

**Всесоюзный
Центр
Переводов**

И. И. УБИН

**АВТОМАТИЧЕСКИЙ ПЕРЕВОДНОЙ СЛОВАРЬ.
ПРИНЦИПЫ ПОСТРОЕНИЯ**

(Методическое пособие)



МОСКВА • 1989

Государственный комитет СССР
по науке и технике

Академия наук
СССР

ВСЕСОЮЗНЫЙ ЦЕНТР ПЕРЕВОДОВ
НАУЧНО-ТЕХНИЧЕСКОЙ ЛИТЕРАТУРЫ И ДОКУМЕНТАЦИИ

И.И. УБИН

АВТОМАТИЧЕСКИЙ ПЕРЕВОДНОЙ СЛОВАРЬ.

ПРИНЦИПЫ ПОСТРОЕНИЯ

/Лингвистические концепции/

/Методическое пособие/

Москва 1989

УДК 801:323,9:681:3:06

Ответственный редактор
к.т.н. Иванов В.А.

Рецензент
д.ф.н. Шайкевич А.Я.

ВВЕДЕНИЕ

Машинный перевод переживает в настоящее время бурный период в своей уже почти полувековой истории. Если всего 5-8 лет назад работы по машинному переводу были сконцентрированы в основном в университетах и в немногих специализированных организациях /Гренобль /Франция/, Саарбрюкен /ФРГ/, Монреаль /Канада/, Техас /США/, Киото /Япония/, проект Евротра /Европейское экономическое сообщество/, Москва, Ленинград /СССР/, то сейчас проблемами машинного перевода в самом широком понимании этого термина занимаются в десятках различных организаций во многих странах. Повышенный интерес к использованию ЭВМ и в первую очередь персональных компьютеров к проблемам перевода проявляют частные коммерческие фирмы в Японии, США и Европе. Общее количество специалистов, работающих в области МП /без учета социалистических стран/, достигает сегодня примерно 1500-2000 человек, из них только в Японии - 800-900 человек /Hutchins, 1988/.

Расширился также перечень типов разрабатываемых систем для автоматизации перевода. Если на начальном этапе развития МП /примерно до середины 60-х гг./ внимание уделялось только созданию "пакетных" систем МП, рассчитанных либо на полностью автоматический режим работы, либо на полуавтоматический режим с использованием человека на этапах пред- и постредактирования, то затем появились крупные автоматические многоязычные словари и терминологические банки /середина 60-х, 70-е гг./. В настоящее время все большее внимание уделяется созданию интерактивных систем МП, разнообразных автоматических словарей и автоматизированных рабочих мест /АРМ/ переводчика.

Принципиальное различие между системами МП, с одной стороны, и разнообразными автоматическими переводными словарями /АПС/ и АРМ переводчика - с другой, заключается в том, что системы МП предназначены для полного перевода всего текста и в конечном счете нацелены на замену человека-переводчика, в то время как АПС и АРМ переводчика предназначены для поиска переводных эквивалентов отдельных слов и словосочетаний и нацелены на диалог с человеком в процессе перевода или чтения иноязычного текста. АПС и АРМ

переводчика, который можно рассматривать как специфический тип АПС, занимают особое место среди средств автоматизации перевода и, на наш взгляд, дают право говорить о появлении нового направления в лексикографии – переводной вычислительной лексикографии. Специфика АПС заключается в промежуточном положении, которое они занимают между переводными словарями в традиционной книжной форме, с одной стороны, и системами МП – с другой. В силу специфичности своих задач и целевой направленности АПС обладают рядом характеристик, в целом не свойственных системам МП и традиционным словарям. К таким свойствам следует в первую очередь отнести **многоязычие, обратимость, гибкость, динамичность и состав.** Анализ этих специфических свойств АПС посвящена более ранняя работа автора /см. Убин, 1986/.

Оптимальная конфигурация АПС в значительной степени зависит от того, насколько разработчики смогли учесть особенности производственной сферы его применения, а именно: лингвистические характеристики обрабатываемых текстов, направление перевода, свойства основных категорий пользователей* и их требования к АПС и др. В настоящей работе на примере АПС Всесоюзного центра переводов рассматриваются вопросы поиска оптимальных путей построения АПС и предлагаются некоторые типовые решения основных лингвистических задач, возникающих при создании автоматических переводных словарей.

*Следуя сложившимся лексикографическим традициям, мы будем называть человека, работающего с традиционным словарем, читателем словаря, а с АПС – пользователем словаря.

Г л а в а I

ОПРЕДЕЛЕНИЕ ОСНОВНЫХ ЭКСПЛУАТАЦИОННЫХ ХАРАКТЕРИСТИК АВТОМАТИЧЕСКОГО ПЕРЕВОДНОГО СЛОВАРЯ НА ОСНОВЕ АНАЛИЗА ПРОИЗВОДСТВЕННОЙ СФЕРЫ НАУЧНО-ТЕХНИЧЕСКОГО ПЕРЕВОДА

1.1. Место автоматического переводного словаря в процессе перевода

Под производственной сферой научно-технического перевода мы понимаем ситуацию, возникающую при выполнении перевода, а именно: условия, в которых он выполняется; свойства исходного материала и требования к конечному продукту /в наших условиях это текст на языке оригинала и его перевод на другом языке/, основные характеристики исполнителей перевода. Анализ особенностей производственной сферы научно-технического перевода в пределах одной конкретной организации или класса однотипных организаций позволит определить основные эксплуатационные характеристики, которыми должна обладать автоматизированная система /в нашем случае автоматический переводной словарь/, проектируемая для данной производственной сферы.

Для того чтобы определить место АПС в процессе перевода, необходимо первоначально решить вопрос о разделении всего процесса перевода на отдельные последовательные этапы. Следует отметить, что при решении этого вопроса в теории МП и в теории традиционного перевода существуют заметные различия. В практике машинного перевода, начиная с создания самых первых систем МП, естественным путем выделялись этапы вычленения лексических единиц в тексте, морфологического и синтаксического анализа, поиска по словарю входного языка для определения переводных эквивалентов, синтаксического и морфологического синтеза и другие более мелкие и специфические этапы процесса МП. Все эти стадии в различных теоретических моделях МП сводятся к трем основным этапам: этапу анализа, промежуточному этапу, основной задачей которого является выбор переводных эквивалентов для лексических единиц входного языка, и этапу синтеза. В промежуточный этап включают поиск по словарю, разрешение неоднозначности лексическими средствами, обработку устойчивых и фразеологических словосоче-

таний, установление лексических и синтаксических соответствий между входным и выходным языками. Этот этап получил в ряде работ по МП название "трансфера" /Марчук, 1985, с. 44/.

В теоретических работах по традиционному переводу в этом вопросе нет единодушия, характерного для теоретических исследований по МП. Некоторые теории перевода в своих построениях не делают в явном виде процесс перевода на отдельные этапы. Так, А.Д.Швейцер пишет, что перевод, как и любая речевая деятельность, не осуществляется по единой модели. В процессе перевода находят применение грамматические трансформации, лексико-грамматические перифразы и семантические преобразования. При этом выбор оптимального способа анализа исходного текста и построение соответствующего текста на другом языке диктуются конкретными условиями коммуникативного акта /Швейцер, 1973, с. 61/.

В теории закономерных соответствий, разработанной Я.И.Рецкером, основное внимание уделяется сравнительному изучению языковых явлений и установлению определенных соответствий между языком оригинала и языком перевода. Соответствия в области лексики, фразеологии, синтаксиса и стиля, по мнению Я.И.Рецкера, составляют лингвистическую основу перевода. В этой теории и в теории уровней эквивалентности не развивается тезис о раздельности этапов анализа и синтеза в процессе перевода /Рецкер, 1974; Комиссаров, 1973/.

В то же время другие исследователи перевода вполне однозначно выделяют в процессе перевода последовательные этапы. Так, чешский ученый Иржи Левый в своих работах по теории художественного перевода весьма четко выделяет три фазы переводческого труда: 1/ постижение подлинника, 2/ интерпретация подлинника, 3/ перевыражение подлинника /Левый, 1974, с. 59/.

Немецкий исследователь О.Каде рассматривает перевод как один из видов двуязычной коммуникации, который начинается с восприятия текста на языке оригинала и заканчивается реализацией текста на языке перевода. Важнейшей фазой этого процесса является замена кода языка оригинала на код языка перевода. Для успешного выполнения перевода переводчик должен решить следующие три задачи:

1/ декодирование текста на языке оригинала с целью перекодирования;

2/ перекодирование, т.е. подстановка знаков языка перевода вместо знаков языка оригинала;

3/ реализация текста на языке перевода / Каде, 1978, с. 73/.

В своей модели синхронного перевода А.Ф.Ширяев также выделяет в процессе перевода три взаимосвязанных этапа: процесс ориентирования в исходном тексте, процесс поиска и принятия переводческих решений и процесс осуществления переводческих действий / Ширяев, 1979, с. 100/.

Таким образом, анализируя разные взгляды на процесс перевода, мы пришли к убеждению, что процесс перевода человеком подразделяется на несколько взаимосвязанных последовательных этапов. В самом общем виде можно выделить следующие три этапа: 1/ анализ текста оригинала; 2/ замена кода языка оригинала на код языка перевода; 3/ синтез текста перевода. Понимая все различия между проблемами теории традиционного перевода и проблемами, которые стоят перед теорией машинного перевода, все же можно провести параллель между тремя фазами традиционного перевода и тремя этапами работы системы МП /анализ, трансфер, синтез/. Такое разделение процесса перевода достаточно четко определяет в нем место переводного словаря, а именно этап замены кода языка оригинала на код языка перевода. Структура и состав переводных словарей в традиционном книжном оформлении подтверждают этот вывод.

Основная задача традиционного переводного словаря - дать соответствия типа "лексическая единица языка оригинала - лексическая единица /или единицы/ языка перевода". Переводный словарь косвенным образом помогает решать некоторые задачи этапов анализа текста оригинала за счет включения в предисловие грамматического описания входного языка, а также за счет включения в словник словаря в качестве отсылочных словарных статей некоторых словоформ входного языка, образуемых нерегулярными способами. Составители некоторых переводных словарей включают в них грамматический очерк выходного языка. В таких случаях можно говорить, что эти словари до некоторой степени помогают также решать задачу синтеза текста.

Тем не менее решение проблем анализа входного и синтеза выходного текстов не являются основными задачами традиционных переводных словарей. Составители традиционных терминологических словарей задач такого типа, как правило, перед собой даже не ставят.



