

братского единства, на вполне добровольном согласии» [3]. Именно на этой основе образовался Советский Союз, который сыграл важнейшую роль в формировании и расцвете национальной по форме, социалистической по содержанию культуры всех народов нашей страны. Равноправие языков является краеугольным камнем всей политики КПСС в области национальных взаимоотношений. При этом диалектически связаны развитие национальных языков и распространение русского языка в качестве межнационального языка. Это вполне обоснованная историческая необходимость, без чего затруднились бы сотрудничество и взаимная помощь народов СССР, их взаимный обмен культурными ценностями. Без этого затормозилось бы само развитие народов СССР, дело коммунистического строительства.

Авторы данной книги критически рассматривают различные теории зарубежных ученых (Д. Хаймса, Д. Дж. Гамперца, М. Коула, С. Скрибнера и др.), которые неприемлемы для объективного изучения вопросов развития и взаимодействия языка и культуры.

Среди важнейших методов исследования взаимоотношений между развитием языка и культуры следует выделить метод моделирования, который впервые используется и подробно анализируется во второй главе (с. 128—160). В качестве конкретного примера анализа развития языков и культур союзных республик авторы книги остановились на четырех из них (азербайджанский, литовский, украинский и белорусский), что, конечно, лишь частично отражает многосторонние и сложные аспекты данной проблематики в ее применении к конкретным условиям жизни советских народов. Важнейшие итоги и перспективы взаимоотношений развития языков и культур в

зрелом социалистическом обществе подведены в заключительной, четвертой главе рецензируемой книги (с. 229—318).

Необходимо было бы более четко выяснить вопросы формирования новой социальной и интернациональной общности людей — советского народа и исследовать проблемы развития национальных языков советских народов в их взаимоотношениях с языком межнационального общения, русским языком, т. к. по этому поводу имеется ряд неверных трактовок у зарубежных «советологов». Важно было бы четко сформулировать социальные функции языка межнационального общения и разграничить эти функции от функций «языка-посредника» (*lingua franca*).

В целом надо подчеркнуть большое значение рецензируемой книги для правильной марксистско-ленинской трактовки важнейших вопросов взаимоотношений развития языков и культур в современном, зрелом социалистическом обществе: впервые в советской науке на конкретном материале некоторых национальных языков советских народов и их культур рассматриваются важнейшие вопросы коммунистического строительства.

Корляню Н. Г.

ЛИТЕРАТУРА

1. Маркс К. Капитал. Предисловие к первому изданию. — Маркс К. и Энгельс Ф. Соч. 2-е изд., т. 23, с. 10.
2. Ленин В. И. Нужен ли обязательный государственный язык? — Полн. собр. соч., т. 24, с. 293.
3. Ленин В. И. Письмо к рабочим и крестьянам Украины по поводу побед над Деникиным. — Полн. собр. соч., т. 40, с. 43.

Караулов Ю. Н. Частотный словарь семантических множителей русского языка. — М.: Наука, 1980. 208 с.

Современные искания в области теории и практики лексикографии все чаще выдвигают на передний план понятие лексикографической деятельности. Вопрос «как делать словари?» в равной степени актуален и для традиционных, и для новых типов словарей, ибо он связан с обоснованностью лексикографических процедур. Эта актуальная проблема и составляет внутренний стержень рецензируемой работы Ю. Н. Караулова. Исследование, приведшее к появлению первого частотного словаря семантических множителей русского языка, составленного с помощью ЭВМ, основано на эксперименте, имеющем двоякую направленность: это одновременно и анализ существующих словарей, и этап в создании нового типа словаря. Книга тесно связана с другими трудами Ю. Н. Караулова, посвященными изучению фундаментальных закономерностей лексикографической практики.

