные 70,  
72, 73, 111, 124  
правило шести шагов 20, 186,  
231, 232, 270, 290  
предикат (см. также аргумент)  
143, 279, 332—339, 240, 344,  
346—349  
пропозиция 279, 283, 284, 289,  
339, 340  
размерность (измерение) семантического пространства 149,  
155, 158, 160, 164, 168, 185,  
199  
релевантные семантические множители (ср. избыточные и нулевые) 261, 303, 304, 315,  
322, 340, 345  
селективный частотный критерий 239, 260, 261, 263, 293,  
295, 311, 329, 343  
семантическая запись по квазиосновам («кодирование») 19,  
26, 171, 236—238, 240, 242—  
247, 251, 294, 299, 303, 305,  
309, 311, 316, 320, 335—338,  
340, 343, 344, 349  
семантическая непрерывность словаря 225, 231, 270  
семантическая связь 171, 176  
182, 225, 227, 233—236, 239,  
248, 249, 256, 263, 305, 311,  
312, 318  
семантический множитель 22,  
26, 54, 171, 176, 237—239,  
241, 242, 244—247, 249—257,  
259—261, 270, 338  
семантических множителей дизъюнкция 290, 293, 294, 300  
семантических множителей каталог 241, 248, 249, 253—256  
семантических множителей конъюнкция 290, 292, 300  
семантическое поле 27, 55, 144,  
163, 171, 176, 181, 192, 200,  
206, 207, 246, 249, 256, 257,  
269—272, 286, 289, 311—313,  
318  
семантическое ядро русской лексики 250, 255, 256, 270,  
271  
семантической связи сила 225,  
236, 263, 269  
словник тезауруса (=каталог слов) 227, 230—233, 241, 242,  
246, 249, 253—257, 339  
темо-рематические отношения 279, 280, 284, 285  
усреднение 24—26, 31, 251  
центробежная направленность (тенденции к лексикографической параметризации языка) 60, 61, 110  
центро斯特ремительная направленность (тенденции к лексикографической параметризации языка) 60, 65, 110

# ОГЛАВЛЕНИЕ

От редактора . . . . .	3
Введение. Цели и проблемы . . . . .	6

## Глава I ПРИНЦИПЫ

### ЛИНГВИСТИЧЕСКОГО КОНСТРУИРОВАНИЯ ПРИМЕНИТЕЛЬНО К СЛОВАРОСТРОЕНИЮ

§ 1. Тенденция к лексикографической параметризации языка . . . . .	34
§ 2. Параметрический анализ словарей . . . . .	74
§ 3. Пути автоматической обработки лексикографических параметров . . . . .	109

## Глава II

### КАК УСТРОЕН ТЕЗАУРУС, КАК ОН РАБОТАЕТ И ЗАЧЕМ НУЖЕН

§ 4. Тезаурус как способ представления семантического пространства . . . . .	148
§ 5. Сфера использования тезауруса, или задачи, которые он решает . . . . .	166

## Глава III

### ОПЫТ КОНСТРУИРОВАНИЯ РУССКОГО ТЕЗАУРУСА С ПОМОЩЬЮ ЭВМ

§ G. Феноменологическая модель тезауруса и поставленные ею проблемы . . . . .	219
§ 7. Асимптотическая модель тезауруса: реализация и интерпретация . . . . .	249

## Глава IV

### НЕКОТОРЫЕ ВОПРОСЫ СООТНОШЕНИЯ СЕМАНТИКИ И СИНТАКСИСА С ПОЗИЦИЙ ЛИНГВИСТИЧЕСКОГО КОНСТРУИРОВАНИЯ

§ 8. Статус и смысл дефинитивных предложений . . . . .	273
§ 9. Слагаемые смысла и лексический компонент словарных определений (денонативный аспект) . . . . .	290
§ 10. Отражение в дефиниции логико-семантических отношений реальности (реляционное значение) . . . . .	305
§ 11. Синтаксический компонент в семантике дефиниции	325
Summary . . . . .	351
Указатель имён . . . . .	358
Указатель ключевых слов . . . . .	363