Рис. 1. Место словаря в процессе перевода

Условные обозначения:

- ~~~~ место традиционного переводного терминологического словаря;
- — место традиционного переводного словаря общей лексики;
- — место автоматического переводного словаря

В настоящее время основной задачей автоматических переводных словарей, как и словарей в традиционной книжной форме, остается установление соответствий "лексическая единица языка 1 - лексическая единица /единицы/ языка 2". Однако сейчас многие АПС в большей степени, чем традиционные словари, решают с помощью различных лингвистических алгоритмов некоторые из задач этапов анализа входного текста, такие как сведение словоформ к словарным формам слова, вычленение в тексте лексических единиц, словообразовательный анализ, что и нашло отражение на приведенном выше рисунке.

1.2. Краткая характеристика Всесоюзного центра переводов

1.2.1. Анализ потока переводов по языкам и тематикам

АПС, который описывается в настоящей работе, предназначен в первую очередь для специализирован-

ных переводческих и информационных организаций и подразделений, поэтому целесообразно предварительно рассмотреть основные аспекты переводческой деятельности, а именно: основные языки, с которых и на которые осуществляется научно-технический перевод, профилирующие тематики текстов и типы источников, некоторые организационные вопросы /выполнение перевода штатным или штатным переводчиком, уровень его языковой и инженерно-технической компетенции/. Ответы на эти и другие вопросы, связанные с традиционной переводческой деятельностью, в значительной степени влияют на тип АПС и определяют ряд лингвистических решений в нем.

Всесоюзный центр переводов научно-технической литературы и документации является одной из ведущих переводческих организаций в стране. Объем научно-технических переводов, выполняемых ВЦП, постоянно растет. Так, если в 1974 г. ВЦП выполнил 20 тыс. а.л. переводов, то в 1988 г. свыше 80 тыс. а.л. ВЦП выполняет переводы по заказам государственных учреждений всех министерств и ведомств. Специфика и разнообразие задач, стоящих перед этими организациями, определяет состав языков и тематику переводов. Переводы ВЦП составляют примерно 70% всего годового поступления в государственный фонд научно-технических переводов. На основании этих данных можно, по нашему мнению, рассматривать переводческую ситуацию в ВЦП как модель переводческой обстановки в целом и предположить, что тенденции, проявляющиеся в ВЦП, свойственны в значительной степени всей системе научно-технического перевода в стране. Мы полагаем, что приведенные выше данные дают основание рассматривать ВЦП как базовую организацию для анализа производственной сферы научно-технического перевода*.

* Всесоюзный центр переводов выполняет переводы только научно-технической литературы и документации, поэтому в дальнейшем, говоря о ВЦП или системе научно-технической информации страны, автор будет в целях удобства использовать термин "перевод" вместо "научно-технический перевод".

В 1985 г. ВЦП выполнил переводы с 35 иностранных языков на русский и с русского на 31 иностранный язык. Объемы переводов, выполняемых с того или иного иностранного языка, весьма различны, что отражает место этого языка в мировой системе научно-технической информации. Так, на долю 10 ведущих языков приходится примерно 95% всего объема переводов, а на долю остальных 25 языков приходится только 5%. Даже среди десяти основных языков резко выделяются английский /62,5%/, немецкий /15,4%/, японский /6,4%/, и французский /5,8%/, языки /Смирнов, 1987, с. 27/.

В 1987 г. объем переводов с русского на иностранные языки составил 7,2% от общего объема переводов. Основная часть этой категории переводов приходится на английский /5,3%/, испанский /0,7%/, немецкий /0,3%/, и французский /0,4%/, языки.

Разделение переводов на переводы с иностранных языков и на иностранные обуславливает различие в целях перевода. По мнению американского исследователя Дж.Слокума, перевод по своим целям подразделяется на два основных направления: перевод для сбора информации и перевод для распространения информации /Слокум, 1987, с. 87/. Перевод с иностранных языков на родной в своей подавляющей части предназначен для сбора информации. Основными потребителями таких переводов являются ученые и специалисты, которые нуждаются в зарубежной информации для поиска решений или для проверки уже принятых решений стоящих перед ними проблем. Основопологающим фактором здесь является оперативность выполнения перевода. Перевод на иностранные языки нацелен в основном на распространение информации. Здесь на первое место выходит качество перевода.

Ценность всякого словаря, в том числе и автоматического, в значительной степени зависит от того, насколько точно решены в нем проблемы целевой направленности, т.е. насколько точно он отвечает потребностям основных категорий своих читателей или пользователей и до какой степени помогает решать стоящие перед ними задачи. Тематика переводимых текстов, вид источника, уровень языковой и профессиональной подготовки читателя /пользователя/, условия работы обязательно должны учитываться при определении макро- и микроструктуры словаря.

1.2.2. Типы материалов, поступающих на перевод

Тип документа, с которым работает пользователь АПС, также оказывает свое влияние на характеристики этого словаря. В самом обобщенном виде можно выделить следующие основные типы материалов и документов, с которыми чаще всего приходится иметь дело переводчикам и техническим специалистам:

текстовые материалы, к которым относятся книги, статьи из журналов и научных сборников, диссертации, доклады, отчеты и т.п.;

техническая документация /описание устройства оборудования, условий ее монтажа, эксплуатации и ремонта, программная документация, различные фирменные материалы и т.п./;

импортно-экспортная документация /перечни комплектующих деталей оборудования, накладные, каталоги запасных частей, списки оборудования, узлов, отдельных деталей для продажи или приобретения и т.п./.

Всесоюзному центру переводов приходится в основном иметь дело с первыми двумя категориями материалов и документов. Текстовые материалы первого типа мы относим к открытым по лексическому составу текстам. В таких текстах могут встречаться слова всех грамматических классов во всех теоретических возможных словоформах или в большинстве из них. Импортно-экспортную документацию мы относим к закрытым с точки зрения лексики текстам. Лексика таких текстов весьма существенно ограничена как по составу, например одной тематикой, так и по форме представления, например только словарной формой. Техническая документация занимает промежуточное положение между этими двумя типами текстовых материалов. Открытые тексты требуют от разработчика АПС включения в него лексики нескольких смежных тематик, общеупотребительной лексики, решения проблемы лемматизации в полном объеме, т.е. для всех теоретически возможных словоформ. С другой стороны, закрытые тексты предъявляют к АПС меньше требований как с точки зрения широты его лексического состава, так и с точки зрения полноты решения проблемы лемматизации. Это означает, что на практике можно добиться высокой степени покрытия словарем закрытого текста, включив в АПС лексику только одной или двух тематик, проблему лемматизации в самом крайнем случае можно даже снять и огра-

ничить запрос только словарной формой слова.

Анализ потока переводов, который проводится в ВЦП регулярно каждые три-пять лет, показал, что основным типом источников переводов в ВЦП стабильно являются периодические издания /журналы, сборники статей по отдельным отраслям науки и техники/. Так, журнальные статьи составляют 96% всей переводимой с японского языка научно-технической литературы. По другим языкам перевод статьи из журнала составляет соответственно: с английского языка - 57%, с немецкого языка - 64%, с французского языка - 73%, с венгерского языка - 84%, с польского языка - 72%, с чешского, словацкого и болгарского языков - 78%.
/Анализ и обобщение, 1986/.

Резюмируя этот краткий раздел об основных характеристиках ВЦП как переводческой организации, следует подчеркнуть, что в ВЦП работают штатные и нештатные сотрудники, перевод осуществляется с нескольких десятков иностранных языков на русский язык и с русского на эти иностранные языки, ведущими среди которых являются английский, немецкий, французский и японский языки. Основными тематиками переводов являются вычислительная техника, машиностроение, робототехника, электроника, металлургия и химия. Ведущим источником текстов для перевода является статья из периодического издания /Смирнов, 1987/. Все эти данные должны быть учтены при разработке АПС ВЦП.

1.3. Необходимость автоматизации перевода в специализированной переводческой организации

1.3.1. Необходимость создания автоматического словаря для переводческой организации

Сложившаяся на современном этапе в мире обстановка в области перевода научно-технической литературы и документации, стремительное ускорение развития практически всех отраслей науки и техники настоятельно требуют поиска путей автоматизации научно-технического перевода. ВЦП в этом отношении не является исключением. Специфика переводческого дела, так как она отражается в ВЦП, дает нам право сделать вывод, что решения проблемы ускорения на-

учно-технического перевода следует искать на пути оптимального сочетания, с одной стороны, систем МП, а с другой – автоматических переводных словарей разных типов.

В современных условиях при огромных объемах иноязычной информации, поступающей на обработку, не хватает ни средств, ни возможностей на тщательный перевод каждого документа традиционными средствами. Число научных публикаций настолько возросло, что не хватает времени на ознакомление со всеми иноязычными текстами. Усилия, затрачиваемые на тщательный перевод всей необходимой литературы, весьма часто не оправдывают себя, поскольку вероятность встретить действительно необходимую информацию невелика. Цели сбора информации можно достичь с помощью оперативного, пусть даже не очень высокого по качеству перевода по интересующей и хорошо знакомой потребителю тематике /Слокум, 1987, с. 87/. Переводы такого типа начинают все шире использоваться в системе научно-технической информации за рубежом и у нас в стране и получили название информационных или сигнальных переводов. Под сигнальным переводом понимается перевод, выполненный самыми оперативными средствами и дающий пользователю представление о содержании текста оригинала. Сигнальный перевод в принципе может иметь серьезные стилистические и даже отдельные смысловые ошибки, он предназначен в основном для принятия решения о ценности и новизне, с точки зрения конкретного пользователя, информации, содержащейся в обрабатываемом документе. В большинстве случаев потребитель может понять с помощью сигнального перевода содержание исходного документа. В случае необходимости сигнальный перевод подвергается необходимому редактированию, при котором устраняются в первую очередь все смысловые, а также основные стилистические ошибки. Отдельные фрагменты текста оригинала могут подвергаться полному и тщательному переводу.

Сигнальные переводы могут выполняться как вручную, так и с помощью автоматизированных средств. Наиболее эффективным средством получения сигнальных переводов, по нашему мнению, являются системы МП. Наиболее ярко преимущество систем МП проявляется при переводе текстов большого объема, предваритель-

но записанных на машиночитаемый носитель. Таких текстов среди переводов, выполняемых в стране, достаточно много, и в первую очередь среди различного вида документации. Следует отметить, что эти же самые тексты, как правило, монотематичны и имеют существенные лексические и синтаксические ограничения, что облегчает настройку на них систем МП /Тихомиров, 1987/.

Однако основной поток заказов на перевод, поступающий в ВЦП в настоящее время, составляют тексты на традиционном бумажном носителе /статьи из журналов, главы из книг, диссертаций и т.п./ средним объемом около 1 а.л., и выполняются они силами штатных переводчиков. Перевод таких текстов системой МП с последующим редактированием менее оптимален, чем перевод человеком-переводчиком, работающим в режиме диалога с АПС, либо с помощью текстоориентированного глоссария, полученного на базе этого АПС. Это объясняется в первую очередь тем, что для перевода таких текстов с помощью системы МП их необходимо предварительно перенести на машиночитаемый носитель. Малые объемы текстов этой категории не дают возможности системам МП реализовать свое основное преимущество - высокую скорость выполнения перевода.

АПС в принципе может использоваться при работе со всеми тремя основными категориями текстов, поступающих на перевод в ВЦП, - текстовыми материалами, технической документацией и импортно-экспортной документацией. По нашему мнению, АПС в условиях типовой переводческой организации будет иметь преимущество перед системой МП при переводе открытых текстов, однако он мало пригоден для оперативного получения сигнальных переводов текстов с магнитной ленты. При переводе импортно-экспортной документации с одинаковым успехом могут использоваться АПС и системы МП с последующим постредактированием текста перевода. Таким образом, мы полагаем, что в ВЦП существуют объективная необходимость создания и экономическая база для эксплуатации автоматического переводного словаря.

1.3.2. Модели основных категорий пользователей автоматического переводного словаря

Без учета требований, уровня языковой и специальной профессиональной подготовки потенциальных пользователей словаря разработчики АПС не смогут решить по-

ставленных перед собой задач. Поэтому для этой цели в первую очередь необходимо определить круг пользователей словаря и хотя бы в самом общем виде построить модели основных категорий пользователей АПС.

Для традиционных терминологических словарей выделяют следующие основные категории читателей: специалисты с высшим образованием, переводчики художественной и научной литературы, сотрудники НИИ, студенты филологических вузов, специалисты среднего звена в промышленности и сельском хозяйстве, работники радио и печати /Герд, 1986, с. 9/.

В свою очередь создатели автоматических словарей и терминологических банков выделяют следующие основные категории пользователей: переводчики, редакторы, терминологи, стандартизаторы терминологии, научные работники, технические специалисты, издатели, разработчики автоматизированных систем переработки текста /Sager, McNaught, 1980, p. 6; Felber, 1983; Galinski, 1988; Туовинен, 1985, с. 219; Скороходько, Стогний, 1986, с. 29/.

АПС ВЦП создается в переводческой организации и предназначен в основном для автоматизации перевода, поэтому он направлен в первую очередь на удовлетворение потребностей переводчика и редактора /штатного и нештатного/ научно-технической литературы. Помимо переводчика и редактора в качестве потенциального пользователя нашего словаря мы также выделяем технического специалиста, который довольно часто самостоятельно работает с иноязычной литературой в целях сбора информации. АПС ВЦП может также представлять интерес для других категорий пользователей /терминологов, стандартизаторов терминологии, издателей/.

Переводчики, редакторы и технические специалисты различаются между собой по уровню языковой и специальной профессиональной подготовки, по своим целям и задачам при работе с научно-технической литературой, поэтому, по нашему мнению, их целесообразно разделить на три отдельные категории пользователей.

1. Переводчики. Основная задача переводчика научно-технической литературы заключается в выполнении перевода текста с целью полной, точной и адекватной передачи содержания текста на языке оригинала средствами языка перевода. Следует различать переводчика с иностранного языка на родной /в нашем случае на русский/ и с родного на иностранный язык.

Практика ВЦП показала, что "среднестатистический переводчик" с иностранного языка на русский является по образованию техническим специалистом и уступает по уровню знания иностранного языка "среднестатистическому переводчику" с русского языка на иностранный, который, как правило, является специалистом с гуманитарным образованием, но превосходит последнего по уровню компетентности в сфере, по которой осуществляется перевод. Различия в уровне технической и языковой подготовки этих категорий переводчиков находят свое отражение в том, с какими запросами они обращаются к словарям. Основные трудности для переводчиков обеих категорий представляют новые, еще не получившие широкого распространения и не зафиксированные в словарях или других доступных лексикографических источниках термины, терминологические словосочетания, фирменные названия новой продукции /в самом широком понимании этого слова/ и сокращения. Переводчик с иностранного языка на русский дополнительно выделяет многозначную общеупотребительную лексику, а переводчика с русского языка на иностранный эта лексика интересует в меньшей степени. По уровню своей подготовки и требованиям к словарям переводчики с иностранного языка на родной стоят ближе к техническим специалистам, а переводчики на иностранный язык - к редакторам. Переводчики обеих категорий обладают высоким уровнем знания иностранного языка и навыками работы с переводными словарями. В переводном словаре переводчик ищет в первую очередь переводные эквиваленты для незнакомой ему лексики. При работе с новой терминологией его интересуют также толкования, а при работе с сокращениями - развернутая форма сокращенных лексических единиц, а также их полные и сокращенные иноязычные эквиваленты /Shaikevich, Oubine, 1988/. Контексты, источники и дата поступления терминологии в словарь переводчиков практически не интересуют.

2. Редакторы. Следует отметить, что разработчики АПС и организации, эксплуатирующие крупные автоматические словари, не выделяют редакторов научно-технического перевода в качестве отдельной категории пользователей своих словарей. Однако, учитывая специфику ВЦП, мы полагаем целесообразным выделить редакторов в отдельную категорию пользователей АПС.

Основная задача редактора ВЦП - определение качества перевода /его адекватности, точности и полноты/. "Среднестатистический редактор" ВЦП обладает

глубоким знанием иностранного языка и специальности, переводы по которой он контролирует, опытом переводческой и редакторской работы.

По своим требованиям к составу словаря и содержанию словарной статьи редактор не отличается от переводчика высокой квалификации. Некоторые дополнительные доводы в пользу выделения редакторов в отдельную от переводчиков категорию пользователей АПС дают различия в условиях их работы. Редакторы в основном являются штатными сотрудниками ВЦП и должны иметь возможность оперативного обращения к словарю, в то время как все переводчики ВЦП являются нештатными сотрудниками, работают вне стен ВЦП и такой возможности практически лишены. Вместе с тем следует сказать, что выделение редакторов в отдельную категорию пользователей основывается в настоящее время в основном на теоретических рассуждениях, и только после накопления необходимых статистических данных по промышленной эксплуатации АПС можно будет окончательно определить правомерность этого решения.

3. Технические специалисты. Основная задача технического специалиста при работе с иноязычной научно-технической литературой – это оперативное получение информации из первоисточника. В большинстве случаев технический специалист ограничивается чтением научно-технического текста, в некоторых случаях прибегает к выборочному переводу. При выполнении полного перевода технический специалист переходит в категорию переводчика. Отличительной особенностью "среднестатистического технического специалиста" является более высокий по сравнению с переводчиком уровень специальной профессиональной подготовки и одновременно с этим, как правило, более слабое знание иностранного языка и приемов перевода. Отсутствие навыков работы с переводными словарями и недостаточное знание иностранного языка затрудняют, а для языков с развитой системой словоизменения и большим количеством нерегулярно образованных словоформ делают практически невозможным самостоятельную работу технического специалиста с иноязычным научно-техническим текстом. Эти факторы также должны быть учтены в создаваемом АПС.

Автоматический переводной словарь ВЦП, как и все крупные АПС и автоматизированные терминологические банки данных, является одновременно словарем и своеобразной автоматизированной информационно-поисковой

системой, поэтому в ограничения, свойственные пользователю АПС, мы включаем ограничения, свойственные читателям традиционных переводных научно-технических словарей, а также ограничения, свойственные пользователям автоматизированных естественно-языковых информационных систем /Попов, 1982, с. 82/. В качестве основных ограничений пользователя АПС можно выделить:

1/ невозможность знать всю необходимую ему лексику даже в пределах одной тематической области;

2/ невозможность безошибочной работы пользователя с автоматическим словарем;

3/ неспособность некоторых категорий пользователей безошибочно выделять в тексте раздельно оформленные лексические единицы и сводить их к словарным формам и, как следствие, правильно оформлять запросы.

Разрабатываемый в ВЦП АПС относится к категории лингвистических автоматов, и поэтому должен создаваться с учетом требований теории и практики построения и использования систем этого типа. Исследователи выделяют следующие четыре основных принципа, которые должны учитываться при построении лингвистических автоматов ближайшего будущего:

1. Принцип интерактивной работы лингвистических автоматов. Смысл его состоит, с одной стороны, в том, чтобы обеспечить контроль человека над функционированием лингвистического автомата, а с другой стороны, чтобы разумно распределить лингвистические функции между лингвистическим автоматом и человеком. При этом автомат должен осуществлять массовые рутинные и полурутинные операции по отождествлению единиц текста с единицами словаря, осуществлять простейший грамматический и смысловой анализ входного текста. В то же время тонкие стилистические и семантико-синтаксические операции остаются пока в компетенции человека.

2. Принцип модульной архитектуры, который предусматривает построение лингвистического автомата в виде ансамбля программных модулей, каждый из которых воспроизводит определенный уровень и/или определенный аспект речемыслительной деятельности человека.

3. Принцип развития, который состоит в построении лингвистического автомата в виде открытой системы, развивающейся путем включения новых или обновления старых модулей.

4. Принцип самостоятельности функционирования модулей лингвистического автомата. Реализация этого

принципа позволяет отдельным модулям лингвистического автомата и их наборам работать автономно, выполняя при этом разнообразные виды автоматической переработки текста /Пиотровский, 1987, с. 71/.

Эти четыре основополагающих принципа построения лингвистических автоматов, а также учет ограничений, свойственных основным категориям пользователей АПС, специфики их главных задач и целей при работе с научно-технической литературой реализуются в следующих конкретных требованиях, которые мы предъявляем нашему автоматическому переводному словарю:

1. АПС должен включать лексические массивы тех языков, по которым выполняется основная часть переводов в организации, для которой создается данный АПС.

2. АПС должен быть обратимым, т.е. пользователь должен иметь возможность произвольно, по своему желанию выбирать входной и выходной языки словаря.

3. Словарные массивы не должны ограничиваться одной лексико-грамматической категорией, например только существительными и номинативными словосочетаниями или одной тематической областью, и должны включать лексику, представляющую наибольший интерес для основных категорий пользователей.