В рассматриваемой работе представлен не просто словарь семантических множителей, но и определенный метаязык, надстроенный над первичным метаязыком дефиниций толкового словаря (этот вторичный метаязык включает правила построения единиц с помощью операций над словами из первичных дефиниций, а также правила кодирования дефиниции в целом). Книга дает, следовательно, обобщенные сведения о словах, которые встречаются в дефинициях и — шире — которые играют важную роль в семантической структуре русского языка. Количественные индексы характеризуют активность каждого данного семантического множителя в построении толкований, т. е. показывают, в определении скольких исходных лексем участвует данный семантический множитель (ср. [1]); по этой причине (а также ввиду закреплённости термина «частота» за встречаемостью единиц в речи) можно было бы,

на наш взгляд, называть используемый автором показатель индексом дефиниционной продуктивности. Методика определения таких индексов, как и вся обработка первичных дефиниций, нацелена на использование словаря прежде всего для установления семантических связей между исходными словами в рамках задачи построения тезауруса.

Из сказанного ясно, что «Частотный словарь семантических множителей русского языка» представляет собой вклад в теорию и практику семантических метаязыков, в теорию и практику статистической семантики, в теорию и практику тезаурусостроения, включая сюда теорию семантических расстояний, а также в теорию и практику автоматизации лексикографических работ [ср. 2—5]: рецензируемый словарь пополняет большой класс «машинных словарей» — как общезыковых, так и специальных (особенно многочисленных в области информационно-поисковых систем).

В подготовке материала для машинной обработки приняли участие студенты кафедры структурной и прикладной лингвистики филологического факультета МГУ; «в разработке вопросов машинного анализа метаязыка толкового словаря, подготовке комплекса соответствующих программ для ЭВМ..., в придании частотному словарю его настоящего вида в машинной печати заслуга принадлежит сотрудникам Института социологических исследований АН СССР В. И. Молчанову, В. А. Афанасьеву, Н. В. Михалеву (с. 3).

Книга открывается предисловием «От автора», далее идут вступительная статья «Истоки словаря, его назначение и правила пользования им» и четыре лексико-семантических инвентаря. Во вступительной статье прежде всего разъясняется, о каких множителях идет речь в словаре: это «однозначные слова, использованные в правой части толкового словаря» (с. 6). Эмпирической базой исследования послужили дефиниции 11 000 слов, специально отобранных для построения тезауруса русского языка. Всего выявлено и подвергнуто количественному описанию 8415 семантических множителей. В так называемой феноменологической модели семантическая связь между двумя словами устанавливается, если они имеют хотя бы один общий множитель. При этом принято рабочее решение игнорировать синтаксис дефиниций, а также представлять их лексический состав в обобщенном виде — множители записываются на уровне особых «квазиоснов». Далее, уточненная «асимптотическая модель» связности опирается уже на полученный частотный словарь множителей: информация об их частоте позволяет ранжировать множители в дефиниции по их информативности (весу) и осуществлять сжатие дефиниции за счет отбрасывания второстепенных множителей.

Исходные (толкуемые) слова образуют два множества: каталог дескрипторов, полученный на основе компиляции классификационных схем идеографических словарей разных языков и включающий 1621 единицу; словник, составленный на базе словаря-минимума для иностранцев

(под ред. В. В. Розановой) и расширенный почти вдвое. Для единиц словника дефиниции брались по словарю С. И. Ожегова, а для дескрипторов — также и по словарю Д. Н. Ушакова. Поскольку методика автоматического построения тезауруса предусматривает распределение словника по семантическим полям, возглавляемым дескрипторами, путем выявления общих семантических множителей, то встает задача приведения дефиниций к виду, удобному для формального сравнения. Из возможных путей представления семантической информации выбран тот, при котором «начало слова оставлялось без изменения, а сокращению, усечению подвергалась правая часть» (с. 10), причем «правая граница кода никак не связана с морфемным членением слова» (с. 14). Получаемая таким образом квазиоснова обозначает единицу смысла, соответствующую слову (или гнезду слов, встречающихся в дефинициях: «бол» — *болезнь, болезненный, бо-лет, боль, больно, загад — загадка, загадывать, поздр — поздравить, подр-вление, гада — гадать, гадание, гадалка*. Подобный подход к кодированию семантической информации широко используется в прикладной лингвистике. Ю. Н. Караулов построил алгоритмы кодирования дефиниции, предусматривающих, в частности, следующие процедуры: внесение самого входного (толкуемого) слова в состав его собственных семантических множителей, устранение дублирования множителей в рамках одной и той же дефиниции, а также исключение нулевых множителей — «интуитивно кажущихся высоковероятными единиц» (с. 12), например: *явление, совокупность, место, величина, свойство*. Все эти правила и определяют состав рецензируемого словаря семантических множителей.