# CONTENTS

Editor's preface . . . . .	3
Introduction. Aims vs. problems . . . . .	6
Chapter I	
THE PRINCIPLES OF LINGUISTIC CONSTRUCTING	
IN REFERENCE TO DICTIONARY MAKING	
§ 1. Trend to lexicographic parameterization of language . . . . .	34
§ 2. Dictionaries parametric analysis . . . . .	74
§ 3. Ways to lexicographic parameters automatic processing	109
Chapter II	
HOW THE THESAURUS IS ORGANIZED, HOW	
IT WORKS AND WHAT IT IS NECESSARY FOR	
§ 4. The thesaurus as a way of the semantic space representation . . . . .	148
§ 5. Lines of the thesaurus employment or problems which it helps to solve. . . . .	166
Chapter III	
THE EXPERIMENT OF THE RUSSIAN THESAURUS	
COMPUTER-CONSTRUCTING	
§ 6. Phenomenological model of thesaurus: problems arisen. . . . .	219
§ 7. Asymptotic model of thesaurus: realization and interpretation . . . . .	249
Chapter IV	
SOME ASPECTS OF SYNTAX-SEMANTICS	
RELATIONS FROM THE LINGUISTIC	
CONSTRUCTING POINT OF VIEW	
§ 8. Status and sense of definition-sentences . . . . .	273
§ 9. The constituents of sense and lexical part of explanatory texts in dictionaries (denotative aspect) . . . . .	290
§ 10. Logical-semantic relations of reality and its reflection in definitions (relational significance). . . . .	305
§ 11. Syntactical constituent in the definition semantics . . . . .	325
Summary. . . . .	351
Names index. . . . .	358
Index of clue-terms. . . . .	363

**Юрий Николаевич Караулов**  
**ЛИНГВИСТИЧЕСКОЕ**  
**КОНСТРУИРОВАНИЕ**  
**И ТЕЗАУРУС**  
**ЛИТЕРАТУРНОГО ЯЗЫКА**

Утверждено к печати  
Институтом языкоznания  
АН СССР

Редактор издательства *Н. Н. Барская*  
Художник *А. П. Кобрин*  
Художественный редактор *Т. П. Поленова*  
Технический редактор *Ф. М. Хенок*  
• Корректоры *М. В. Борткова, Л. П. Стрельчук*

ИБ Я» 22345

Сдано в набор 03.11.80. Подписано к печати 11.05.81  
Формат 84×108<sup>1/3</sup>, № 2  
Бумага типографская № 2  
Гарнитура обыкновенная  
Печать высокая  
Усл. печ. л. 19,2. Уч.-изд. л. 21,3  
Тираж 3300 экз. Тип. зон. 2089  
Цена 2 р. 50 к.

Издательство «Наука»  
117864, ГСП-7, Москва, В-485, Профсоюзная ул., 90  
Ордена Трудового Красного Знамени  
Первая типография издательства «Наука»  
199034, Ленинград, В-34, 9 линия, 12

27.12.03

Наименование	Модель	К-во	Цена	Сумма
Корпус	Case 250W 14 USB	1	32	32
Материнская плата	Asus KM400	1	66	66
Процессор	Duron 1400 + Cooler	1	48	48
Оперативная память	DDR 256Mb 2700	1	37	37
Жесткий диск	HDD 40Gb 7200 Samsung	1	64	64
Дисковод	FDD	1	9	9
Привод CD дисков	LG52	1	18	18
Видеокарта	onboard	1	0	0
Сетевая карта	onboard	1	0	0
Звуковая карта	onboard	1	0	0
Колонки	Genius	1	8	8
Клавиатура	Keyboard PS/2	1	7	7
Манипулятор	Genius PS/2	1	0	4
Коврик	Cork	1	1	1
Монитор	Samsung 755DFX	1	153	153
Сборка		1	10	10
<b>Итого компьютер \$</b>				<b>457</b>
		0	0	0
сканер		0	0	56
Filter	Sven	1	5	5
	Win XP, Office, films, music	0	0	0
		0	0	0
		0	0	0
<b>Всего, \$</b>			<b>518</b>	
<b>Всего, гр.</b>				<b>2 797,20</b>