4. АПС должен допускать обращение в виде отдельных словоформ и текстовых форм словосочетаний.

5. АПС должен иметь возможность устранять в автоматическом режиме наиболее характерные ошибки пользователя при оформлении запроса.

6. В связи с разными требованиями пользователей к объему и составу информации в словарной статье АПС должен иметь возможность варьировать объем и состав информации в словарной статье по требованию пользователя.

7. АПС должен иметь возможность выдавать частичные ответы на сверхдлинные запросы.

8. АПС должен иметь возможность оперативной коррекции /изменения, пополнения, устранения/ лексических массивов, а также состава и объема информации в словарной статье.

1.3.3. Цели и задачи автоматического переводного словаря.

Основной целью нашего АПС является повышение производительности труда переводчика и редактора при

работе с текстами разных категорий, но в первую очередь с открытыми текстами.

Второй целью является частичная автоматизация лексикографических работ, например, для оперативно-го пополнения словарных компонентов систем машинного перевода и издания традиционных переводных терминологических словарей.

Поставленные цели достигаются в результате выполнения автоматическим переводным словарем следующих задач:

- предоставления пользователю переводных эквивалентов интересующей его лексики в режиме диалога либо в виде текстоориентированного словаря;
- автоматический перевод списочных частей научно-технической литературы и документации;
- сбор новой терминологии путем автоматической регистрации запрашиваемых и отсутствующих в автоматическом словаре терминов с последующим их анализом;
- автоматическая генерация переводных словарей и словников для каждого рабочего языка словаря по широкому набору признаков.

Решение основной задачи АПС обеспечивается благодаря разнообразию и гибкости лингвистических и программных средств АПС, которые позволяют:

- осуществлять запрос на любом из рабочих языков;
- получать ответ на любом из рабочих языков;
- осуществлять запрос в любой текстовой форме;
- работать со словарем в режиме диалога;
- получать текстоориентированные микрословари с размещением запрошенных лексических единиц либо в алфавитном порядке, либо в порядке их появления в тексте оригинала;

регулировать по желанию пользователя объем ответа;

получать без дополнительного обращения к словарю переводные эквиваленты всех слов из состава запрошенных, но не найденных в словаре словосочетаний;

получать без дополнительного обращения к АПС значения и возможные переводные эквиваленты для компонентов отсутствующих в словаре сложных слов;

получать информацию о тематической принадлежности термина;

получать информацию о "надежности" термина, т.е. об уровне его стандартизации или степени его употребительности;

получать толкование термина и примеры его контекстного употребления;

получать грамматическую информацию о запрашиваемом термине и его переводном эквиваленте.

Выделение перевода списочных частей документации в отдельную задачу объясняется спецификой технической, в особенности экспортно-импортной документации, которая насыщена различными перечнями, каталогами деталей, заявками и т.п. Особый режим автоматического словаря позволяет обрабатывать подобные тексты без участия человека, что значительно сокращает время выполнения переводов данного типа.

Важную роль в повышении эффективности переводческой деятельности играет осуществляемая автоматическим словарем задача автоматической регистрации новой терминологии. Бурное развитие многих отраслей науки и техники обуславливает высокую "подвижность" современных терминосистем, за изменениями в которых традиционные словари успеть не в состоянии. Оперативная регистрация новой терминологии с помощью АПС, регистрация частоты обращения к термину и введение индекса надежности дают возможность:

непрерывно и своевременно пополнять лексические массивы АПС за счет новой терминологии, поступающей в виде запросов пользователей;

способствовать изданию новых терминов в виде глоссариев, словарей, тетрадей и т.п. в традиционной книжной форме;

пополнять словарные компоненты систем автоматической обработки текстовой информации;

анализировать динамику терминосистем и накапливать данные для стандартизации терминологии;

ускорять издание и переиздание терминологических переводных словарей в традиционной книжной форме.

1.4. Режимы эксплуатации автоматического переводного словаря

Учитывая то, что переводы в ВЦП и в ряде других переводческих организаций страны выполняются силами штатных и нештатных сотрудников, АПС должен иметь разные режимы эксплуатации, отвечающие разным условиям их работы. АПС ВЦП может эксплуатироваться в режиме диалога и в режиме пакетной обработки запросов.

Режим диалога. При этом режиме ввод запросов и вывод ответов осуществляется через дисплей, выдача переводных эквивалентов производится отдельно каждой запрошенной лексической единице /одиночный поиск/ или одновременно несколькими лексическими единицами /групповой поиск/. При этом вся отображаемая на экране дисплея информация может одновременно выводиться на АЦПУ. Наиболее эффективен такой режим для выполнения срочных переводов при одновременной работе с АПС нескольких пользователей. Ввод запросов в режиме диалога, как правило, осуществляется самим пользователем. Этот режим предназначен в основном для штатных работников. Автоматический словарь может также использоваться в этом режиме в библиотеках, оснащенных необходимой вычислительной техникой.

Режим пакетной обработки запросов. При этом режиме осуществляется ввод сразу нескольких запросов с машиночитаемого носителя или дисплейного терминала. Выдача ответов производится одновременно всем или несколькими запрошенными лексическими единицами на АЦПУ или на магнитный носитель. При пакетном режиме эксплуатации АПС возникает разрыв между моментом осуществления запроса и моментом получения ответа, поэтому на первый взгляд создается впечатление, что такой режим противоречит одному из основных требований, предъявляемых к автоматическим словарям, а именно - повысить оперативность выполнения переводов. Однако если предположить, что обычный цикл такой эксплуатации не должен в нормальных условиях превышать двух дней /из расчета один день на заполнение запроса пользователем, перенос запроса на машиночитаемый носитель и прогон его через АПС и один - на выдачу текстоориентированного словаря пользователю/, то подобная задержка оказывается практически несущественной при выполнении переводов объемом в 3 а.л. и более. Этот вывод основывается на том, что для несрочных переводов штатному переводчику выделяется 10 дней на перевод текста объемом в 1 а.л. Опыт эксплуатации крупных АПС показал, что использование подобных текстоориентированных микрословарей, полученных на базе этих АПС, ведет не только к сокращению времени на перевод на 50-70%, но также способствует повышению качества перевода в первую очередь за счет унификации используемой терминологии /Язык и машины, 1968, с. 31; Мкртчян, 1979; Krollmann, Schuck, Winkler, 1965, s. 15/.

1.5. Многоязычие автоматического переводного словаря

Как было показано в разд. 1.2.1, ВЦП выполняет переводы с более чем 30 иностранных языков на русский и с русского почти на 30 иностранных языков. Из этого следует естественный вывод, что АПС ВЦП должен быть многоязычным. Однако создание автоматического переводного словаря с рабочими лексическими массивами для 30 языков - задача в настоящее время практически неразрешимая. По состоянию на начало 80-х гг. промышленные АПС и терминологические банки включают максимально 8-10 языков /Sager, McNaught, 1980/. Одновременно с этим в ВЦП нет практической необходимости создания АПС с большим количеством языков. Объем переводов, выполняемых с иностранных языков на русский, распределяется весьма неравномерно. Так, четыре ведущих языка - английский, немецкий, японский и французский - охватывают 90,5% всех переводов на русский язык. С русского языка основная часть переводов выполняется на английский /примерно 80%/. Таким образом, с практической точки зрения достаточно ограничить количество языков в словаре 3-5 основными языками.

В настоящее время АПС ВЦП разрабатывается как многоязычный словарь с лексическими массивами и лингвистическими алгоритмами для английского, немецкого, французского и русского языков. Отсутствие японского языка объясняется тем, что проблемы ввода и вывода иероглифической письменности представляют для нас в настоящее время серьезные технические и лингвистические сложности. Логика лингвистического и программного обеспечения АПС ВЦП не препятствует увеличению количества разных языков в словаре.

Основополагающим принципом при решении проблемы многоязычия автоматического словаря является способ организации связей между разноязычными лексическими единицами в данном словаре. В нашем АПС используется принцип независимого описания разноязычных лексических единиц, когда каждая лексическая единица описывается независимо для каждого языка и только средствами и по законам этого же языка. Это означает, что такие основные виды сведений о слове, как его грамматическая информация, описание омонимии и многозначности, структурирование разных значений многозначного слова в словарной статье, ограничительная и уточ-

няя информация, индекс надежности и тематическая принадлежность, должны приписываться, например, английскому слову только на основании законов английского языка. Информация, которую получает русское слово, являющееся переводным эквивалентом упомянутого выше английского слова, определяется на основании русского языкового материала и по законам русского языка и в принципе может весьма существенно отличаться от соответствующей информации английской лексической единицы по многим зонам. Словарная статья в таких АПС обычно состоит из значительного количества разных зон, каждая из которых предназначена для хранения однотипной лингвистической или экстралингвистической информации и хранится в памяти ЭВМ отдельно. На логическом уровне все зоны лексических единиц одного языка как бы объединены в единый одноязычный словарь, который связан системой связей с одноязычными словарями других языков, образуя таким образом многоязычный словарь.

Принцип независимого описания разноязычных лексических единиц в многоязычном автоматическом словаре имеет свои положительные и отрицательные стороны. Основным преимуществом такого подхода является то, что он позволяет на теоретическом уровне точно, глубоко и непротиворечиво описывать лексику разных языков своими собственными средствами, не навязывая законов и традиций описания одного языка другому. Как следствие этого, можно достичь обратимости переводного словаря, а также генерировать одноязычные словари с различными словарными статьями, чего в условиях традиционной лексикографии или при использовании других принципов решения проблемы многоязычия в АПС добиться значительно сложнее.

Отрицательные стороны независимого описания разноязычной лексики сводятся в основном к трудностям практической реализации этого подхода. Основными источниками пополнения автоматических словарей в большинстве случаев служат переводные словари, глоссарии, Тетради новых терминов и другие подобные виды лексикографических изданий в традиционной книжной форме, в которых уже теоретически нарушен принцип независимого описания лексики разных языков, так как лексика входного языка описывается с помощью лексики выходного языка и через призму этого же языка, что наиболее ярко проявляется в словарях интенсивного типа,

т.е. словарях с богатой словарной статьей. Практически вся информация в словарной статье переводного словаря относится к заглавному слову, а именно: стилистические и оценочные пометы, неформальные ограничения на употребительность, уточнения значений и контекстные ограничения и разъяснения. Грамматическая и фонетическая информация дается в подавляющем большинстве переводных словарей только заглавному слову. Автоматический перенос этой информации с заглавного слова на его переводные эквиваленты приведет к искажению значений лексических единиц выходного языка. Поэтому при использовании переводных словарей в качестве источников пополнения лексических массивов автоматического переводного словаря лексика выходного языка пополняется необходимой информацией в соответствии с требованиями лингвистического обеспечения АПС. Так, всем русским словам, поступившим в наш АПС в качестве переводных эквивалентов иноязычной лексики, приписывается следующая информация: тематическая рубрика и подрубрика, стилистическая принадлежность, индекс надежности, лексикограмматический класс. В зоне помет указывается синтаксическая, лексическая, грамматическая и другая информация, которая необходима для разграничения многозначности и омонимии, а также уточняет сферу употребления данного русского слова. На каждую русскую лексическую единицу составляется лемматизационная словарная статья, в которой указывается словоизменительная основа /или основы/, порядковые номера основ и словоизменительный класс. Если данная русская лексическая единица является отдельным словом, то создается, как правило, одна лемматизационная словарная статья. Если эта лексическая единица – словосочетание, то составляются лемматизационные словарные статьи на каждое слово, входящее в это словосочетание. Информация, приписанная русским лексическим единицам в зоне помет, используется затем при структурировании словарных статей многозначных русских слов, когда они уже становятся заглавными словами.

1.6. Обратимость автоматического переводного словаря

Под обратимостью автоматического переводного словаря понимается возможность произвольно в процес-

се его эксплуатации по желанию пользователя менять местами входной и выходной языки. Необходимость придания автоматическому словарю свойства обратимости предопределяется тем, что специализированные переводческие организации, как правило, осуществляют перевод в обе стороны: с иностранных языков на родной и с родного на иностранные. В условиях ВЦП объем переводов на иностранные языки заметно уступает переводам с иностранных языков, однако недопустимо недооценивать роль этого вида перевода. Этот вид перевода относится к переводу для распространения информации, поэтому требования к такому виду перевода весьма высокие. Здесь неприемлем информационный или сигнальный перевод. Перевод для распространения информации должен выполняться человеком-переводчиком, вооруженным традиционными или автоматическими словарями, системой МП интерактивного типа либо системой МП традиционного типа с обязательным тщательным постредактированием машинного перевода человеком. Следует также учитывать возможность того, что в других переводческих организациях перевод на иностранные языки может играть более заметную роль. Таким образом, для ВЦП, как и для большинства других специализированных переводческих организаций, необходимы АПС для перевода с иностранных языков на родной и с родного на иностранные. Это требование может быть выполнено либо с помощью двух необратимых словарей для каждой пары языков, либо одним обратимым словарем для каждой пары языков, либо же одним обратимым многоязычным словарем для нескольких /в принципе всех/ языков. Мы считаем более оптимальным создание обратимых словарей.

В многоязычном словаре проблему обратимости можно рассматривать на двух уровнях: на уровне языков и на уровне отдельных лексических единиц. Обратимость многоязычного словаря на уровне языков /языковая обратимость многоязычного словаря/ означает, что языки, входящие в состав этого словаря, могут выступать как в качестве языков запроса, так и в качестве языков ответа. В том случае, когда все языки из состава словаря могут быть как входными, так и выходными, можно говорить о полной обратимости этого словаря на уровне языков. Если это требование удовлетворяется не для всех языков, то в этом случае обратимость будем называть частичной.

Многоязычный автоматический словарь ВЦП на уровне языков обладает полной обратимостью, т.е. позволяет осуществлять запрос и получать ответ на любом из четырех включенных в словарь языков: русском, английском, французском и немецком. Лингвистическое и программное обеспечение разработаны таким образом, что в них не содержится принципиальных препятствий для эксплуатации словаря в любом из теоретически возможных двуязычных вариантов: англо-русском, русско-английском, немецко-русском, русско-немецком, французско-русском, русско-французском, англо-немецком, немецко-английском, англо-французском, французско-английском, немецко-французском и французско-немецком. Реальные условия работы ВЦП не требуют эксплуатации автоматического словаря во всех его возможных вариантах. Практически не нужны варианты, когда входной и выходной языки являются иностранными. Однако в других условиях или при добавлении еще одного языка, например болгарского или венгерского, такая возможность может возникнуть. Поэтому мы считаем целесообразным сохранить полную обратимость нашего АПС на уровне языков.

Обратимость многоязычного словаря на уровне отдельных лексических единиц /лексическая обратимость многоязычного словаря/ означает, что каждая лексическая единица может выступать в качестве запроса и в качестве ответа. В том случае, когда любая лексическая единица любого из языков многоязычного словаря может быть входной и выходной, можно говорить о полной обратимости этого словаря на уровне лексических единиц. При полной лексической обратимости словаря каждая лексическая единица, выступая в качестве единицы входного языка, должна иметь переводные эквиваленты на всех остальных языках, входящих в состав многоязычного словаря.

Полная лексическая обратимость многоязычного словаря достижима при эквивалентности лексических массивов всех языков. Если же разноязычные лексические массивы неэквивалентны, то можно говорить об обратимости словаря только для языковых пар, в нашем случае, например, для английского и русского, французского и русского языков.

Немаловажным фактором, влияющим на лексическую обратимость АПС, является качественный состав его лексических массивов. Полной лексической обратимости можно достичь только при условии, что в качестве главных слов и их переводных эквивалентов всегда вы-

ступают отдельные слова или словосочетания, являющиеся раздельнооформленными лексическими единицами. Однако это условие в переводной терминологической лексикографии выполняется далеко не всегда. Нередко в переводных терминологических словарях в качестве переводных эквивалентов вместо лексических единиц используются так называемые развернутые переводные эквиваленты, которые передают значение заглавной лексической единицы описательным способом и более соответствуют по своей структуре и величине толкованиям, чем переводным эквивалентам.

В качестве примеров такого явления можно привести следующие переводные эквиваленты, взятые из опубликованных терминологических словарей и Тетрадей новых терминов: *central diagnostics* - диагностика /технологического оборудования, например гибкой производственной системы/, выполняемая центральной ЭВМ /обслуживающей эту систему/; *quadrant sawing* - распиловка бревна с получением четырехкантного бруса и последующей распиловкой его на четыре доски с поворотом бревна после каждого реза на 180° и 90° ; *coefficient d'utilisation pratique* - коэффициент относительной работоспособности взрывчатого вещества; *médecin embaucheur* - врач, ведающий осмотром лиц, поступающих на работу в шахту.

Словари со словарными статьями, в которых используются развернутые переводные эквиваленты, могут функционировать в обе стороны, т.е. быть обратимыми, только при условии, что все развернутые переводные эквиваленты запрашиваются точно в той же форме, в какой они зафиксированы в этих словарях. Однако для практической работы это слишком жесткое условие. Невозможно требовать от пользователя, чтобы он точно помнил все подобные переводные эквиваленты, либо рассчитывать на то, что они будут встречаться в тексте именно в том виде, в каком уже зафиксированы в словаре. Словари со словарными статьями такого типа обладают частичной лексической обратимостью. Они будут работать от заглавной лексической единицы к развернутому переводному эквиваленту, но обратный путь в обычных условиях практически не реализуем. В связи с тем, что основным источником лексических массивов для нашего АПС служат переводные терминологические словари, Тетради новых терминов, различные глоссарии, составители которых нередко прибегают к использованию развернутых пере-

водных эквивалентов, наш АПС обладает частичной лексической обратимостью. Некоторые лексикографические приемы, используемые для уменьшения влияния развернутых переводных эквивалентов на лексическую обратимость словаря, будут рассмотрены в разд. 2.1.4.

Для уменьшения остроты этой проблемы в разрабатываемом АПС в перспективе предусматривается использование приема, характерного для информационно-поисковых систем, а именно поиск словосочетаний по ключевым словам. В тех случаях, когда пользователь не в состоянии получить требуемую ему словарную статью из-за большой длины развернутого переводного эквивалента, он может перейти на особый режим и запросить все словосочетания, в которые входят интересующие его слова. Поиск при этом осуществляется по зоне переводных эквивалентов и по зоне толкований. Так, при эксплуатации словаря в русско-французском варианте в режиме поиска ключевых слов на примере приведенных выше словарных статей при запросе слов "врач" и "шахта" пользователь получит на первом этапе развернутый переводной эквивалент "врач, ведающий осмотром лиц, поступающих на работу в шахту", а затем и его французский переводной эквивалент *médecin embaucheur*, а при запросе "распиловка" и "бревно" - соответственно "распиловка бревна с получением четырехкантного бруса и т.д." и затем его английский эквивалент *quadrant sawing*. Следует отметить, что при использовании режима запроса по ключевым словам пользователь может получить помимо необходимой ему и другие словарные статьи. Этот естественный информационный шум мы считаем разумной платой за достижение обратимости АПС в таких сложных случаях.

Необходимость поиска по этим двум зонам объясняется тем, что во многих случаях развернутый переводной эквивалент занимает промежуточное положение между собственно переводным эквивалентом и толкованием. Это приводит к тому, что одни составители словарных статей оставляют такие развернутые ПЭ в зоне переводных эквивалентов, а другие относят их в зону толкований. В связи с тем что лексические массивы достаточно велики, а составители словарных статей разобщены территориально, весьма сложно добиться полной унификации в решении этого вопроса для всех словарных статей. Однако такое на первый взгляд непоследовательное решение не сказывается на эксплуата-

ционных характеристиках АПС. Мы считаем такой путь естественным рабочим компромиссом при создании крупных промышленных автоматизированных систем. Гибкость лингвистического и программного обеспечения АПС позволяет в любое время корректировать все зоны словарных статей и добиваться унификации в решении этой проблемы, если в этом возникнет необходимость.

Резюмируя этот раздел, можно сказать, что наш многоязычный автоматический словарь потенциально обладает полной обратимостью на уровне языков. Однако в условиях ВЦП обратимость словаря реализуется в языковых парах с участием русского языка, т.е. английский-русский, немецкий-русский и французский-русский языки. На лексическом уровне словарь обладает неполной обратимостью для всех языковых пар.

1.7. Динамичность и гибкость АПС