Рассматривая вопрос о применении частотного словаря семантических множителей, Ю. Н. Караулов прежде всего подчеркивает специфику частотного анализа метаязыка лексикографии по сравнению с более обычными задачами лингвостатистики, выявляя вместе с тем и определенное сходство между ними: «Совокупность дефиниций оказывается адекватной совокупности текстов, репрезентирующих язык в целом, тогда как одна дефиниция, строго говоря, не равна тексту, поскольку не существует дефиниций, превышающих по своей длине одно предложение» (с. 21). Это противоречивое свойство автор называет «эффектом словаря».

Основное внимание уделяется использованию частотного словаря множителей для разработки асимптотической модели связности как основы машинного составления тезауруса. По результатам пробных экспериментов Ю. Н. Караулов вводит селективный параметр, позволяющий оценивать семантические множители по их роли в распределении слов по семантическим полям. Таким параметром является частота на множестве дефиниций дескрипторов. Весь спектр таких частот эмпирическим путем разделяется на три зоны: 1) частота от 1 до 6 (зона релевант-

ности семантических множителей как показателей связи), 2) частота от 7 до 14 (зона релаксации), 3) частота от 15 до 69 (зона нерелевантности). Эти критерии дают, однако, в ряде случаев результаты, противоречащие интуиции; пути устранения подобных ошибок автор видит в улучшении селективного параметра, в развитии понятия «сила связи».

Указываются еще две области применения созданного словаря: он может служить вспомогательным материалом для лексикографа, способствуя упорядочению процесса составления дефиниций, и может также выступать в качестве «семантического ядра русской лексики», «источника порождения базового (оптимального) словника» (с. 27). Как показал предварительный эксперимент, список множителей дает при его развертывании 30 000 слов.

Из четырех инвентарей, приводимых в книге, два составляют основной корпус словаря, а два носят вспомогательный характер. Первый инвентарь — это список семантических множителей, расположенных по убыванию частоты. В него вошли лишь те множители (всего их 5526), которые встретились в дефинициях дескрипторов. Словарная статья (строка) содержит четыре зоны: порядковый номер (ранг), частота на массиве словника, частота на массиве дескрипторов, буквенная запись множителя. Упорядочение идет по частоте на массиве дескрипторов.

Второй инвентарь — это алфавитно-частотный список множителей. Он включает в се множители, встретившиеся в проанализированных дефинициях, и позволяет для любого из них определить его количественные характеристики.

Два вспомогательных инвентаря представляют собой каталог дескрипторов в алфавитном и систематическом (с разбиением по дескрипторным областям) виде. Читатель имеет возможность судить о степени охвата этими основными понятиями различных классов русской лексики.

Важным достоинством рецензируемой работы Ю. Н. Караулова является ее идейная насыщенность: в концентрированной и вместе с тем наглядной форме в ней дается целый ряд методологических и методических положений, намечающих, по существу, новые пути разработки научных основ лексикографии.

Инвентаризация семантических множителей русского языка осуществляется как органическая часть общей стратегии решения определенной практической задачи — создания тезауруса с помощью ЭВМ. Избранный путь ее решения отличается научной смелостью: естественная гипотеза об определении смысловой близости слов через сравнение их толкований проверяется на огромном противоречивом материале, тающем в себе множество «подводных камней».

Получающиеся результаты зачастую не укладываются в стройную схему, но все наблюдаемые отклонения явным образом представлены в книге и отражают реальные исследовательские поиски взаимосвязей между семантической и статистической структурами. Существенное

значение имеет принцип проведения пробных экспериментов и, в частности, идея сочетания обработки дефиниций с ассоциативным экспериментом. Автор точно указывает те моменты процедуры, которые опираются на интуицию, дополняющую применение формальных методов. Читатель получает наглядное представление о том, как соединяются «ручные» и автоматизированные операции при построении словаря. Отмечаются пути дальнейшей автоматизации отдельных этапов лексикографической деятельности (с. 32, 35 и др.).