Проблема динамичности словаря, под которой обычно понимают возможность оперативной коррекции словника словаря и содержания словарной статьи в нем, является одной из самых сложных и трудноразрешимых в традиционной лексикографии. Практика показывает, что целый комплекс разнообразных причин - лексикографических, издательских, технических, экономических и др. - не позволяет оперативно переиздавать нужные читателю словари. Анализ динамики переиздания словарей в нашей стране показал, что переводные словари даже по таким перспективным и динамично развивающимся отраслям науки и техники, как биология, вычислительная техника и автоматика, переиздаются у нас в среднем один раз в десять лет. Примерно такое же положение и в нашей одноязычной лексикографии /Шайкевич, 1986/. В этом проявляется одно из основных свойств традиционных словарей - их статичность. Действительно, список слов /а также содержание словарной статьи - И.И.У./ в любом изданном или завершённом словаре является закрытым, хотя от издания к изданию он может обновляться /Караулов, 1976, с. 70/.

На основании данных о периодичности переиздания переводных и одноязычных словарей можно с полной уверенностью утверждать, что традиционная лексикография в среднем на 7-10 лет отстает от процесса фактического обогащения или изменения терминологических

систем современных активно развивающихся направлений науки и техники. В качестве одного из возможных путей частичного решения проблемы коррекции состава традиционных словарей используется выпуск различных дополнений к словарям, которые выходят либо самостоятельно, либо в составе того же словаря при его стереотипном переиздании. Той же цели служат Тетради новых терминов, терминологические бюллетени и другие формы оперативного издания новой лексики. Тем не менее эта проблема в целом пока остается в традиционной лексикографии не решенной. Запаздывание с включением новой лексики отмечается переводчиками как один из основных недостатков традиционных словарей. Словарная статья "бумажного" словаря также является статичной как по своей структуре, так и по содержанию. Это означает, что любая неточность или ошибка могут быть исправлены, любая дополнительная информация может быть внесена только при очередном переиздании словаря.

В отличие от традиционных словарей словари систем МП и АПС потенциально обладают свойством динамичности. Отбор лексики не прекращается по завершении создания АПС, а продолжается в течение всего времени его функционирования. Постоянно происходит исключение оказавшихся ненужными слов и включение новой лексики, коррекция и пополнение информации в любой зоне словарной статьи. Переменность состава и способность воспринимать изменения в ходе функционирования являются принципиальными характеристиками, отличающими машинные словари от обычных /Марчук, 1983, с. 133/. Динамичность состава автоматических словарей достигается в основном за счет гибкости программного обеспечения. На лингвистическом уровне достижение динамичности словаря требует разработки четкой структуры словарной статьи и записи однотипной информации строго в одну и ту же зону во всех словарных статьях. На организационном уровне резко возрастает роль слежения за изменениями словника словаря, состава и объема информации в словарных статьях.

Под гибкостью АПС автор понимает комплекс лингвистических и программных приемов, которые упрощают обращение к словарю, расширяют возможности пользователя при работе с иноязычным текстом, снижают требования к уровню знания входного языка, а также позволяют использовать АПС не только для хранения лексических массивов, но и для решения других задач ав-

томатической обработки текстов и автоматизации лексикографических работ. Конкретно под гибкостью АПС, создаваемого для нужд переводческой организации, мы будем понимать возможность:

- эксплуатировать АПС в разных режимах ввода запроса и получения ответа;
- осуществлять запрос лексической единицы как в словарной, так и в текстовой форме;
- варьировать ответ по объему и по составу информации в зависимости от категории пользователя и типа задачи, стоящей перед ним;
- автоматически корректировать наиболее характерные ошибки пользователя.

Особым проявлением гибкости автоматического переводного словаря является его обратимость.

Наиболее важным средством повышения гибкости АПС является упрощение формы запроса, т.е. обращения к словарю. В традиционных переводных словарях форма обращения совпадает со словарной формой, т.е. с формой хранения лексических единиц в словаре. Это означает, что пользователь должен уметь переходить от словоформ к словарным формам. Однако практика показывает, что значительная часть пользователей не может надежно решать эту задачу. Трудности определения словарной формы слова возникают даже в тех случаях, когда входной язык является для пользователя родным, но их еще больше при работе с неродными языками. Особенно остро эта проблема стоит для языков синтетического типа. Для русского языка наибольшие сложности возникают при обработке видовых форм глагола, причастий, супплетивных форм, приставочных форм глаголов. Такие же трудности возникают в английском языке при обработке неправильных глаголов, во французском языке - всех глаголов и неправильных глаголов в особенности, в немецком языке - сильных глаголов, форм существительных и прилагательных, образуемых с помощью умляута. Сложности усугубляются также тем, что несловарные формы слова встречаются в текстах значительно чаще, чем их словарные формы. Так, русский и французский глаголы реализуются в текстах в словарной форме в среднем в 5 раз реже, чем в несловарных формах, русское существительное в словарной форме встречается в текстах почти в три раза реже, чем в несловарных формах /Кожина, 1972, с. 139; Убин, 1984, с. 169/. Автоматический переводной словарь, рассчитанный на широкий круг

пользователей с разными уровнями знания входного иностранного языка, должен иметь возможности автоматического решения подобных лингвистических задач.

Гибкость и динамичность АПС требуют решения целого ряда лингвистических проблем. Среди них - разработка новой словарной статьи, которая должна обладать разветвленной и строгой структурой, чтобы, с одной стороны, имелась возможность фиксировать самую разнообразную информацию, необходимую разным категориям пользователей, а с другой - автоматически обрабатывать эту информацию /корректировать, устранять, расширять, переносить и т.п./. Гибкость обращения к лексическим массивам АПС требует разработки и включения в его состав лингвистического аппарата, который отсутствует в традиционных словарях и свойствен автоматизированным системам обработки текстовой информации. К такому аппарату в первую очередь относятся лингвистические алгоритмы, решающие задачу сведения текстовых форм лексических единиц к их словарным формам. АПС должен также иметь возможность автоматически осуществлять первичный анализ обращений к нему на разных уровнях, например осуществлять контроль за "спрашиваемостью" лексики по тематике или источникам. Подобная информация позволит объективно и оперативно корректировать лексические массивы по результатам эксплуатации.

Гибкость и динамичность автоматического словаря являются существенными преимуществами АПС перед традиционными словарями. Гибкость автоматического словаря позволяет рассматривать его не только как хранилище лексической информации, но и как особый инструмент по ее обработке. Динамичность АПС дает возможность разработчику и пользователю постоянно поддерживать актуальность и четкую целевую направленность его лексических массивов.

1.8. Выводы

В настоящем разделе дан краткий анализ производственной сферы научно-технического перевода на примере крупной переводческой организации, для которой проектируется АПС. Производственная сфера научно-технического перевода представляет собой всю совокупность объектов, задач и условий выполнения научно-

технического перевода. Основными среди них являются исходный материал и конечный продукт /в наших условиях - это научно-технический текст на языке оригинала и его перевод на другом языке/, основные исполнители перевода, основные языки и направление перевода. Учет требований производственной сферы позволили определить диапазон задач, на решение которых должен быть нацелен автоматический переводной словарь, условия, в которых он будет разрабатываться и эксплуатироваться.

Анализ производственной сферы выполнения научно-технического перевода в условиях ВЦП привел нас к выводу, что автоматический переводной словарь, проектируемый для этой сферы, должен обладать следующими обязательными свойствами - многоязычием, обратимостью, гибкостью и динамичностью. В силу того что на примере ВЦП отражается обстановка в системе научно-технического перевода в стране в целом, можно предположить, что АПС, обладающий такими свойствами, будет полезен и другим переводческим, информационным и научно-исследовательским организациям, выполняющим научно-технический перевод.

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ АВТОМАТИЧЕСКОГО
ПЕРЕВОДНОГО СЛОВАРЯ ВЦП

Под лингвистическим обеспечением АПС мы понимаем совокупность приемов и методов решения конкретных лингвистических задач, встающих при создании многоязычного АПС, а также сами эти решения в виде лингвистических алгоритмов, грамматических описаний языков, входящих в состав АПС, массивов языковых объектов разных уровней, представленных в форме, необходимой для функционирования лингвистических алгоритмов данного АПС. К сфере лингвистического обеспечения относятся также создание структуры словарной статьи для всех типов языковых единиц АПС, определение состава и объема информации в словарной статье, поиск путей гибкого варьирования состава и объема информации в словарной статье АПС в соответствии с потребностями конкретного пользователя.

2.1. Состав и объем лексических единиц АПС ВЦП2.1.1. Типы самостоятельных единиц состава АПС

Под самостоятельными единицами состава АПС мы понимаем единицы, способные выступать в роли заглавных единиц в словарных статьях этого словаря. Номенклатура самостоятельных единиц АПС является производной от задач и целей АПС и принципов построения его лингвистического обеспечения. В нашем словаре в качестве самостоятельных выступают следующие типы языковых единиц: морфемы, цельнооформленные лексические единицы /отдельные слова/, раздельнооформленные лексические единицы /словосочетания/, фрагменты текста. Понятия цельнооформленности и раздельнооформленности лексических единиц определяются в данной работе в соответствии с концепциями А.И. Смирницкого /Смирницкий, 1981/.

Морфема. Самостоятельные словарные статьи для морфем создаются с целью обеспечения работы алгоритмов словообразовательного анализа. Массив словообразовательных морфем /префиксов и суффиксов/ в АПС включает для каждого языка как традиционно выделяемые словообразовательные элементы, так и наиболее ха-

рактерные для научно-технических терминсистем аффиксы и полуаффиксы. На каждую словообразовательную морфему составляется отдельная словарная статья, которая по своей структуре весьма близка словарной статье отдельного слова.

Цельнооформленная лексическая единица /отдельное слово/ теоретически и композиционно является основной единицей состава АПС. Базовая структура словарной статьи разрабатывалась для цельнооформленных лексических единиц /ЛЕ/ и реализуется в полном объеме в подавляющем большинстве случаев при описании этого типа ЛЕ. К отдельным словам осуществляется привязка всех словосочетаний, в которые входят эти слова. Однако в количественном отношении цельнооформленная ЛЕ уступает в нашем АПС словосочетанию.

Раздельнооформленная лексическая единица /словосочетание/ в количественном отношении является основной единицей АПС. Словарная статья словосочетания обычно беднее словарной статьи отдельного слова и, как правило, включает только зоны, обязательные для заполнения. Количественное преобладание словосочетаний в нашем словаре объясняется двумя основными причинами: 1/ основная часть лексического состава АПС — это терминологическая лексика, которая в свою очередь представлена в основном словосочетаниями; 2/ необходимость учета особенностей обращения пользователей к словарю, поскольку примерно 80% обращений к автоматическому словарю составляют словосочетания /Krollmann, Schuck, Winkler, 1965, s. 22/.

Промежуточное положение между цельнооформленными и раздельнооформленными ЛЕ занимают аббревиатуры и сокращения. В практике переводной терминологической лексикографии весьма часто понятия "аббревиатура" и "сокращение" рассматриваются как идентичные. В большинстве словарей используется для обозначения этих языковых явлений термин "сокращение". В своей работе мы также используем только один термин — "сокращение", понимая его как слово, образованное сложением начальных букв или звуков слов исходного словосочетания. Специфической особенностью сокращений является то, что с позиций лексикографического описания они имеют две формы: основную — собственно сокращение и развернутую, под которой мы будем понимать исходное словосочетание, на основе которого создано данное сокращение. Сокращение в своей основной форме описы-

вается в нашем АПС как отдельное слово, а в развернутой форме — как словосочетание.

Сокращения занимают особое место в переводных терминологических словарях, и многие авторы выносят их в отдельный раздел словаря. В нашем АПС этим лексическим единицам также уделяется повышенное внимание. Это необходимо в связи с тем, что сокращения представляют одну из основных трудностей при работе с иноязычной научно-технической литературой. Многие авторы научно-технических публикаций весьма широко используют различные сокращения, как общеизвестные и стандартизованные, так и узкоспециальные, характерные только для одного направления исследований, одной научной школы или фирмы или даже одного исследователя. Нередко сокращения вводятся в текст без ссылки на исходное словосочетание, что создает дополнительные трудности из-за большой омонимии сокращенных форм терминологических словосочетаний и отсутствия этих сокращений в традиционных переводных словарях. Для уменьшения остроты этой проблемы каждое сокращение в нашем АПС имеет отсылку к исходному словосочетанию, которое выдается пользователю при запросе этого сокращения одновременно с его переводными эквивалентами, которые в свою очередь также фиксируются в его словарной статье в двух формах: сокращенной и развернутой.

Фрагменты текста. Включение фрагмента текста в АПС ВЦП в качестве самостоятельных ЛЕ отражает существующую тенденцию расширения состава автоматических словарей за счет единиц текста: развернутых словосочетаний, отдельных предложений, фрагментов и даже целых текстов /Schulz, 1978, р. 159/. Подобный подход к комплектованию словника переводного словаря в традиционной лексикографии не используется. Однако в вычислительной лексикографии такое решение в определенных условиях оказывается весьма удобным. Дело в том, что АПС показали себя эффективным инструментом полуавтоматического /перевод с помощью ЭВМ/ перевода стандартизованных текстов, например товаросопроводительной документации или юридических документов, в которых определенные части регулярно повторяются без существенных изменений.

В наш АПС фрагменты текста чаще попадают в качестве русских, так называемых развернутых переводных эквивалентов новых иноязычных терминов. В самостоятельные вокабулы они превращаются в силу обратимости

АПС. Возникновение в переводной терминологической лексикографии такого явления, как развернутый переводной эквивалент /ПЭ/, который более соответствует краткому толкованию, чем переводному эквиваленту в обычном понимании, объясняется, на наш взгляд, в первую очередь неравномерностью развития терминосистем одной и той же науки в разных языках. Возникновение нового термина на родном языке его создателя весьма часто намного опережает возникновение и всеобщее принятие его эквивалента на других языках. Немаловажно также то, что нередко первыми с новым иноязычным термином сталкиваются не терминологи, а переводчики и лексикографы. Составление переводных терминологических словарей, Тетрадей новых терминов, глоссариев и т.п. заметно опережает процесс становления и стандартизации терминологии, поэтому авторам таких переводных терминологических изданий часто приходится искать переводные эквиваленты новых иноязычных терминов в научно-технических переводах или создавать их самим, используя для этого транслитерацию или буквальный перевод. Однако в ряде случаев эти привычные приемы не могут помочь лексикографу, и тогда он вынужден прибегать к использованию так называемых развернутых переводных эквивалентов. На наш взгляд, развернутый переводной эквивалент является неизбежным лексикографическим злом на начальном этапе формирования терминосистемы той или иной науки, однако по мере развития терминосистемы и хода стандартизации терминологии развернутые переводные эквиваленты следует заменять в переводных словарях краткими общепринятыми переводными эквивалентами.

Современные промышленные АПС создаются на базе мощных ЭВМ, имеющих огромные ресурсы памяти, которые позволяют разместить и успешно работать с лексическими массивами в сотни тысяч словосочетаний. Вместе с тем нестрогий подход к верхней границе словосочетаний приводит к тому, что эксплуатационные характеристики АПС существенно ухудшаются: например, увеличивается время поиска запрошенных ЛЕ в массивах АПС, растет количество отказов, поэтому следует избегать перенасыщения АПС свободными словосочетаниями. Единицы текста целесообразно использовать в качестве самостоятельных единиц АПС только в тех специфических случаях, когда это

дает безусловный выигрыш в качестве и скорости перевода исходного текста. Развивая мысль В.П. Беркова о традиционных двуязычных словарях, можно сказать, что на уровне словосочетаний переводные автоматические словари должны быть в значительной степени дифференциальными и должны, за некоторыми исключениями, фиксировать в эксплицитной форме лишь случаи расхождений, число которых, по-видимому, на много порядков меньше числа словосочетаний, находящихся в поэлементном соответствии /Берков, 1973, с. 117/.

Количество свободных словосочетаний и единиц текста в автоматическом словаре определяется в первую очередь типом словаря и его задачами. Чем больше ограничены цели АПС, чем уже он тематически, тем больше такой словарь является текстоориентированным, тем больше в нем может и должно быть свободных словосочетаний и единиц текста. Чем шире диапазон задач АПС, чем шире тематический спектр лексического состава АПС, тем более такой словарь должен быть системоориентированным, тем меньше в нем должно быть свободных словосочетаний и единиц текста.

2.1.2. Распределение лексического состава по частям речи

Состав лексических массивов имеет основополагающее значение для любого, в том числе и автоматического словаря и в значительной степени определяет его успех или неудачу. Для АПС отбор лексики обуславливается прежде всего тем, для решения каких задач создается этот словарь и на какой круг пользователей он ориентируется. АПС ВЦП предназначен для решения достаточно широкого круга задач, главной из которых является повышение производительности труда переводчика, редактора и технического специалиста при работе с иноязычной литературой, поэтому наш АПС включает в основном лексику трех грамматических классов: существительное, глагол и прилагательное.

Распределение лексического состава
по частям речи /в %/ в АПС ВЦП

Языки	Существительное	Прилагательное	Глагол	Прочие отдельные слова	Номинативные словосочетания
Английский	28	9	28	3	32
Французский	16	5	3	2	74
Немецкий	41	3	1	2	53

В класс прочих отдельных слов отнесены названия команд, операторов, наречия, предлоги, союзы, частицы, местоимения, аббревиатуры и числительные.

Представленные выше данные являются предварительными, поскольку работа по наращиванию объемов лексических массивов АПС продолжается непрерывно. Вместе с тем некоторые выводы по составу АПС ВЦП можно сделать. Первое – это преобладание в нашем словаре существительных и номинативных словосочетаний. По нашему мнению, этот факт объясняется тем, что наш АПС в первую очередь является специализированным терминологическим словарем, а терминология в своей основной части представлена именно этим классом лексики. Обращает на себя внимание достаточно высокая доля глагольной лексики в английской части словаря, особенно по сравнению с французской и немецкой частями словаря. Французская и немецкая части АПС ВЦП комплектуются в настоящее время пока только на основании традиционных терминологических словарей, поэтому полученные данные фактически отражают специфику лексического состава этих словарей, для которых характерно преобладание существительных и номинативных словосочетаний.

С точки зрения профессиональных переводчиков бедность в показе несубстантивных ЛЕ является серьезным недостатком существующих традиционных терминологических словарей в целом /Shaikovich, Oubine, 1988/. Среди основных категорий пользователей нашего АПС мы выделяем технических специалистов, которым глагольные

ЛЕ более важны для понимания всего текста, чем даже переводчикам, поэтому мы стремимся устранить или хотя частично ослабить этот недостаток традиционных терминологических словарей. В английской части АПС нам частично удалось это сделать за счет увеличения доли глагольной лексики. В дальнейшем это же будет сделано для французской и немецкой частей.

Суммарный вес субстантивных ЛЕ /отдельных слов и словосочетаний/ в немецкой и французской частях нашего АПС примерно равны. Однако в немецкой части в количественном отношении преобладают цельнооформленные ЛЕ, а во французской части преобладают словосочетания. На наш взгляд, это является отражением объективных различий, существующих между этими языками на лексическом уровне, а именно тем, что многие термины, отражающие одни и те же понятия, в немецком языке оформляются как сложные слова, а во французском – как словосочетания.

2.1.3. Тематическая принадлежность лексических единиц

Отбор лексики по тематическому признаку в АПС ВЦП определяется тем, на обработку каких текстов он в первую очередь нацелен. На основании результатов анализа тематики переводов ВЦП, а также с учетом некоторых внешних факторов было принято решение комплектовать лексические массивы АПС по следующим основным тематическим направлениям: электроника, вычислительная техника, программирование, авиация. Во всех тех случаях, когда возникают сомнения в отнесении конкретной лексической единицы к выбранной или к одной из смежных тематик, применяется расширительный подход, т.е. данная лексическая единица включалась в словарь с указанием ей индексов двух тематик одновременно. Такое решение представляется нам целесообразным, поскольку лексика смежных областей науки и техники в значительной степени пересекается /К вопросу о структуре..., 1986/.

Зона тематической принадлежности в нашем АПС имеет два уровня глубины описания: на уровне тематической рубрики и далее на уровне тематической подрубрики. Тематические рубрики совпадают с тематическими областями, указанными выше. Каждая тематическая руб-

рика подразделяется на основании Рубрикатора ГАСНТИ на отдельные подрубрики. Для рубрики "Вычислительная техника" выделено 22 тематические подрубрики, например цифровые ЭВМ, аналоговые ЭВМ, принцип обработки, элементы, узлы и устройство ЭВМ, математическое обеспечение, вычислительные сети и комплексы и т.п. Для рубрики "Авиация" выделено 18 подрубрик: авиационные конструкции, космонавтика, бортовое оборудование, электрооборудование, силовые установки, типы летательных аппаратов и т.п.

Структура словарной статьи АПС ВЦП предусматривает обязательную простановку индекса тематической принадлежности для каждого переводного эквивалента лексической единицы. Для записи используются общепринятые сокращения, например ВЧТ - вычислительная техника, лингв. - лингвистика. Приписывание тематической пометы переводному эквиваленту, а не всей ЛЕ в целом, позволяет выделять в пределах одной и той же заглавной ЛЕ разные значения и даже разные переводные эквиваленты в пределах одного и того же значения, относящиеся к разным тематическим областям. Такой подход дает возможность более подробно описывать значение лексической единицы.

В связи с ориентацией АПС на обработку открытых текстов, а также с учетом уровня знания иностранного языка техническим специалистом в словаре заметно представлена общеупотребительная лексика как в виде самостоятельных словарных статей, так и в виде нетерминологических значений многозначных слов. Общеупотребительной лексике проставляется индекс ОБЩ.

Выделение переводов списочных частей технической документации как особой задачи АПС требует включения в состав АПС всей лексики этого типа текстов вне зависимости от ее тематической принадлежности. В нашем случае такая задача ставилась только перед англо-русской версией АПС по авиационной тематике, поэтому в его состав помимо лексики из традиционных англо-русских словарей по авиационно-космической тематике вошла вся лексика из документации, содержащей подлежащие переводу списочные части.

2.1.4. Новая лексика

Новая лексика представляет в настоящее время одну из самых злободневных проблем терминологической

лексикографии. Бурное и одновременное развитие многих отраслей науки и техники ведет к появлению большого количества новых терминов, заимствованию терминологии из соседних смежных дисциплин, расширению значений "старых" терминов и приданию терминологических значений общеупотребительной лексике. Наиболее остро проблема новых терминов встает при работе /переводе, редактировании, чтении и т.п./ с иноязычной научно-технической литературой, так как в этом случае к задаче понимания значения нового термина на языке оригинала добавляется задача поиска или порождения его переводного эквивалента на языке перевода. Вероятно, поэтому переводчики и научные работники, отвечая на вопросы нашей анкеты о достоинствах и недостатках переводных терминологических словарей, практически единодушно выделили проблему новой терминологии как одну из самых острых проблем современной терминологической лексикографии, требующих оперативного решения. Достоинства переводного терминологического словаря прямо связываются с тем, насколько он охватывает наиболее важную и в первую очередь новую терминологию по своей тематике. Оперативность издания новых терминов приобретает первостепенное значение для терминологической лексикографии. Выдвигаются требования издания новых терминов в любой форме с периодичностью не реже 1/2 - 1 год /Shaikevich, Oubine, 1988/.

Не меньшую трудность новые термины представляют также для систем МП и автоматических переводных словарей. Мы не имеем точных данных о доле новой терминологии в текстах, переведенных системами МП или с помощью АПС за рубежом или у нас в стране. Однако наши данные, полученные при эксплуатации англо-русской системы МП АМПАР, говорят, что 12-15% всех новых для системы МП /т.е. отсутствующих в словарях системы/ английских лексических единиц являются фактически новыми терминами. Предварительные данные по эксплуатации АПС ВЦП показывают, что новые термины составляют около 25% всех отказов словаря, т.е. лексических единиц, запрошенных пользователем и отсутствующих в АПС.

Проблема новых слов /неологизмов/ привлекает в настоящее время пристальное внимание лексикографов, однако здесь, как и по ряду других вопросов, мнения лексикографов весьма существенно расходятся. "Одни понимают неологизм как стилистическую категорию, где

главное - "ощущение новизны" при восприятии слова; другие считают, что неологизмы возникают при обозначении появляющихся с развитием науки и техники новых понятий, предметов материальной культуры; третьи называют неологизмами слова, не отмеченные словарями; четвертые относят к неологизмам всякое вновь появившееся в языке слово или словосочетание"/Степанов, 1986, с. 66/.

В лексикографической работе во Всесоюзном центре переводов новыми терминами считаются термины, не зафиксированные в общедоступных дву- и многоязычных лексикографических источниках на момент их обнаружения в иноязычном тексте. Такое понимание нового термина на первый взгляд может показаться весьма упрощенным. Однако мы считаем его правомерным, поскольку подходим к этой проблеме с позиций переводчика, для которого особую трудность представляет перевод иноязычного текста, содержащего лексику, не зафиксированную в доступных ему переводных лексикографических источниках, вне зависимости от того, была ли эта лексика ранее зафиксирована в одноязычных словарях, справочниках и других публикациях или в недоступных ему переводных словарях. Проблема новой терминологической лексики заслуживает отдельного исследования, поэтому мы ограничимся здесь только перечислением основных с позиций терминологической лексикографии типов новых терминов:

1. Цельнооформленные лексические единицы /слова/, ранее не существовавшие и созданные специально для обозначения нового понятия, или предмета, либо не имевшие ранее терминологических значений, например, англ. язык: parametro - технологический макроцикл /с параметрическим заданием данных/, datuming - установка /рабочего органа станка/ в исходное положение, browser - программа ускоренного просмотра базы данных; нем. язык: Fangen - определение вызывающего абонента, Einsatz - сменный модуль, Steller - установочный элемент.

2. Новые значения для "старой" лексики /как отдельных слов, так и словосочетаний/, например, англ. язык: mezzanine - стеллажи накопителя /деталей/; candidate - вариант, beauty - достоинство, history - изменение; нем. язык: Karte - плата /ЭВМ/, Satz - кадр /станки с числовым программным управлением/; фр. язык: balayage - прогон /перемещение

курсора на экране дисплея в конец строки/, courbe - поворотная секция конвейера, pointeur - точка входа в программу.

3. Терминологические словосочетания, состоящие как из вновь созданных, так и "старых" слов. К этой же группе можно отнести словосочетания и сложные слова, например, англ. язык: idiot proof - защита от некорректных действий оператора, dirty time - служебное время /полета/, side writer - печатающее устройство с боковым перемещением каретки и бумаги; фр. язык: roue à caractères - печатающее колесо, trempé négative - высокий отпуск с последующим быстрым охлаждением, joue de touret - щека барабана; нем. язык: Bildpunkt - примитив /элемент изображения в графических схемах/, Kuchendiagramm - псевдообъемная диаграмма.

Определенную часть новых терминов составляет профессиональная ненормативная лексика, которая проникает в речь специалистов в основном из технических описаний на иностранном языке и выступает, как правило, в качестве разговорных синонимов рекомендованных терминов. В большинстве случаев такие новые термины в русском языке создаются путем транслитерации соответствующих иноязычных терминов: например,abend - абенд /аварийное завершение программы/, dump - дамповать /делать распечатку областей памяти ЭВМ/, driver - драйвер /стандартная программа связи с периферийным устройством/, cancel - канцелировать /снять задание/ и др. Традиционные иностранно-русские терминологические словари пока весьма скупо представляют этот пласт лексики. На наш взгляд, такой подход затрудняет работу с научно-технической литературой, поскольку подобные профессиональные жаргонизмы начинают все чаще проникать из устной в письменную речь технических специалистов и поэтому должны найти свое отражение в переводных словарях как стилистически сниженные синонимы стандартизованных терминов.

Как справедливо пишет А.С. Герд, "...процесс унификации и стандартизации терминов не терпит поспешности... Термины должны пройти стадию вхождения в употребление" /Герд, 1978, с. 4/. Действительно, заранее трудно сказать, какой из двух терминов - стилистически нейтральный или разговорный - войдет в употребление, а какой окажется термином-однодневкой. Примерами этого могут послужить термины "дисплей", который 10-15 лет назад рассматривался как нерекomenдованный

термин, а сейчас уже стандартизован, и "рутина" /стандартная подпрограмма/, который, несмотря на противодействие многих терминологов, широко используется в устной речи программистов, в технической документации по программированию и нашел свое отражение в переводных словарях. В нашем АПС профессионализмы и стилистически сниженные /разговорные/ термины включаются в состав словарных статей в качестве синонимов стилистически нейтральных терминов.

Основной трудностью при лексикографической обработке новых терминов является отсутствие надежных переводных эквивалентов и нередко необходимость создания их своими силами. Выход из такого положения многие лексикографы и переводчики видят в использовании развернутых переводных эквивалентов. Однако мы должны отметить, что широкое использование развернутых ПЭ в словарных статьях автоматического словаря ухудшает его достоинства как переводного словаря, а также заметно ослабляет его обратимость. По этим причинам мы стремились заменять их на краткие эквиваленты, которые помещали в зону ПЭ, а развернутый переводной эквивалент устранили либо переносили в зону толкования. Часто мы не могли найти точных и кратких ПЭ для новых терминов, и тогда в отдельных, наиболее простых случаях разработчики АПС сами создавали путем транслитерации или калькирования необходимые краткие русские переводные эквиваленты. Однако в большинстве случаев мы не брали на себя ответственность создавать и рекомендовать к употреблению новые термины. Наша осторожность в этом вопросе объясняется тем, что АПС ВЦП является в первую очередь фиксирующим, а не нормирующим словарем, и наша основная задача - найти и зафиксировать новый термин, а не породить его и рекомендовать к употреблению.

2.1.5. Индекс_надежности

В предыдущем разделе мы показали необходимость комплектования АПС как устоявшейся, так и новой лексики, не прошедшей проверки лексикографической практикой. Такой подход, с одной стороны, уменьшает количество отказов АПС при работе с текстами самых разных типов, а с другой стороны, приводит к тому, что пользователь, получая информацию, собранную из самых

разных источников, не знает степени ее надежности и вынужден решать эту задачу сам. Особенно остро эта проблема встает, когда одна лексическая единица языка оригинала получает несколько переводных эквивалентов, взятых из разных по уровню надежности источников. Решение этой проблемы смыкается с решением проблемы стандартизации терминологии – проблемой, которая привлекает в настоящее время пристальное внимание терминологов, но пока далека от окончательного решения. Здесь следует четко определить нашу позицию: при комплектовании состава нашего АПС мы не ставили перед собой задачу стандартизации терминологии. Как и в случае с порождением термина, наша задача состоит не в стандартизации термина, а в его лексикографической фиксации и доведении его до сведения пользователя. Вместе с тем мы не имеем права полностью оставить в стороне проблему стандартизации терминологии. Для того чтобы пользователь испытывал к словарю доверие, необходимо в явном виде указывать степень надежности предлагаемой ему информации. Для этой цели в нашем АПС используется индекс надежности, который приписывается каждому переводному эквиваленту и ставится непосредственно перед ним /см. Приложение, примеры 1, 2/.

В основе индекса надежности в нашем АПС лежит тип источника, откуда поступила данная словарная статья или ее часть.

Индекс надежности имеет четыре уровня:

Индекс А – означает, что данный ПЭ зафиксирован в каком-либо стандарте /ГОСТ, международный стандарт/ либо рекомендован к употреблению авторитетными международными организациями.

Индекс Б – означает, что данный ПЭ зафиксирован в авторитетных переводных словарях. Степень надежности определяется для каждого словаря отдельно и может меняться в зависимости от тематической принадлежности переводного эквивалента. Это означает, что лексика по основной тематической направленности словаря получает индекс Б, а лексика других тематических направлений, включенная в этот же словарь, получает индекс В.

Индекс В – означает, что ПЭ зафиксирован в недостаточно апробированных переводных словарях или в Тетрадах новых терминов, экспресс-выпусках новой терминологии и т.п.

Индекс Г - означает, что ПЭ ранее не был зафиксирован в лексикографических изданиях и обнаружен в научно-техническом тексте, переводе или создан специально для данной словарной статьи.

Одна и та же заглавная ЛЕ может иметь для разных значений или даже в рамках одного значения несколько ПЭ с разными индексами надежности, которые не являются раз и навсегда зафиксированными и могут, а при необходимости должны, меняться в соответствии с изменением статуса заглавной лексической единицы и ее переводных эквивалентов. Если заглавная ЛЕ или ее переводные эквиваленты зафиксированы в источниках разных категорий, то им приписывается более высокий индекс надежности. При переходе ЛЕ из одного статуса в другой, например из Тетради новых терминов в надежный словарь, а затем в стандарт, должны соответствующим образом меняться индексы надежности. Подобная "нежесткость" индекса надежности позволяет оперативно учитывать изменения в терминосистемах развивающихся отраслей науки и техники. Система стилистических помет и индекс надежности, используемые в нашем АПС, не предусматривают фиксацию устаревших и nereкомендуемых терминов. Отсутствие специального указания на термины этой категории является сознательным решением. Такая лексика не должна включаться в наш АПС, поскольку он предназначен на данном этапе своего создания для перевода и чтения современной научно-технической литературы, а не для исследования изменения терминосистемы во времени. Критерием невключения или изъятия лексической единицы из состава словаря является ее отсутствие в письменных текстах или устной речи специалистов по данной тематике.

Индекс надежности, применяемый нами в настоящее время, является однокомпонентным и приписывается переводному эквиваленту. Однако фактически он характеризует не только надежность ПЭ, но и надежность всей триады: "лексическая единица 'Языка 1 - отношение переводной эквивалентности - лексическая единица Языка 2", т.е. надежность двуязычной словарной статьи целиком или ее части. Использование однокомпонентного индекса надежности на основе типа лексикографического или другого источника является весьма удобным практическим приемом фиксации уровня нормированности терминологической лексики в автоматических переводных словарях. Вместе с тем однокомпонентный индекс надежности

"надежно работает" только при переносе словарной статьи целиком или ее основной части /заглавная единица - переводной эквивалент/ из одного лексикографического источника в АПС. Например, когда в АПС включается вся словарная статья из авторитетного традиционного переводного словаря, то все ее ПЭ получают индекс Б. Когда для этой же заглавной ЛЕ обнаруживаются новые ПЭ в словарях другого уровня или же не зафиксированные ранее, то они включаются в словарную статью с индексом В или Г, соответственно. Сложнее обстоит дело, когда словарная статья не заимствуется целиком или по частям, а создается заново для автоматического словаря. В этом случае однокомпонентный индекс надежности оказывается недостаточно удобным инструментом.

При создании словарной статьи АПС теоретически возможны варианты, когда заглавная лексическая единица взята из одноязычного стандарта языка оригинала, переводной эквивалент взят из одноязычного стандарта языка перевода, но ни один лексикографический источник не зафиксировал их в качестве переводных эквивалентов друг друга, а это сделал составитель данной словарной статьи. Для этого случая целесообразно ввести трехкомпонентный индекс надежности, с помощью которого можно будет фиксировать уровень надежности каждого из компонентов триады независимо друг от друга. Например, для приведенного выше примера должен использоваться индекс АГА, который означает: заглавная ЛЕ зафиксирована в стандарте языка оригинала, отношение переводной эквивалентности установлено составителем данной словарной статьи, переводной эквивалент зафиксирован в стандарте языка перевода. Индекс ГГГ будет означать, что словарная статья во всех ее компонентах построена на основе нелексикографических источников. Индекс БГБ охватывает случаи, когда двуязычная словарная статья создается на основе авторитетных одноязычных словарей языка оригинала и языка перевода. Трехуровневый индекс надежности позволяет точнее фиксировать сходства и различия в уровне нормированности всех трех компонентов триады "заглавная единица - отношение переводной эквивалентности - переводной эквивалент".

Использование трехкомпонентного индекса надежности позволит четко ограничить дву- и многоязычные терминологические стандарты от дву- и многоязычных

словарных статей, заимствованных из различных лексикографических источников или же созданных разработчиками АПС на базе одноязычного исходного материала. Такое решение делает АПС более приспособленным для стандартизации терминологии.

2.1.6. Требование эквивалентности разноязычных лексических массивов в многоязычном автоматическом словаре

АПС ВЦП является многоязычным и обратимым. Эти две важные характеристики словаря оказывают существенное влияние на состав его лексических массивов. В разд. 1.7 гл. 1 сформулировано понятие лексической обратимости АПС, которое достигается при эквивалентности лексических массивов соответствующих языков.

Лексические массивы двух языков в АПС будем считать эквивалентными, если каждая лексическая единица Языка 1 имеет переводной эквивалент в лексическом массиве Языка 2 и наоборот. Лексические массивы, для которых такое требование не выполняется, будем считать неэквивалентными. В упрощенном виде эту характеристику разноязычных лексических массивов можно показать на схемах эквивалентности-неэквивалентности лексических массивов разных языков в многоязычном автоматическом словаре:

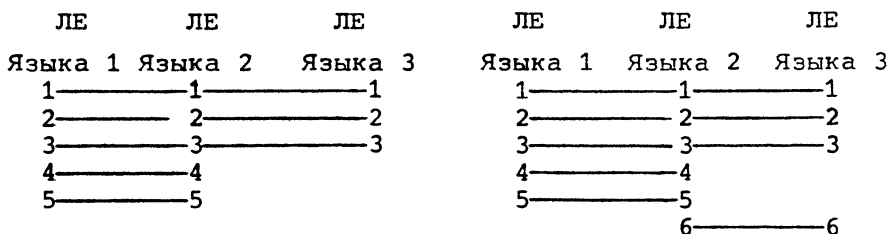


Схема 1

Схема 2

В схемах 1 и 2 рассматриваются гипотетические случаи, когда все ЛЕ всех трех языков однозначны. На схеме 1 показан случай эквивалентности и равенства по объему лексических массивов Языков 1 и 2. Лексический массив Языка 3 меньше по объему и неэквивалентен лексическим массивам Языков 1 и 2. Схема 2 показывает более обобщенный и реальный случай, когда лексические массивы всех трех языков не равны по объемам и не

эквивалентны. Множество лексических единиц Языка 2, являющегося в данной схеме стержневым языком /языком-посредником/, равно объединению множеств лексических единиц Языков 1 и 3. В рассмотренных случаях сознательно не учитывается фактор многозначности лексических единиц. Автор отдает себе отчет, что в принципе возможно создание разных по объему, но эквивалентных лексических массивов, эквивалентность которых будет достигаться за счет более высокого уровня многозначности ЛЕ одного из языков. Однако такой случай может возникнуть только на лексических массивах малых объемов, а также при сознательно практикуемом различии в глубине описания многозначной лексики разных языков. При соблюдении единого подхода к глубине описания многозначной лексики основным фактором, определяющим эквивалентность лексических массивов в крупном многоязычном словаре, является их объем.

С ростом объема словаря влияние фактора многозначности лексических единиц на эквивалентность лексических массивов уменьшается, вместе с тем одновременно становится значительно сложнее достичь полной эквивалентности лексических массивов всех языков. Этот факт достаточно четко просматривается на примере нашего многоязычного словаря. В настоящее время лексические массивы по вычислительной технике в словаре достигли следующих объемов: массив английских ЛЕ - 67 тыс. единиц; массив французских ЛЕ - 35,5 тыс. единиц; массив немецких ЛЕ - 42 тыс. единиц. Массив русских лексических единиц является объединением переводных эквивалентов всех иноязычных входных ЛЕ. Его объем в настоящее время пока не подсчитан. Уже из простого сопоставления объемов ясно, что разноязычные лексические массивы неэквивалентны. Неэквивалентность разноязычных лексических объемов в АПС ВЦП объясняется целым рядом причин, основными из которых являются сознательное выделение англо-русской версии АПС как основной, а также характер источников пополнения АПС. Концентрация усилий разработчиков на англо-русской версии словаря и преимущественное наращивание его объемов объясняются тем, что переводы с английского языка и на английский составляют около 70% всех переводов, выполняемых в нашей организации.

Основными источниками пополнения АПС являются различные двуязычные иностранно-русские словари, Тетради новых терминов и другие глоссарии по одной и той же или близким тематикам. Несмотря на совпадение

или близость тематической ориентации этих источников, они весьма существенно различаются как по составу входных лексических единиц, так и по составу их русских переводных эквивалентов. Это в свою очередь приводит к тому, что, во-первых, непропорционально растет объем русских ЛЕ, а во-вторых, при объединении нескольких иностранно-русских лексикографических источников с разными входными языками в одном многоязычном словаре значительное количество русских ЛЕ не получают полного набора переводных эквивалентов на всех иностранных языках.

Такое положение не является нормальным, и необходимо проведение особого цикла работ для достижения полной эквивалентности либо уменьшения степени неэквивалентности разноязычных лексических массивов. В АПС ВЦП это решается следующим образом. В АПС вводится определенный массив англо-русских, немецко-русских и французско-русских словарных статей по одной тематике, на основе которых формируется информационная база АПС. После этого для этих массивов создается русско-иностранный многоязычный массив. В силу несогласованности состава исходных двуязычных словарей определенное количество многоязычных русско-иностранных словарных статей будут иметь неполный состав иноязычных переводных эквивалентов. Многоязычный русско-иностранный лексический массив распечатывается таким образом, чтобы для каждой русской ЛЕ было четко видно, из какого двуязычного массива она поступила /см. Приложение, пример № 3/.

Такая многоязычная русско-иностранная распечатка является исходной точкой индивидуального пополнения АПС, основная цель которого - достижение эквивалентности разноязычных лексических массивов. Суть этой работы сводится к тому, что если у некоторой русской ЛЕ отсутствуют переводные эквиваленты одного или двух языков, лексикограф должен найти эти эквиваленты и составить иностранно-русские словарные статьи, в которых заглавной ЛЕ является данная иноязычная ЛЕ, а в качестве ее переводного эквивалента выступает исходная русская лексическая единица. Например, из приведенных в Приложении многоязычных словарных статей видно, что русское слово "барьер" не имеет в нашем АПС английских и немецких переводных эквивалентов, а слово "база" не имеет немецкого эквивалента. Если лексикографические источники, на основании которых

составляется необходимая иностранно-русская словарная статья, дают заглавной ЛЕ в рамках одного и того же значения несколько русских ПЭ помимо того, который указан в многоязычной русско-иностранной распечатке, необходимо проверить, входят ли эти русские ЛЕ в эту распечатку или в состав всего АПС. Если да, то их можно оставить в создаваемой иностранно-русской словарной статье, а всю словарную статью можно вводить в АПС. Если же эти русские ЛЕ не зафиксированы в многоязычной распечатке и не были ранее введены в АПС, то на данном этапе их не следует включать в создаваемую иностранно-русскую словарную статью, так как в противном случае это приведет к дальнейшему росту объема русского лексического массива и, как следствие, к увеличению неэквивалентности между русским массивом и массивами двух других иностранных языков. После того как по этой технологии на основе данной многоязычной русско-иностранной распечатки будут составлены и введены в АПС все необходимые иностранно-русские словарные статьи, можно будет сказать, что на данном объеме АПС достигнута эквивалентность всех разноязычных лексических массивов.