Ю. Н. Караулов характеризует семантические множители как элементы «познавательного характера» (с. 4). И предложенный им условный метаязык показывает возможности экспериментального моделирования различных способов представления знаний, зафиксированных в лексике естественного языка [6]. Оригинальность приведенного в книге описания заключается в последовательном изучении эффективности максимально простых средств семантического анализа.

В рецензируемой работе, как и во всяком большом практическом деле, есть, разумеется, ряд дискуссионных моментов. На наш взгляд, слишком категорично подчеркиваются элементарность и универсальность семантических множителей (с. 4—5), тем более, что конкретное исследование автора демонстрирует возможность установления разного порога дробления компонентов и их определенной зависимости от специфики русской лексики (ср. огромное количество приставочных глаголов, используемых в дефинициях).

Нам представляется неоправданной также твердая вера в существование жесткого «селективного параметра» (с. 24—25): вряд ли целесообразно считать, что в лексико-семантической структуре языка имеется некий оптимальный уровень иерархии, на котором и строятся самые правильные семантические поля.

Остается не совсем ясным, как учитывается полисемия исходных слов (и на этапе обработки дефиниций, и на этапе установления связей между ними). Имеются случаи, когда многозначность остается неразрешенной и на уровне семантических множителей. Это сказывается на построении тезауруса, а также затрудняет использование словаря семантических множителей с другими целями. Так, именно благодаря многозначности множителя «налож» устанавливается ложная связь между *арестовать* и *заплата* (с. 23).

На наш взгляд, требует также уточнения алгоритм кодирования семантических множителей: хотя при формировании квазиоснов и сделано существенное укрупнение единиц толкования, все же остались случаи излишнего дробления (например, 15—8 «берег», 1—0 «побереж» и 1—2 «прибреж»). Не вполне обоснованным представляется утверждение о «семантической опустошенности» так называемых «нулевых множителей» (с. 22, 30, 32—33). Их следует, может быть, называть избыточными и более обстоятель-

но подвергать семантическому анализу [ср. 7].

Что касается статистической стороны словаря, то она испытала на себе влияние ряда прикладных решений, связанных с принятой методикой построения тезауруса. На статистику повлияли особая процедура обработки дефиниций дескрипторов (отличная от процедуры обработки прочих слов), решение вносить само исходное слово в состав его собственных семантических множителей, применение отрицательных трансформаций и другие моменты. В целом же количественные данные достаточно точно характеризуют дефиниционный аспект толкового словаря. И еще одно замечание терминологического характера: термин «дескриптор» широко используется в информатике для обозначения класса ключевых слов, приравниваемых друг к другу по смыслу в рамках определенной информационной задачи. Поскольку такое понимание явно отличается от принятого в концепции Ю. Н. Караулова, то целесообразно было бы употреблять в книге другой термин.

Расценивая данную работу как фундаментальный результат в лексикографии, считаю необходимым наметить некоторые пути ее опосредованного и комплексного использования.

На основе опубликованных данных мы можем получать новые списки семантических множителей, упорядочивая их по другим количественным показателям. Во-первых, здесь надо подчеркнуть, что, помимо частоты на массиве дескрипторов, существенный интерес представляет частота на массиве словника (ср.: 2—5 «гибк», 19—5 «наблю», 28—5 «изда», 35—5 «учеб»). Во-вторых, особенно важна, на наш взгляд, общая дефиниционная активность (совокупно по дескрипторам и словнику). Произведенный нами подсчет показывает, что по этому показателю наиболее активные множители будут упорядочены следующим образом: 183—52 «дви» (235), 180—47 «часть» (227), 181—41 «врем» (222), 160—60 «друг» (220), 180—38 «мест» (218), 161—47 «действ» (208), 112—67 «раст» (179), 107—66 «животн» (173), 129—43 «обмещ» (172), 102—69 «зем» (171), 136—33 «направ» (169), 137—21 «работ» (158), 101—51 «жи» (152), 91—59 «чел» (150), 91—53 «поверхн» (144).