Возможен и другой путь достижения эквивалентности лексических массивов, при котором лексикограф сохраняет все новые русские ЛЕ, появившиеся в качестве переводных эквивалентов при составлении иностранно-русской словарной статьи, например англо-русской, но он должен помнить, что этим он делает необходимым создание дополнительных немецко-русских и французско-русских словарных статей, в которых эти новые русские ЛЕ должны выступать в качестве переводных эквивалентов. На этом этапе ситуация может вновь повториться, и тем самым такой путь к достижению полной эквивалентности разноязычных лексических массивов во многоязычном словаре становится практически бесконечным. Ситуация еще больше осложняется из-за трудноподдающейся описанию системы взаимоотношений между многозначными разноязычными лексическими единицами в таком словаре.

Эти же русско-иностранные многоязычные словарные статьи используются для контроля за единообразием при определении лексикограмматических классов /ЛГК/ русских лексических единиц, используемых в качестве переводных эквивалентов при составлении иностранно-русских словарных статей. Расхождения в определении

ЛГК русских переводных эквивалентов возникают из-за того, что англо-русские, немецко-русские и французско-русские словарные статьи составляются разными людьми в разное время и в разных местах. Например, из словарной статьи "действие" / см. Приложение, пример № 3/ видно, что в одном случае слово "действие" получило ЛГК 0004, а в другом случае - 0013. Все лексические массивы поступают в ВЦП на редактирование, где эти и другие ошибки должны быть исправлены.

Таким образом, стремление к достижению полной эквивалентности всех лексических массивов в нашем АПС требует пополнения и доведения всех иноязычных лексических массивов по их составу и объему до уровня лексического массива русского языка и тем самым серьезным образом влияет на отбор лексики в АПС.

2.1.7. Динамичность состава словаря

Как и все системы машинного перевода и автоматические словари, наш АПС обладает свойством динамичности. Разработчик словаря постоянно имеет возможность осуществлять коррекцию лексических составов АПС как в сторону пополнения, так и в сторону устранения любых лексических единиц. Строго говоря, подобное свойство АПС является в первую очередь порождением достоинств его программного обеспечения. АПС ВЦП имеет ряд сервисных программ, которые позволяют накапливать данные по количеству обращений к каждой лексической единице, а также собирать "отказы", т.е. лексику, запрошенную пользователем, но отсутствующую в АПС, с указанием числа запросов к каждой отсутствующей лексической единице. Накопленные данные могут затем использоваться для коррекции состава АПС. Коррекции могут подвергаться как отдельные лексические единицы, так и целые массивы лексических единиц.

2.2. Структура словарной статьи АПС ВЦП

Структура, состав и объем информации в словарной статье АПС ВЦП определяются его основной целевой направленностью /повышение производительности человека при переводе или чтении иноязычной научно-

технической литературы/, свойствами и требованиями основных категорий пользователей /переводчиков, редакторов и технических специалистов/ и возможностями разработчиков АПС. Целевая направленность на перевод и чтение иноязычной литературы требуют включения переводных эквивалентов и информации, определяющей выбор нужного эквивалента из нескольких предложенных. Ориентирование на пользователей с разным уровнем знания входного языка требует включения в словарную статью грамматической информации, а также специальной информации, необходимой для работы различных лингвистических алгоритмов. Мы предполагаем, что АПС может представлять интерес для лексикографов и терминологов, поэтому включили в словарную статью специальные зоны для такой информации, как контекст, лексическая сочетаемость, источник поступления, дата составления и фамилия составителя или редактора словарной статьи.

Ограниченные возможности разработчиков АПС в значительной степени определяют стратегию выбора объема и формы представления некоторых видов информации. Так, например, набор переводных эквивалентов, пометы, толкования, разграничение значений многозначных слов и их ранжирование в исходных словарных статьях не создаются самими разработчиками, а заимствуются из опубликованных словарей. Поскольку базовая словарная статья составляется автоматически на основе исходных словарных статей, то ограничения на объем и форму представления информации распространяются и на базовую словарную статью АПС. Кроме того, традиционные словари, к сожалению, не имеют единых и полностью совместимых принципов описания ЛЕ, и поэтому в базовой словарной статье у одной и той же ЛЕ встречаются различные варианты ранжирования ее значений или самая разная описательная информация к ним. Такие словарные статьи необходимо перед включением в АПС передавать на переработку лексикографам.

Ориентирование на диалог с человеком предопределяет и форму представления информации в словарной статье, которая в целом имеет вид, приближенный к словарным статьям традиционных словарей. Однако на организацию словарной статьи АПС и всего АПС влияет ЭВМ как его материальный носитель. На уровне структуры словарной статьи это проявляется в первую очередь в ее жесткой зональной расчлененности.

Словарная статья АПС описывает отношения между иноязычными и русскими ЛЕ. Базу словаря составляют

массивы англо-русских, французско-русских и немецко-русских словарных статей. Все остальные варианты двуязычных словарей формируются автоматически, а затем корректируются вручную.

Исходная словарная статья автоматического словаря разделена на 13 зон. Каждая зона содержит строго определенную информацию и имеет свои правила заполнения.

Зона 1. Зона заголовка. Ключевой зоной словарной статьи, по которой осуществляется вход в словарную статью, производится ее формирование и поиск, является зона заголовка. В качестве такого может выступать морфема, слово, словосочетание в своей словарной /канонической/ форме или фрагмент текста. Требование записи в словарной форме на фрагменты текста не распространяется. Для описания слов-омонимов создается несколько словарных статей. В этом случае заголовок словарной статьи сопровождается порядковым номером /см. Приложение, пример № 4/. Здесь имеются в виду лишь полные омонимы. Если же совпадают словарные формы разных частей речи, то функцию различения соответствующих словарных статей выполняет индекс лексикограмматического класса.

Зона 2. Зона лексико-грамматического класса заголовка. За заголовком следует индекс его ЛГК. Цельнооформленные лексические единицы подразделяются на группы в зависимости от своей частеречной принадлежности, а внутри этих групп классифицируются по ряду категориальных признаков. Словосочетания делятся на типы в зависимости от функции, выполняемой ими в предложении.

Каждый из рабочих языков словаря имеет свою систему ЛГК, созданную с учетом его грамматического строя. Названия ЛГК при составлении словарной статьи записываются в цифровых кодах, а при выдаче ответа пользователю даются в стандартных сокращениях, принятых в традиционной лексикографии. Исключение составляет расшифровка кодов именных и глагольных словосочетаний, где помимо типа словосочетания указывается и лексико-грамматический класс основного слова. Лексико-грамматическая информация помогает построить грамматически правильную фразу с заглавной лексической единицей. Кроме того, зона 2 служит для различения лексико-грамматических омонимов, имеющих одинаковую словарную форму, но относящихся к различным частям речи. /см. Приложение, пример № 5/.

Зона 2а. Зона лексико-грамматического класса русского переводного эквивалента. Непосредственно за русским ПЭ записывается индекс его лексико-грамматического класса. По требованию пользователя при выдаче словарной статьи эта информация может не выдаваться, но само ее написание в исходной словарной статье необходимо для "перевертывания" словарной статьи, т.е. автоматического формирования русско-иностранной словарной статьи.

Зона 3. Зона морфограмматической информации /МГИ/. Приводимая в словарной статье морфограмматическая информация служит для построения всей словоизменительной парадигмы заглавной единицы и, таким образом, позволяет осуществлять поиск любой текстовой формы слова или словосочетания. Правила заполнения этой зоны зависят от выбранного для данного языка пути сведения текстовых форм ЛЕ к их словарным формам /подробнее об этом см. разд. 2.4.3/.

Если сведение текстовых форм к словарным осуществляется методом поискового массива, то в этой же зоне записываются все словоформы заголовка, не совпадающие по написанию с самим заголовком. Они включаются в поисковый массив, т.е. по ним может вестись поиск словарной статьи. /см. Приложение, пример № 6/.

Если же используется алгоритмический метод, то в зоне морфограмматической информации для французских ЛЕ записываются их машинные основы с указанием их порядковых номеров и номеров словоизменительных классов, а для русских слов составляются отдельные лемматизационные статьи. Описанные выше правила заполнения зоны МГИ распространяются на словарные статьи, где в качестве заголовка выступает отдельное слово. Если же в качестве заголовка выступает словосочетание, то зона МГИ не заполняется, поскольку для приведения текстовых форм словосочетаний к словарному виду используется алгоритмический метод. Использование алгоритма лемматизации словосочетаний дает значительный выигрыш памяти, так как нет необходимости хранить в памяти ЭВМ все возможные формы русских и иноязычных многокомпонентных лексических единиц.

Внутри зоны МГИ предусмотрена подзона сокращений, куда записываются все варианты сокращений заглавной ЛЕ как для отдельных слов, так и для словосочетаний.

Зона МГИ не распечатывается при выдаче ответа пользователю, чтобы не перегружать его практически ненужной ему информацией. Исключение составляют сокращения, которые выдаются пользователю при любом режиме ответа: при запросе полной лексической единицы пользователь без специального требования получает и ее сокращенный вариант, а при запросе сокращения - и его расшифровку.

Сокращения, занесенные в зону МГИ, не выступают в качестве самостоятельных словарных статей, поскольку они включаются в поисковый массив и по ним может осуществляться поиск словарной статьи. Сокращение может выступать в роли заглавной ЛЕ лишь в том случае, если неизвестна его расшифровка.

Зона 3а. Зона сокращений в русском переводном эквиваленте /РПЭ/. В словарных статьях, где русские переводные эквиваленты имеют сокращения, введена зона 3а, которая следует сразу за зоной 2а /зоной лексико-грамматического класса русского ПЭ/. Она необходима для выполнения следующих функций: 1/ сохранения общей структуры словарной статьи при автоматическом создании русско-иностранных словарных статей и 2/ для обеспечения выдачи пользователю полной лексической единицы и ее сокращений.

Зона 4. Зона тематической рубрики, подрубрики и стиля. В этой зоне приводятся тематические и стилистические пометы, уточняющие область и условия употребления заглавной ЛЕ в значении, соответствующем конкретному русскому ПЭ. Стилистические пометы в основном соответствуют принятым в традиционных словарях, например: разг. - разговорный вариант, проф. - профессионализм. Все три вида информации записываются в цифровых кодах. Тематические рубрики, тематические подрубрики и стилистические пометы формируют три различных списка, открытых для пополнения. При выдаче ответа пользователю цифровые коды преобразуются в стандартные буквенные сокращения и записываются перед индексом надежности. /См. Приложение, пример № 7/.

В зоне 4 в подзоне стиля может быть дано указание на язык-источник для заимствованной лексики. В этом случае при выдаче на печать эта информация ставится перед переводным эквивалентом, что соответствует лексикографической традиции. Если в этой подзоне дано указание на стиливые особенности, то при выдаче на печать эта информация также записывается перед переводным эквивалентом /см. Приложение, примеры №8,7/.


Тематические и стилистические пометы являются параметрами, по которым может осуществляться поиск. Так, из словарной статьи запрашиваемого слова по специальному требованию пользователя могут быть извлечены и выданы на печать переводные эквиваленты, относящиеся лишь к определенной тематике или имеющие определенные стилистические пометы. На основе этих помет могут быть сформированы монотематические словари или списки слов, имеющих определенные стилистические особенности.

Зона 5. Зона индекса надежности. Каждому русскому переводному эквиваленту предшествует индекс надежности, отражающий степень "общепринятости" данного переводного эквивалента. Этот индекс определяется надежностью лексикографического источника, из которого он взят /см. Приложение, примеры № 1, 2/.

Зона 6. Зона русских переводных эквивалентов Русские переводные эквиваленты представляют собой отдельные слова и словосочетания, записанные в канонической форме. Русских переводных эквивалентов может быть один или несколько. Все РПЭ снабжаются порядковыми номерами. Синонимичные РПЭ имеют одинаковые порядковые номера, а РГЭ, выражающие разные значения заглавной ЛЕ, - разные номера, соответственно. Порядок следования русских переводных эквивалентов обуславливается как лексикографическими традициями, так и рядом особенностей данного АПС /его ориентированностью на научно-техническую лексику, наличием различных режимов выдачи ответа и т.п./.

Для русских глаголов, образующих совершенный вид бесприставочным способом, в зоне РПЭ приводятся оба вида /см. Приложение, пример № 9/. Это позволяет пользователю со слабым знанием русского языка запрашивать русские глаголы совершенного вида в текстовой форме, не приводя их к традиционной словарной форме - к несовершенному виду. Если совершенный вид русского глагола образуется приставочным способом, то при его запросе словообразовательная приставка будет отделена автоматически; поэтому для таких русских глаголов в качестве переводного эквивалента записывается только форма несовершенного вида.

В традиционной лексикографии существует понятие безэквивалентной лексики. Сюда относятся лексические единицы, не имеющие перевода на другой язык. Это или общеупотребительная лексика, обозначающая понятия, отсутствующие в других странах /например, "рушник",

"варенец" и т.п./, или новые термины, для которых еще нет соответствующего термина в другом языке. В АПС ВЦП относительно часто встречается безэквивалентная лексика второго типа. В словарных статьях таких лексических единиц в зоне РПЭ ставится знак /о  /, указывающий на отсутствие переводного эквивалента, но обязательно заполняется зона толкований, где дается по возможности полное и точное описание значения, выражаемое данной ЛЕ. /см. Приложение, пример № 10/. Если со временем новому термину будет найден соответствующий переводной эквивалент, его необходимо внести в зону РПЭ, и тогда словарная статья будет иметь стандартный вид.

Зона 7. Зона пояснительных помет. В словарных статьях нашего АПС, так же как и в традиционных словарях, пояснительные пометы служат для уточнения значений лексических единиц и описания правил их употребления. Эта информация необходима для правильного выбора переводного эквивалента. Пояснительные пометы могут быть приписаны русским ПЭ как однозначных, так и многозначных слов и словосочетаний. По своему содержанию их можно условно разделить на семантические, лексические и синтаксические.

Семантические пометы описывают семантическую сочетаемость РПЭ в терминах семантических требований, предъявляемых данной лексической единицей к своему потенциальному главному актанту, т.е. указывается, представители каких семантических классов могут заполнять позицию главного актанта при соответствующем русском ПЭ. /См. Приложение, пример № 11/. Эта же зона используется для фиксации семантического класса заглавной ЛЕ и ее переводных эквивалентов /см. Приложение, пример № 12/.

Лексические пометы описывают лексическую сочетаемость русских ПЭ в терминах конкретных лексем /см. Приложение, пример № 13/. Синтаксические пометы фиксируют элементы синтаксической модели управления русских ПЭ в терминах синтаксических позиций и особенностей грамматического поведения заглавной ЛЕ в данном значении /см. Приложение, пример № 14/.

На данном этапе создания АПС ВЦП пояснительные пометы записываются на русском языке, который выбран в АПС ВЦП в качестве метаязыка описания. Мы понимаем, что такое решение не является оптимальным для нашего переводного словаря в первую очередь из-за его обратимости и многоязычия. Так, при автоматическом формиро-

вании русско-иностранной словарной статьи семантические пометы сохраняют свою привязку к переводному эквиваленту, в роли которого в этом случае уже выступает иностранная ЛЕ. Еще более серьезные трудности возникают при генерации иностранно-иностраных версий АПС, например, в англо-немецком варианте, когда семантический класс заглавной английской ЛЕ или ее немецкого ПЭ фиксируется с помощью русских лексем. В дальнейшем мы планируем записывать пояснительные пометы, относящиеся к заглавной ЛЕ на языке оригинала, относящиеся к ее переводным эквивалентам - на языке перевода.

Зона 8. Зона толкований. Для описания значений терминологических единиц могут быть использованы толкования. Толкования не обязательно имеют строгую логическую структуру и могут представлять собой некоторое пояснение в произвольной форме. Прежде всего это относится к новым терминам. Иногда толкование используется для разрешения многозначности заглавной ЛЕ, особенно в случаях, когда русский ПЭ является также многозначным /см. Приложение, пример № 15/. В настоящее время толкование записывается на русском языке. Мы полагаем, что в переводном словаре толкование значения заглавной ЛЕ целесообразно фиксировать на выходном языке. Обратимость нашего АПС ставит перед нами задачу записи толкования заглавной ЛЕ также на входном языке.

Зона 9. Зона примеров употребления. Русские переводные эквиваленты могут сопровождаться примерами употребления заглавной лексической единицы в соответствующих значениях. Зона 9 выполняет две функции: иллюстративную и смысловозначительную. Однозначные лексические единицы с индексами надежности А и Б, как правило, даются без примеров. Для лексики с индексом надежности В примеры желательны, а для ЛЕ с индексом надежности Г - обязательны. В словарной статье многозначного слова примеры могут выполнять смысловозначительную функцию, способствуя разграничению разных значений заглавной ЛЕ или же выбору нужного ПЭ в пределах одного значения. В качестве примеров могут использоваться не только законченные предложения, но и отдельные словосочетания. Примеры даются на входном языке и в ряде случаев сопровождаются русским переводом /см. Приложение, пример № 1/. Примеры-словосочетания обычно заимствуются из тра-

диционных словарей, примеры-предложения подбираются на основе конкордансов, составленных по текстам соответствующей тематики.

Зона 10. Зона фразеологии. Зона состоит из двух подзон:

10а/ подзона, в которой записывается информация, характеризующая заглавную ЛЕ с точки зрения ее положения в лексической системе языка. Нередко встречаются ЛЕ, имеющие несколько вариантов написания. В этом случае составляется столько словарных статей, сколько существует вариантов написания, при этом в каждой словарной статье в подзоне 10а в качестве синонимов перечисляются все остальные варианты. Эта зона открыта для пополнения любой другой информацией, которая может быть полезна для пользователя, но не учтена в общей структуре словарной статьи /следует, однако, иметь в виду, что эта информация должна относиться только к данной заглавной лексической единице, и поэтому при "перевертывании" словарной статьи она не выдается/.

10б/ В отличие от подзоны 10а, эта подзона составляется автоматически по запросу пользователя и представляет собой перечень всех словосочетаний из АПС, в которых содержится заглавная ЛЕ /только отдельное слово/ в любой текстовой форме /см. Приложение, пример № 16/.

Зона 11. Имя составителя словарной статьи. Зона 13. Дата составления словарной статьи. Базовая словарная статья формируется автоматически из введенных из разных источников исходных словарных статей, имеющих одинаковую заглавную ЛЕ. Такая базовая словарная статья представляет собой неупорядоченный перечень переводных эквивалентов, поэтому специалист-лексикограф должен ее отредактировать. После этого в зоне 11 он ставит свою фамилию, а в зоне 13 дату редактирования.

Зона 12. Название источника. В базовой словарной статье каждому переводному эквиваленту указывается номер источника или источников, откуда он заимствован /см. Приложение, примеры № 13-15 и др./.

2.3. Описание многозначных лексических единиц

2.3.1. Принципы разграничения многозначности

Проблема лексической многозначности является одной из основных в традиционной двуязычной лексикографии. Не снимается она и для многоязычных автоматических словарей. Более того, многоязычие и обратимость автоматических словарей существенно усложняют эту проблему. Разрабатывая наш АПС, мы исходили из того, что, несмотря на всю свою специфичность, автоматические переводные словари в первую очередь все же остаются переводными словарями, и поэтому основные положения теории традиционной переводной лексикографии распространяются и на них.

В традиционной переводной лексикографии существуют два основных подхода к решению проблемы многозначности лексических единиц входного языка. Первый из этих подходов заключается в том, что описание смысловой структуры лексики входного языка не должно зависеть от законов выходного языка. Смысловая структура слова должна быть в принципе одинаковой во всех двуязычных словарях с данным языком и должна совпадать со смысловой структурой этого слова в одноязычных толковых словарях /Фельдман, 1957а, с. 14/.

Второй подход состоит в том, что разграничение значений лексики входного языка в значительной степени зависит от лексики выходного языка. "Слово имеет столько переводных значений, скольким несинонимическим эквивалентам в переводящем языке оно соответствует" /Durovič, 1961, с. 83; цит. по Берков, 1971, с. 417/. Доводы в пользу такого решения сводятся прежде всего к тому, что двуязычный словарь - это в первую очередь практическое пособие для перевода и его основной задачей является показ эквивалентов слов входного языка в лексике выходного языка. Такой подход не требует разработки смысловой структуры лексики входного языка.

Мы не разделяем точку зрения ученых второго направления и полагаем, что, несмотря на определенные практические удобства и простоту решений при создании словарных статей многозначной лексики, подобный подход ведет к искажению объективной картины действительности. При этом подходе одно и то же слово при одинаковой глубине описания может фактически получить несколько разных смысловых структур, что, по нашему

мнению, неверно. Смысловая структура слова любого языка есть факт этого же языка, и от других языков она не зависит. Это весьма четко просматривается при построении многоязычных словарей /как традиционных, так и автоматических/. Если мы попытаемся руководствоваться вторым подходом и будем описывать смысловую структуру заглавного слова в зависимости от количества его переводных эквивалентов в выходном языке, мы потеряем возможность получить объективную картину, так как выходные языки, а их несколько, будут навязывать свои, часто несовместимые требования. Например, если мы рассмотрим весь аспект значений слов "бежать", "run", "courir", которые традиционно считаются прямыми переводными эквивалентами друг друга, то обнаружим, что между этими словами различий не меньше, чем сходств. Так, в слове "бежать" выделяется в разных словарях 5-6 значений с английскими ПЭ run, hurry, escape, be fast, avoid; в слове run выделяется 12-15 значений с русскими ПЭ "бежать", "работать", "функционировать", "эксплуатировать", "управлять", "течь", "курсировать", "спасаться", "гласить", "действовать" /быть действенным/, "помещать" /в газете/, "баллотироваться" и др.; в слове courir выделяется 6-8 значений с русскими ПЭ "бежать", "течь", "длиться", "продолжаться", "курсировать", "преследовать" и др. Как видно из этого примера, в каждом из этих слов есть значения, отсутствующие в других, отсюда и несовпадения в наборах переводных эквивалентов в других языках. Мы полагаем, что объективной картины можно добиться только при независимом описании, когда лексика любого языка описывается только в пределах и по законам этого же языка. В нашем АПС мы стремимся руководствоваться этими принципами при разрешении многозначности единиц всех уровней.

Вместе с тем такой подход сам по себе не снимает различий в семантических структурах ЛЕ, которые являются эквивалентами друг друга в разных языках. "Однако несовпадение системы значений слов двух языков, одинаковых по одному своему значению, само по себе не создает трудности для перевода их постольку, поскольку каждое значение слова одного языка может быть по отдельности выражено словом другого языка" /Фельдман, 1957б, с.90/. Для того чтобы преодолеть трудности, возникающие из-за несоответствия систем значений многозначных слов разных языков, в

нашем АПС используется метод раздельного хранения значений многозначных ЛЕ, когда словарная статья многозначной ЛЕ не хранится в АПС целиком, а расчленяется на отдельные значения, каждое из которых имеет свое описание, т.е. свои переводные эквиваленты, пояснительные и тематические пометы, индекс надежности, толкование, примеры употребления и т.п. Межъязыковые соответствия, таким образом, устанавливаются не между целыми многозначными ЛЕ, например "бежать", run, courir, а между отдельными значениями этих слов. Теоретически эти разноязычные словозначения также не идентичны, но различия между ними невелики, и на практике их можно не учитывать. Этой особенности переводной лексикографии есть параллель в лексикографии одноязычной, которая заключается в том, что стремление к максимальной точности в описании семантики отдельного слова ведет к искажениям в системе в целом. "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" /Караулов, 1981/.

Полная словарная статья многозначного слова в нашем АПС синтезируется из своих отдельно хранящихся значений, каждое из которых может иметь в качестве переводного эквивалента отдельные значения как из состава многозначных слов, которые традиционно считаются их переводными эквивалентами в других языках, так и из состава других лексических единиц. Такое решение позволяет сохранить принцип обратимости словаря в двуязычном и многоязычном вариантах АПС. Так, при запросе слова "управлять" пользователь получит среди других английских эквивалентов слово run, но не в полном объеме, а только в соответствующем значении: run an office /управлять учреждением/.

2.3.2. Средства разграничения многозначности и омонимии в автоматическом словаре ВЦП

Создавая массив многозначной лексики в нашем АПС, мы исходили из того, что наиболее точно смысловая структура многозначных слов описывается в одноязычных толковых словарях. Однако на практике в ус-

ловиях весьма жестких временных ограничений чрезвычайно трудно строить словарные статьи, заимствуя смысловую структуру слова из одноязычного толкового словаря, а его иноязычные эквиваленты из соответствующих двуязычных словарей. Как правило, мы брали за основу словарную статью многозначного слова из надежного двуязычного словаря и вкладывали ее содержание в структуру нашей словарной статьи, сохраняя при этом в большинстве случаев без изменений его смысловую структуру. С формальной точки зрения омонимия отличается в нашем словаре от полисемии тем, что разным омонимам в словаре соответствуют разные словарные статьи, а разные значения одного многозначного слова объединяются в одну статью. Критерии разграничения омонимии и полисемии в нашем словаре соответствуют критериям, принятым в традиционной лексикографии.

Структура словарной статьи многозначного слова аналогична структуре словарных статей однозначных слов, однако по составу информации в словарной статье и ее объему многозначная лексика, как и в традиционных словарях, заметно превосходит однозначную. Для разграничения омонимии и полисемии в нашем АПС используются следующие зоны словарной статьи: зона 4 - тематическая рубрика и подрубрика, зона 6 - переводные эквиваленты, зона 7 - пояснительные пометы, зона 8 - толкование, зона 9 - примеры употребления.

Т е м а т и ч е с к а я р у б р и к а

Указание на тематическую принадлежность является весьма распространенным и, пожалуй, самым экономным средством решения проблемы многозначности в автоматических словарях и терминологических банках. В ряде случаев действительно достаточно указать тематическую принадлежность, чтобы разграничить значения многозначного слова. В терминологическом банке TERMIUM в основном для разграничения омонимии и полисемии разработана специальная и весьма подробная система кодирования областей науки и техники, состоящая из 26³ классов, разделов и подразделов /Dubuc, 1972/. Такой подход оказался достаточно надежным приемом решения проблемы лексической многозначности в банке, однако разработчики банка не ограничиваются только им. В нашем АПС для этих целей также разработана специальная система тематического кодирования лексики на основе Рубрикатора ГАСНТИ /см. разд. 2.1.3./.

П е р е в о д н о й э к в и в а л е н т

Однако указания на тематическую принадлежность в большинстве случаев оказывается недостаточно для разрешения многозначности. В частности, этот прием "не работает" для общеупотребительной лексики, а также для нетерминологической части значения ЛЕ, совмещающих в себе терминологические и нетерминологические значения. За зоной тематической принадлежности в нашем АПС следует зона ПЭ. Переводной эквивалент — это основное средство семантизации лексики входного языка в дву- и многозначных словарях. С помощью переводного эквивалента можно наиболее экономно передать значение заглавной ЛЕ. Мы отмечали, что некоторые ученые рассматривают переводной эквивалент как основное средство разграничения многозначности входной лексики. В нашем АПС мы не используем переводной эквивалент в этих целях. Вместе с тем следует сказать, что правильно подобранные ПЭ косвенным образом отражают смысловую структуру заглавного слова. Мы полагаем, что этим мы не противоречим тезису о независимом внутриязыковом разграничении многозначности слова. Переводной эквивалент не есть средство разграничения значений, а лишь средство его фиксации. Надо сказать, что в словарях экстенсивного типа, в которых, как правило, нет ни толкований, ни смыслоразличительных помет, это чуть ли не единственное средство показа смысловой структуры заглавного слова. Разные значения в словарной статье нашего словаря нумеруются арабскими цифрами 1,2,3... В пределах одного и того же значения разные переводные эквиваленты получают один и тот же номер /см. Приложение, примеры № 1, 10, 13/.

П о я с н и т е л ь н ы е п о м е т ы

В ряде случаев переводные эквиваленты, раскрывающие значение иноязычного слова, не определяют точно круга его употребления и не позволяют полностью и с неоспоримой ясностью разграничить значения заглавного слова. В таких случаях необходимы дополнительные пояснения, пометы и иллюстрации /Фельдман, 1957б, с.84-85/. Для этой цели в структуре словарной статьи нашего АС предусмотрены зоны пояснительных помет и контекстных примеров. Зону пояснительных помет мы рас-

смаатриваем как основное средство для разрешения многозначности заглавного слова и для уточнения сферы его употребления. В этой зоне может фиксироваться любая информация, которая позволяет разграничить значения полисемантического слова.

Мы отказались от идеи использовать для этих целей или для записи толкований некий специальный формальный язык с применением глубинного описания лексической семантики. Мы исходили при этом из того, что существующие сегодня формальные языки такого типа, во-первых, не позволяют однозначно и экономно решать проблемы многозначности в естественных языках в основном из-за своей недостаточной описательной мощности. С другой стороны, стремление увеличить описательные возможности подобных языков приведет, по нашему мнению, к тому, что они по своей сложности будут приближаться к естественным языкам. Во-вторых, использование формальных средств в качестве метаязыка описания в АПС резко снижает его "потребительскую" ценность, поскольку вынуждает пользователя осваивать этот формальный язык, что для пользователя без специальной лингвистической подготовки достаточно сложно и, как правило, вызывает у него протест. На следующем этапе создания АПС мы используем в качестве языка метаописания русский язык, понимая, что это также не является идеальным решением для многоязычного словаря /см. по этому вопросу с. 60/.

Т о л к о в а н и е

Толкование является мощным средством для точного раскрытия значения слова. Толкование в форме краткого определения дается чаще всего в одноязычных словарях общей лексики, и обычно его достаточно для понимания заглавной ЛЕ в соответствующем значении. В переводных словарях общей лексики толкования используются крайне редко, в основном для семантизации безэквивалентной лексики. В переводных терминологических словарях использование толкования — прием достаточно регулярный для семантизации заглавных ЛЕ. В отличие от одноязычных толковых словарей общей лексики в терминологических словарях чаще используются развернутые толкования, иногда совпадающие с энциклопедическими дефинициями. Основной функцией толкования в таких словарях является семантизирующая функция. Реже толкование используется для дифференцирования значений полисемантических

слов. В АПС и терминологических банках /ТБД/ толкованию уделяется особое внимание. Большинство промышленных АПС и ТБД включают толкования на одном или нескольких языках. Там, где толкование в настоящее время еще отсутствует, в качестве первоочередной поставлена задача включения толкования в структуру словарной статьи /Sager, McNaught, 1980/. Как и в традиционных терминологических словарях, основная функция толкования в АПС и ТБД - семантизирующая. Роль толкования при разграничении значений полисемантических слов особо не подчеркивается. Возможно, это объясняется тем, что в настоящее время все эти автоматизированные системы обработки лингвистической информации нацелены на терминологическую лексику, где проблема многозначности не так остра, как в общей лексике.

В нашем АПС толкование имеет две функции - семантизирующую и дифференцирующую. Главной функцией является семантизирующая функция. Это можно объяснить тем, что большая часть состава нашего АПС представлена терминологической лексикой, для которой, по нашему мнению, семантизация значения в АПС преобладает над дифференциацией значения, поскольку 1/ в большинстве случаев пользователи обращаются к переводному словарю при незнании переводных эквивалентов необходимого ему слова; 2/ переводные словари, в том числе и наш АПС, имеют другие, более компактные средства дифференциации значения, чем толкование. В нашем случае - это тематическая рубрика и специальная зона пояснительных помет. Однако в некоторых случаях мы используем толкование для разграничения многозначности заглавных ЛЕ. Наиболее характерным примером такого использования толкования являются случаи, когда русские переводные эквиваленты, соответствующие разным значениям заглавного иноязычного слова, сами являются многозначными и не различаются по форме. Особенно ярко роль толкования в обеих функциях проявляется в словарных статьях некоторых новых иноязычных терминов, для которых еще не найдены общепринятые русские переводные эквиваленты. В таких случаях толкование выполняет и семантизацию значения, и его дифференциацию /см. Приложение, пример словарной статьи FRAME, значения 5 и 6/.

Зона толкования в нашем АПС имеет неограниченную длину. Мы приняли такое решение, поскольку использо-

вание кратких определений в качестве толкований не является надежным средством при описании значений терминологической лексики и в первую очередь новых терминов, поэтому лексикограф вынужден прибегать к использованию развернутых толкований. Частое отсутствие надежных лексикографических источников для новой терминологической лексики, а также жесткие временные рамки, в которых нередко оказывается составитель словарных статей для автоматического словаря, вынуждают его быть многословным. Все эти причины привели нас к снятию ограничений на длину толкования в нашем словаре и соответственно на длину зоны толкования.

П р и м е р ы у п о т р е б л е н и я

Примеры употребления имеют три основных функции в словаре: 1/ иллюстративную - для показа на конкретном языковом материале примеров употребления заглавной ЛЕ; 2/ семантизирующую - для описания значения заглавной ЛЕ; 3/ дифференцирующую - для разграничения значений полисемантических лексических единиц.

Эти функции реализуются по-разному в словарях разного типа - в зависимости от целей и задач конкретного словаря. Так, в переводной лексикографии контекстные примеры имеют, как правило, две функции: иллюстративную и дифференцирующую. Семантизирующая функция реализуется в основном с помощью переводных эквивалентов. Для переводных словарей свойственно использование дву- трехчленных словосочетаний и беллетристических цитат, целых фраз или предложений. Использование контекстных примеров характерно в основном для интенсивных словарей общеупотребительной лексики. В традиционных научно-технических словарях, большинство которых относится к экстенсивному типу, контекстные примеры, как правило, не используются. Для одноязычных толковых и учебных словарей характерно широкое использование контекстных примеров во всех трех функциях. Для однозначных лексических единиц в таких словарях ведущей является семантизирующая функция, если же заглавное слово многозначно, то семантизирующая функция совмещается с дифференцирующей, причем главной считается последняя /Лексическая основа ..., 1984, с. 12/.

Автоматические переводные словари и терминологические банки весьма существенно различаются между со-

бой по отношению к роли и значению контекста при лексикографировании слова. Наибольшее внимание контексту как важному средству лексикографического описания лексики уделяется в терминологическом банке TERMIUM, в котором каждый термин обязательно сопровождается развернутыми контекстами, четко иллюстрирующими все его значения. /Dubuc, 1972/. Существенное внимание контексту уделяется в таких автоматических словарях, как TEAM, LEXIS, EURODICATOM. С другой стороны, такие ТБД, как EWF, TERMDOK, NORMATERM, не используют контекст в своих словарных статьях /Sager, McNaught, 1980/.

В словарных статьях многозначных лексических единиц нашего АПС контексты выполняют иллюстрирующую и дифференцирующую функции. Зона контекстных примеров имеет в словаре неограниченную длину, что позволяет использовать полные цитаты. Сами цитаты заимствуются из научно-технических текстов по тематике заглавного слова. Отличительной особенностью АПС ВЦП в этом аспекте является то, что в ряде случаев иноязычный контекст сопровождается русским переводом. Функции русского перевода заключаются в основном не в дифференцировании значения заглавного слова, а в показе сферы употребления русского ПЭ, а также частично в семантизации его значения. Русский контекст необходим, когда в силу обратимости АПС данный русский ПЭ становится заглавным словом словарной статьи. При эксплуатации АПС в режиме, когда входными и выходными языками являются иностранные языки, русские контексты должны программным способом устраняться из словарной статьи. Использование русских переводов оригинальных иноязычных контекстов является в настоящее время вынужденным шагом, поскольку основным источником контекстов для нас служили оригинальные иноязычные /английские, немецкие и французские/ научно-технические тексты и их русские переводы. В дальнейшем мы будем избегать использовать переводы и в качестве контекстных примеров для лексических единиц любых типов. Следует сказать, что, поскольку АПС ВЦП является в первую очередь переводным словарем, контексты играют у нас вспомогательную роль при семантизации значения слова и его дифференцировании. Вместе с тем мы считаем целесообразным сохранить в структуре словарной статьи зону контекстных примеров с неограниченной длиной, поскольку при использовании

АПС как терминологического банка или базы лингвистической информации роль контекста при описании различных классов лексики и в первую очередь новой лексики будет более значительной.

2.3.3. Ранжирование значений в словарной статье многозначного слова

Разрешение многозначности не является конечной точкой в работе лексикографа при составлении словарной статьи полисемантического слова. Не менее важной является задача определения порядка следования отдельных значений в словарной статье. "Порядок значений, органически соответствующий истории развития слова, дает наиболее ясную картину его смысловой структуры. С другой стороны, несомненно, что практические соображения требуют ставить на первое место наиболее актуальное и широкоупотребительное значение, каким нередко оказывается позднейшее" /Фельдман, 1957а, с. 25/. Некоторые ученые полагают, что отдельные значения многозначного слова целесообразно размещать в соответствии с частотой их употребления в речи /Borsdorf, Gross, 1981/. В терминологической лексикографии существует точка зрения, что на первое место следует ставить значение, совпадающее с основной тематической направленностью терминологического словаря. Например, в соответствии с таким решением в словаре по фото- и кинотехнике первым у слова FRAME будет стоять значение "кадр", а в словаре лингвистических терминов первым у того же слова должно быть значение "фрейм".

В нашем АПС проблема ранжирования отдельных значений полисемантического слова приобретает дополнительную остроту при составлении "прямых" иностранно-русских словарных статей, и особенно при автоматическом получении русско-иностраннных "перевертышей". В первом случае острота возникает из-за того, что наш АПС комплектуется на основе широкого круга лексикографических и текстовых источников, и поэтому в словарной статье одного многозначного слова нашего словаря сливаются словарные статьи из разных традиционных словарей с самыми разными принципами деления значения слова, с разным количеством отдельных значений и с разным порядком их размещения в словарной статье.

Автоматически объединяются только те значения, которые совпадают полностью по всем зонам: индекса надежности, тематической рубрики переводного эквивалента, пояснительных помет и остальным. При несовпадении этих зон в отдельных значениях одного и того же слова, пришедших из разных словарей, они как бы подклеиваются одно за другим, нарушая тем самым общую структуру словарной статьи. Словарные статьи такого типа должны проходить этап ручной коррекции, на котором лексикограф устраняет избыточную информацию и строит словарную статью в соответствии с принятой структурой.

Еще раз с проблемой структурирования значений в словарной статье разработки сталкиваются при реализации обратимости АПС. В целом обратимость АПС не создает серьезных трудностей при работе с однозначными ЛЕ, но как только происходит автоматическое "перевертывание" многозначной лексики, то возникают дополнительные сложности. Информация для автоматического "перевертыша" поступает из разных словарных статей, и поэтому полученный продукт можно рассматривать только как своего рода полуфабрикат, который следует подвергнуть тщательной доработке. АПС выдает лексикографу неупорядоченный набор переводных эквивалентов со всей сопутствующей информацией, которые объединены только заглавной единицей. Как и в случае с "прямыми" словарными статьями, сливаются только те зоны в словарной статье, которые полностью совпадают. Остальные подсоединяются друг к другу. Лексикографу предстоит распределить переводные эквиваленты в соответствии с выражаемыми ими значениями и ранжировать эти значения определенным образом, убрать дублируемую и, может быть, добавить отсутствующую информацию. Теоретически представляется возможным поставить ранжирование значений в зависимость от частотных характеристик переводных эквивалентов, несущих эти значения, используя для этой цели двуязычные частотные словари типа /Алексеев, 1971; Англо-русский частотный словарь..., 1977 и Русско-английский частотный словарь..., 1977/. Данные словари созданы на основе английских текстов и их русских переводов и содержат много интересных фактов, полезных для лексикографа.

Мы используем эти словари при составлении словарных статей многозначной лексики в нашем АПС. Однако недостаточный объем обработанного материала в этих словарях не позволяет нам считать их основным

критерием в этой работе. Создание же дополнительных словарей такого типа требует больших материальных затрат и времени /Алексеев, 1971, с. 35/, поэтому в практической работе мы больше ориентируемся на существующие авторитетные двуязычные словари. Мы отказались также от идеи ранжировать отдельные значения в зависимости от тематической направленности словаря. По нашему мнению, такое решение представляет определенный интерес в основном для узкотематических словарей в традиционной книжной форме, поскольку естественно предположить, что такие словари в первую очередь используются при работе с текстами данной тематики. Однако для нашего АПС такое решение не может быть рекомендовано, поскольку:

а/ АПС ВЦП в перспективе должен стать политематическим;

б/ программное обеспечение АПС позволяет вычленивать любые тематические значения из состава многозначных слов.

Завершая этот раздел, мы хотели бы еще раз сказать, что, по нашему мнению, автоматические словари в целом и АПС ВЦП в частности обладают необходимым набором средств для семантизации и дифференциации значений многозначной лексики. Структура словарной статьи нашего АПС в принципе дает возможность лексикографу строить такие двуязычные словарные статьи, которые могут органично сочетать свойства переводных и толковых словарей.

2.4. Запрос. Обращение к АПС

2.4.1. Технология осуществления запроса

Под термином ЗАПРОС мы будем понимать обращение пользователя к АПС, а также саму запрошенную лексическую единицу. В нашем АПС в зависимости от режима его эксплуатации предусмотрены следующие способы осуществления запроса:

1. Лично пользователем через дисплейное терминальное устройство ЭВМ. Этот способ используется в основном при диалоговом режиме эксплуатации АПС.

2. С помощью специального бланка, заполняемого пользователем при предварительном просмотре текста оригинала или в процессе перевода. Содержание бланка за-

тем переносится техническими работниками на машиночитаемый носитель и вводится в ЭВМ,

3. Методом подчеркивания запрашиваемой лексики в тексте оригинала с последующим переносом с оригинала на машиночитаемый носитель и вводом в ЭВМ.