Рецензируемый словарь позволяет получать, далее, количественные характеристики для разнообразных укрупненных семантических множителей с последующим изучением их статистического распределения. Это прежде всего обобщение данных, касающихся словообразовательных коррелятов, морфем и их сочетаний. Возможно укрупнение данных и для собственно семантических группировок множителей. Наш подсчет показывает, что условный синонимический ряд *желать, хотеть, стремиться, мечтать, жаждать, возжелеть, желание, воля, стремление, желанный, долгожданный* и т. д. имеет общую частоту 152 (и по дескрипторам, и по словнику). Семантическое поле основных цветообозначений характеризуется показателем 176, причем его члены упорядочиваются по своей дефиниционной активности следующим об-

разом: «бел», «красн», «сер», «черя», «зел», «желт», «син», «голуб», а оранжевый, коричневый и фиолетовый цвета в дефинициях вообще не встретились. Изучение материала в указанном направлении позволит описать семантико-квантитативную структуру словаря множителей.

Работа Ю. Н. Караулова составляет серьезную основу для сопоставления в будущем метаязыков разных толковых словарей, что важно и само по себе, и в плане определения того, насколько модель семантической связности словарного состава зависит от используемого толкового словаря.

Словари семантических множителей — это необходимая эмпирическая база для поисков оптимальных решений при построении семантических сетей и формул в задачах «искусственного интеллекта». Без подобных словарей невозможно исследовать и проблему универсальности семантических мнемонетов.

В заключение отметим, что применение методики Ю. Н. Караулова при проведении подобных работ в будущем допускает совершенствование в нескольких направлениях. Во-первых, в плане отбора исходной лексики (с учетом различных общезыковых словарей-минимумов, а также лексики функционально-стилистических сфер общения и специальных подязыков). Во-вторых, по линии совершенствования предварительной обработки метаязыков дефиниций (здесь полезно учесть например, такие факты: часто отсутствие каких-либо слов в дефиниции обусловлено внешними обстоятельствами или просто случайностью; чем больше толковый словарь по объему, тем активнее в его дефинициях используются синонимы; появление в дефиниции того или иного члена словообразовательного гнезда часто зависит от морфологических, синтаксических или стилистических причин). В-третьих, полезно исследовать иные пути установления связей между исходными словами (скажем, можно предварительно не фиксировать дескрипторы, попытаться учесть совместную встречаемость множителей в дефиниции, использовать методы измерения смысловой близости, сложившиеся в информатике, психолингвистике, практике «контент-анализа»).

Авторская концепция построения тезауруса нашла практическое воплощение в работе 1982 года («Русский семантический словарь. Опыт автоматического построения тезауруса: от понятия к слову»), которая была подвергнута критическому анализу на совместном заседании Советов Института языкознания и Института русского языка АН СССР 20 сентября 1983 г. Большинство участников дискуссии отметило принципиальную перспективность данного направления в автоматизации лексикографических исследований. Были предложены конкретные пути их дальнейшего развития. Недостатки «Русского семантического словаря» в его вышедшем виде послужили предметом критического отклика, опубликованного в «Правде» [8]. Многие из этих недостатков изданного тезауруса, прежде всего неточности в определении смысловых

связей слов, были выявлены и в ходе упомянутого выше обсуждения.

Резюмируя наш разбор, можно констатировать, что рецензируемая работа представляет собой определенный шаг в многотрудном становлении экспериментального подхода к семантике и лексикографии. Ее конкретный результат — это множество семантических элементов, достаточных для описания десятков тысяч русских слов. Это существенный вклад в создаваемый машинный фонд русского языка.

Городецкий Б. Ю.

ЛИТЕРАТУРА

1. West M. Definition vocabulary. Toronto, 1935.
2. Kiss G., Armstrong C., Milroy R., Piper J. An associative thesaurus of

English and its computer analysis.— In: The computer and literary studies. Edinburgh, 1973.

3. Osgood C. E., May W. H., Miron M. S. Cross-cultural universals of affective meaning. Urbana — Chicago — London, 1975.
4. Марчук Ю. Н. Вычислительная лексикография. М., 1976.
5. Андрущенко В. М. Автоматизированные лексикографические системы.— В кн.: Теоретические и прикладные аспекты вычислительной лингвистики. М., 1981.
6. Коршунов А. М. Отражение, деятельность, познание. М., 1979, с. 188—198.
7. Laffal J. A concept dictionary of English. Essex (Connecticut), 1973.
8. Дерягин В. Учить ли ЭВМ писать с ошибками.— Правда, 1983, 27 сент.

Clauson G. An etymological dictionary of pre-thirteenth-century Turkish. Index. Vol. I — Szeged, 1981, XI + 342 p.; Vol. II — Szeged, 1981, 261 p. (Studia Uralo-Altaica, N 14—15)

Вышедший в 1972 г. в Оксфорде «Этимологический словарь тюркского языка до XIII в.» известного английского тюрколога Дж. Клосона (1891—1974) [1] получил высокую оценку со стороны тюркологов (см., например, [2]). Однако этот словарь с довольно сильным креном в сторону истории слов имеет весьма своеобразное расположение лексического материала — почти исключительно по одним согласным, к тому же без их противопоставления по глухости—звонкости и лишь при достаточно схематическом учете гласных, что навевает характерными особенностями старинных тюркских письменностей семитского и индоиранского происхождения, где гласные в составе слова часто обозначались непоследовательно и обычно неоднозначно. Это обстоятельство, хорошо известное специалистам по истории тюркской письменности, обычно затрудняет пользование словарем для нетюркологов, которые недостаточно хорошо знают своеобразие этих старых тюркских письменных систем. Были высказаны настоятельные пожелания о необходимости составления указателя к этому весьма информативному этимологическому и историческому словарю. Два тома строго алфавитных указателей составлены по инициативе известного венгерского тюрколога А. Рона-Таша в Сегедском университете имени Агаты Йожефа с помощью вычислительных машин.

В первом томе содержится один строго алфавитный общий список из 10 842 словарных единиц, представленных в качестве заголовочных и их вариантов в словаре Дж. Клосона (в том числе 203 ошибочно выделенных издателями и исследователями тюркских текстов, 792 вторичных единицы в 2069 смысловых, а остальные 7778 единиц представляют собой независимые словарные единицы). Среди них 3664 глагола. В общем алфавитном указателе-списке, кроме отсылки к соответствующим страницам словаря,

дается итоговая лапидарная индексная этимологическая справка о каждом слове: особыми однобуквенными латинскими индексами указывается его производный (D — derived), сложный (C — compound), заимствованный (F — foreign), вторичный (S — secondary) или даже ошибочный (E — erroneous) характер. Гипотезы, которых много в словаре Дж. Клосона и которые всегда вызывают подозрения в их достоверности, в отдельную группу не выделены. Для слов с неясным чтением (по консонантизму и вокализму) отмечается их филологическая ненадежность (U — uncertain). В случае неясности этих характеристик слово по соответствующим параметрам получает индекс X. Исконные корневые тюркские слова, как и в самом словаре Дж. Клосона, никакими индексами не снабжаются. Следует учесть, что в указателе включены лишь заголовочные слова, извлеченные Дж. Клосоном из письменных памятников до XIII в. Параллели из более поздних памятников и из современных тюркских языков, показывающие распространение этимологизируемых слов в тюркском языковом мире, в указателе не представлены. Это обстоятельство, конечно, затрудняет использование словаря для наведения этимологических справок по словам современных тюркских языков и требует создания указателя тюркских слов из современных языков, которые фигурируют в словаре Дж. Клосона. Сомнительное в каком-либо отношении вариантное чтение отдельных слов из письменных памятников заключается в скобки с отсылкой к более вероятному чтению, которое дается без скобок средствами современного турецкого алфавита на основе латиницы. Фактически количество этимологических словарных статей в словаре Дж. Клосона гораздо меньше, чем можно судить на основании рецензируемого указателя, поскольку самостоятельную нумерацию получили чисто отсылочные слова, а также варианты прочтения, ко-