4. С помощью команды выдать пользователю переводные эквиваленты всех слов текста, либо слов определенной категории, например слов, относящихся к одному грамматическому классу, либо с определенным индексом надежности, либо одной тематики, либо входящих в несколько категорий одновременно. В настоящее время этот способ реализован в нашем АПС только на уровне цельнооформленных ЛЕ. Это означает, что заглавными словами в текстоориентированном словаре, полученном таким способом, могут быть только отдельные слова даже в тех случаях, когда слово входит в словосочетание и это словосочетание имеется в составе АПС. Такое решение безусловно не является оптимальным, и поэтому в настоящее время ведется работа по созданию специального алгоритма автоматического вычленения раздельнооформленных ЛЕ из текста. Второй, третий и четвертый способы предназначены в основном для пакетного режима эксплуатации. Последний способ запроса является наиболее легким и удобным для пользователя, однако в этом случае текстоориентированные словари резко увеличатся в объеме и не будут отвечать языковым потребностям конкретного пользователя. Кроме того, этот способ экономически невыгоден, так как он предполагает либо наличие переводимого текста на машиночитаемом носителе, что в условиях специализированных переводческих организаций бывает пока весьма редко, либо перенос этого текста на машиночитаемый носитель специально для создания текстоориентированного словаря, что без читающего автомата достаточно дорого /Krollman, 1971, p. 121/.

Метод подчеркивания запрашиваемой лексики в тексте оригинала является весьма удобным способом осуществления запроса и не требует у пользователя больших затрат времени. Однако и он имеет свои недостатки, так как после подчеркивания пользователь должен на некоторое время передать текст оригинала в руки технических работников, что может задержать его работу с переводом.

2.4.2. Лингвистическая форма запроса

Среди основных требований к АПС особо следует выделить возможность общения со словарем на ограниченном естественном языке и уменьшения количества отказов словаря при обработке лексических единиц, формально отсутствующих в словаре /сверхдлинных или содержащих посторонние элементы, слов со словообразовательными элементами и др./. В нашем АПС это проявляется в обеспечении возможности обращения к нему в виде словоформ и текстовых форм словосочетаний, а также автоматического анализа нетиповых запросов. На лингвистическом уровне эти задачи решаются с помощью нескольких алгоритмов, которые могут функционировать как независимо, так и во взаимодействии друг с другом.

Алгоритм лемматизации обеспечивает преобразование текстовой формы запроса в словарную, используется при осуществлении пословного перевода запрошенных, но не найденных в словарных массивах словосочетаний, при анализе структуры сложных слов, при выделении словообразовательных элементов в запрошенных, но не найденных лексических единицах.

Алгоритм обработки "некорректных" запросов позволяет разделить запрошенную, но не найденную в АПС лексическую единицу, например многословную последовательность, на более мелкие единицы, а также ликвидировать в запросах посторонние элементы контекста.

Алгоритм словообразовательного анализа позволяет выделять в запрошенных, но не найденных в словарных массивах лексических единицах словообразовательные элементы и выдавать в качестве ответа на запрос переводной эквивалент основы и толкования значений терминообразующих и словообразовательных элементов, а также их возможные эквиваленты на другом языке.

Взаимодействие лингвистических алгоритмов показано на схеме 3



Схема 3

2.4.3. Алгоритмы лемматизации

Наиболее важным шагом, приближающим обращение к АПС к общению с ЭВМ на ограниченном естественном языке, является, по нашему мнению, возможность осуществлять запрос любых ЛЕ в их текстовой форме. Эта задача решается в нашем АПС с помощью алгоритмов лемматизации, т.е. сведения текстовых форм лексических единиц к их словарным /каноническим/ формам.

Следует отметить, что, несмотря на большое сходство между алгоритмами лемматизации и алгоритмами морфологического анализа, которое заключается в совпадении объекта обработки /словоформа/, а также лингвистических приемов, используемых при построении этих алгоритмов, между ними существует принципиальное различие, которое заключается в их целевой направленности. Алгоритм морфологической обработки направлен на получение морфологической и синтаксической информации об обрабатываемом слове. Задача алгоритма лемматизации проще и заключается в поиске для каждой словоформы ее словарной формы, которую мы понимаем в соответствии с канонами, сложившимися в традиционной лексикографии.

В АПС ВЦП используются два метода решения проблемы лемматизации лексики всех входящих в состав АПС языков - это метод поискового массива словоформ и метод словаря основ с блоком морфологической обработки. Оба эти метода имеют свои положительные и отрицательные стороны. При использовании метода поискового массива словоформ в память ЭВМ заносятся все возможные словоформы лексем, включенных в АПС с отсылками к словарной форме, которая также включается в поисковый массив. Распознавание запроса осуществляется путем сравнения реализованной в тексте словоформы со словоформами поискового массива. Положительной стороной данного подхода является то, что запрашиваемая словоформа не подвергается никакому членению на составные части. Это уменьшает количество возможных сбоев и делает работу АПС в целом более надежной. Отрицательной стороной данного подхода являются повышенные требования к объему памяти ЭВМ. Так, поисковый массив словоформ английского языка будет примерно в два раза превышать по объему соответствующий массив основ, а для русского языка это соотношение

увеличится в 10-12 раз /О многоцелевом автоматическом словаре, 1975, с. 156-157/.

Метод словаря основ предполагает распознавание запрашиваемой словоформы путем членения ее на составные части - основу и словоизменительную флексию. Затем следует отсылка к словарной форме запрошенной словоформы, которая хранится в памяти ЭВМ, либо синтез словарной формы из основы и соответствующей флексии. Использование этого метода предполагает создание алгоритма морфологической обработки со словарем основ и таблицами словоизменительных флексий. Отрицательной стороной данного подхода является более высокая по сравнению с методом поискового массива сложность и трудоемкость лингвистических работ /Нелюбин, 1983, с. 150/. С другой стороны, этот подход позволяет решать большое количество задач и экономит память ЭВМ.

Выбор того или иного метода зависит от особенностей системы словоизменения конкретного языка. В принципе возможно использование обоих методов для одного и того же языка. В настоящем разделе изложены принципиальные лингвистические решения, положенные в основу алгоритмов лемматизации английского, русского и французского языков.

2.4.3.1. Лемматизация английской лексики

Для английского языка в нашем АПС используется метод поискового массива словоформ, при котором в память ЭВМ заносятся все словоформы, варианты написания и сокращения английских ЛЕ как отдельных слов, так и слов, входящих в состав словосочетаний, включенных в состав АПС. Основными факторами в пользу выбора метода поискового массива словоформ для английского языка являются слаборазвитая система словоизменения в этом языке, а также тот факт, что словарные формы лексики основных знаменательных частей речи /существительное, глагол, прилагательное/ реализуются в английских научно-технических текстах почти в два раза чаще, чем их несловарные формы.

Большую часть массива английских ЛЕ составляют именные словосочетания. Для того чтобы в целях экономии памяти ЭВМ не вводить в поисковый массив все теоретически возможные текстовые формы словосочетаний, в АПС предусмотрен несложный алгоритм их лемматиза-

ции. На вход алгоритма поступает не найденное в словаре запрошенное словосочетание. Рассматривается последнее слово запроса. Если это слово в том виде, в каком оно встретилось в словосочетании, есть в АПС, то предполагается, что изменено предыдущее слово. Если окажется, что каждое слово запроса имеется в АПС, но тем не менее все словосочетание не найдено, то комбинаторным способом создаются такие варианты сочетаний, где на место каждого слова последовательно ставятся его словоформы. В первую очередь проверяется вариант, когда последнее слово ставится во множественном числе, а предыдущие слова остаются в словарном виде. Если среди полученных комбинаций ни одна не будет найдена в АПС, то алгоритм выдает информацию "словосочетания нет в словаре", при этом могут быть даны словарные статьи отдельных найденных в АПС слов этого словосочетания.

2.4.3.2. Лемматизация русской лексики

Для русского языка проблема запроса в текстовой форме решается с помощью словаря основ и блока морфологической обработки /Соколова, 1984/. Русская словоформа, поступившая на обработку в алгоритм лемматизации, расчленяется на машинную основу и словоизменятельную флексию. Затем следует отсылка к словарной форме одного слова или нескольких слов, если основа и словоизменятельная флексия омонимичны. Выбор такого пути решения проблемы лемматизации для русского языка объясняется тем, что русский язык отличается весьма развитой системой словоизменения, что ведет к тому, что в реальных текстах русские словоформы встречаются в три раза чаще, чем их словарные формы. Положительной стороной этого подхода, как мы уже отмечали ранее, является также экономное расходование памяти ЭВМ. Однако сложность лингвистических работ при этом существенно выше, чем при использовании метода поискового массива словоформ.

Лингвистическое обеспечение алгоритма лемматизации для русского языка включает:

- массивы словарных форм отдельных русских слов и словосочетаний;
- словарь машинных основ отдельных русских слов с указанием их порядковых номеров и соответствующих словоизменятельных классов;

- список окончаний с перечнем соответствующих словоизмерительных классов;

- массив последовательностей кодов русских словосочетаний.

Массив словарных форм русских словосочетаний комплектуется из словосочетаний, являющихся переводными эквивалентами в иностранно-русских исходных словарных статьях. Массив словарных форм отдельных русских слов составляется из заголовков русских лемматизационных словарных статей. Все слова, входящие в массив словарных форм отдельных русских слов, разделяются на следующие крупные лексико-грамматические классы: существительные, глаголы, прилагательные, наречия, предлоги, союзы.

Все причастия рассматриваются как прилагательные, так как переводчику, по нашему мнению, целесообразнее давать аналогичную форму на языке перевода, чем исходную глагольную форму. Например, более корректно на запрос АВТОМАТИЗИРОВАННЫЙ дать ответ АВТОМАТЕД, чем АВТОМАТИЗИРОВАННЫЙ ⇒ АВТОМАТИЗИРОВАТЬ ⇒ АВТОМАТЕ. Кроме того, некоторые причастия имеют в качестве переводного эквивалента прилагательное, а не глагольную форму. Однако в тех случаях, когда запрошенное причастие не было включено в словарь, предусматривается возможность найти глагол, от которого оно образовано. Деепричастия и прилагательные в сравнительной или превосходной степени как самостоятельные единицы в словарь не включаются, а алгоритмическим путем приводятся к исходному глаголу или прилагательному в положительной степени. Сложные слова с дефисом включаются в массив словарных форм в полном виде, например ПРОГРАММА-ДИСПЕТЧЕР, БЛЕДНО-ЗЕЛЕНый и т.п.

В качестве словарной формы глагола выступает неопределенная форма несовершенного вида. Поскольку в русском языке совершенный вид глаголов может образовываться двумя способами: приставочным /"делать-сделать"/ и суффиксальным /"включать-включить"/, то для глаголов второго типа в массив словарных форм включается как форма несовершенного вида, так и форма совершенного вида.

Для глаголов с возвратными частицами -ся, -сь в массив словарных форм включаются формы без возвратных частиц, если эти глаголы могут без них употребляться. Если же глаголы не могут употребляться без этих частиц /"бороться"/ или могут употребляться, но при этом меняют свое значение, в массив словарных форм включается форма с частицей.

Каждое отдельное русское слово в массиве словарных форм имеет жестко фиксированный номер /адрес/. При пополнении АПС эти номера не изменяются. Каждое слово имеет также адресные ссылки к ПЭ на других языках, входящих в состав нашего АПС.

В массиве словарных форм словосочетаний различаются следующие виды словосочетаний: именные, глагольные, атрибутивные и наречные. Они включаются в словарь в полном виде с указанием на тип словосочетания. Так же как и отдельные слова, они имеют адресные отсылки к переводным эквивалентам на других языках.

Словарь машинных основ представляет собой перечень всех основ русских слов, входящих в массив словарных форм отдельных русских слов, а также слов, входящих в состав словосочетаний из массива словарных форм русских словосочетаний. В принятой в словаре грамматической классификации под машинной основой понимается наибольшая неизменяемая часть слова, причем, если в парадигме слова происходят морфологические изменения, выделяется соответствующее число основ, отражающих эти изменения /Соколова, Убин, 1988/. При разработке грамматической классификации мы руководствовались принципом: лучше иметь большее число основ, чем окончаний. В качестве основ неизменяемых слов выступают сами эти слова.

У каждой основы из словаря машинных основ указывается порядковый номер этой основы и номер словоизменительного класса, описывающего словоизменительную парадигму этого слова. Порядковый номер основы определяется в зависимости от порождающей силы данной основы в словоизменительной парадигме. Например, для слова "болт" выделяется основа "болт-", от которой образуются все грамматические категории, поэтому ей приписывается порядковый номер I. Для слова "гайка" выделяются две основы: "гайк-" и "гаек-", причем от второй основы образуются только форма родительного падежа множественного числа, а все остальные грамматические категории образуются от первой основы, поэтому им приписываются следующие порядковые номера: "гайк- 1", "гаек- 2". Номер словоизменительного класса определяется с помощью специальных таблиц.

Словоизменительные классы объединяют слова, одинаково образующие свои словоформы. Номер класса

/НК/ однозначно определяет, какое количество основ имеет конкретное слово для образования всех его грамматических категорий, образующих его парадигму, а также какие окончания и основы участвуют в образовании определенной словоформы. Каждое слово может входить в один словоизменительный класс, за исключением тех случаев, когда языковая норма разрешает различное образование той или иной грамматической категории, вследствие чего одно и то же слово можно отнести к разным словоизменительным классам. Например, слово "полюс" имеет два варианта именительного падежа множественного числа: "полюса" и "полюсь", поэтому это слово будет входить в два словоизменительных класса. В настоящее время выделено 295 словоизменительных классов, которые, по нашему предположению, описывают практически все возможные варианты словоизменения для русских слов.

Для анализа русской словоформы необходимо правильно найти ее основу и окончание. Основным инструментом здесь является список словоизменительных окончаний, которые могут быть организованы в традиционном алфавитном порядке, а в пределах каждой буквы - по возрастанию длины или же по принципу обратного словаря. Для каждого окончания, в свою очередь, составляется список номеров словоизменительных классов, в которых его окончание может присоединяться к первой основе, затем список классов, где данное окончание присоединяется ко второй основе, и т.д. Необходимым и достаточным условием включения того или иного класса в эти списки является хотя бы единичное вхождение данного окончания в словоизменительную парадигму, которую описывает этот класс. В настоящее время выделено 161 окончание. Максимальная длина окончания - четыре буквы.

На вход алгоритма лемматизации отдельных русских слов поступает анализируемая словоформа. На первом этапе работы алгоритма у данной словоформы находятся все возможные окончания, включая нулевое. Для этого с конца словоформы по одной отсекаются буквы, последовательность которых затем проверяется на вхождение в список окончаний. В том случае, если отсеченная последовательность букв не найдена в списке окончаний, но может входить хотя бы в одно окончание большей длины, отсечение букв продолжается. Например, у запрашиваемой словоформы "УСТРОЙСТВАМИ" сначала пред-

полагается возможность нулевого окончания, затем отсекается буква -и, которая входит в список окончаний /эти варианты разбиения на основу и окончание запоминаться/; окончания -ми в списке окончаний нет, поэтому вариант разбиения УСТРОЙСТВА-МИ отвергается как невозможный; однако последовательность букв -ми может входить в окончания большой длины / -ами , -нами и т.д./, поэтому отсекается третья буква с конца и новая последовательность букв отыскивается в списке окончаний. Поскольку в списке есть такая последовательность, то вариант разбиения УСТРОЙСТВ-АМИ запоминается как возможный, а разбиение УСТРОЙСТ-ВАМИ отвергается как невозможное, так как такого окончания в списке окончаний нет.

На втором этапе работы алгоритма лемматизации оставшиеся после отсеечения окончаний части словоформы проверяются на вхождение в массив машинных основ. Если эта гипотетическая основа есть в поисковом массиве, считается, что такой вариант разбиения на основу и окончания является возможным, и он отсылается в накопитель гипотез. Если же такая основа не найдена, то вариант разбиения считается неверным и в дальнейшем не рассматривается. Так, для приведенного выше примера вариант разбиения УСТРОЙСТВ-АМИ считается возможным и будет направлен в накопитель гипотез, так как в поисковом массиве машинных основ есть основа УСТРОЙСТВ-, вариант разбиения УСТРОЙСТВАМ-И будет отброшен как невозможный, так как основы УСТРОЙСТВАМ- в поисковом массиве машинных основ нет.

В тех случаях, когда у анализируемой словоформы не найдено ни одного окончания, она выдается на печать с информацией "нет в словаре". Если же у анализируемой словоформы были найдены возможные окончания, но не удалось найти ни одной основы в массиве основ, она поступает на обработку в блок словообразовательного анализа.

На третьем этапе работы алгоритма лемматизации проверяется правильность вариантов разбиения входной словоформы на основу и окончание, которые поступили в накопитель гипотез. Другими словами, следует определить, совместимо ли найденное окончание с найденной основой. Для этого проверяется, входит ли номер словоизменительного класса, приписанного основе, в список словоизменительных классов найденного окончания. Если этот номер найден в списке, считается, что

данная основа совместима с данным окончанием, и, таким образом, этот вариант разбиения входной словоформы из накопителя гипотез считается верным. В противном случае данный вариант разбиения считается неверным. Если все варианты из накопителя гипотез оказались неверными, то словоформа подается на вход алгоритма словообразовательного анализа.

На последнем этапе работы алгоритма по адресной ссылке, указанной у найденной основы, находится соответствующая каноническая форма. В нашем примере по адресной ссылке основы УСТРОЙСТВ- будет найдена каноническая форма УСТРОЙСТВО, которая в свою очередь имеет адресные ссылки к иноязычным переводным эквивалентам.

Особые случаи при приведении к канонической форме отдельных русских слов представляют омографы и глаголы с возвратными частицами, которые из-за недостатка места мы рассматривать не будем и отошлем читателя к /Соколова, 1984/.

Важную роль при обработке русской лексики играет лемматизация словосочетаний. Для этой цели создан отдельный алгоритм, на вход которого поступает русское словосочетание в текстовой форме. Задача алгоритма - привести это словосочетание к словарному виду. На первом этапе работы алгоритма каждый компонент словосочетания с помощью алгоритма лемматизации отдельных слов приводится к словарному виду. Поскольку каждая каноническая форма в массиве словарных форм имеет жесткофиксированный номер, то можно все компоненты словосочетания заменить их номерами и тем самым получить неизменяемую кодовую последовательность для анализируемого словосочетания. Полученная кодовая последовательность отыскивается в массиве кодовых последовательностей русских словосочетаний. Если она найдена, то по адресной отсылке находится сама словарная форма словосочетания из массива словарных форм с соответствующими адресами переводных эквивалентов на всех входящих в словарь языках. В том случае, если эта последовательность не найдена, запрашиваемое словосочетание переходит в блок обработки "некорректных" запросов. Одним из преимуществ кодовой записи словосочетаний является возможность нахождения всех словосочетаний, в которые входит интересующее пользователя слово, независимо от того, в какой грамматической форме оно встретилось.

Таковы общие принципы построения лингвистического обеспечения блока лемматизации русской лексики в нашем АПС.

2.4.3.3. Лемматизация французской лексики

Проблема сведения текстовых форм лексических единиц к словарным для французского языка решается с помощью комбинации двух уже описанных подходов: для существительных и прилагательных используется метод поискового массива словоформ, для глаголов – алгоритм лемматизации, предполагающий членение словоформы на основу и окончание /Ануфриева, 1984/.

Лингвистическое обеспечение алгоритма лемматизации для французского языка включает:

- поисковый массив словоформ;
- поисковый массив основ глаголов;
- поисковый массив глагольных окончаний;
- таблицы определения словоизменительных классов глаголов.

Поисковый массив словоформ создается автоматически на этапе формирования информационной базы АПС из лексических единиц, содержащихся в зонах 1 и 3 словарных статей всех частей речи, кроме глаголов. Из словарных статей глаголов берутся лишь заголовки. Все единицы поискового массива словоформ снабжаются отсылками к соответствующим единицам из массива заголовков словарных статей.

Словоизменительная классификация французских глаголов, используемая в нашем АПС, включает максимально 48 форм /в том числе и омографичные/. На основе анализа всей парадигмы для каждой словоизменительной модели выделены основы и окончания. В качестве основы выбирается максимальная неизменяемая часть слова. В тех случаях, когда внутри основы происходят фонетические чередования, для описания парадигмы выбирается несколько основ, с тем чтобы не увеличивать длину и число окончаний. При этом учитывается также количество глаголов, спрягающихся по данному образцу, и регулярность соответствующего чередования. В среднем в образовании глагольных парадигм участвуют три основы. Максимальное число основ /10/ имеет глагол *être*. Основы нумеруются в порядке появления их в парадигме /начиная с инфинитива/. Затем вся парадигма записыва-

ется с помощью формул вида " $N + t$ ", N - порядковый номер основы, а t - окончание. Глаголы, у которых наборы формул полностью совпадают, образуют один словоизменительный класс. В настоящее время выделено 70 словоизменительных классов французского глагола. Оказались охваченными практически все типы глагольных парадигм. Однако система в целом является открытой для пополнения.

В силу того что большинство словоизменительных классов включает лишь по нескольку глаголов, которые можно задать списком, а представительные классы хорошо поддаются содержательному описанию для определения принадлежности глагола к тому или иному словоизменительному классу, используются специальные таблицы, в которых комбинируются оба упомянутых выше принципа, а именно:

- описание классов, основанное на двух критериях: 1/ принадлежность глаголов к одной из трех групп /1-й, 2-й или 3-й/, традиционно выделяемых во французской грамматике; 2/ ряд фонетических особенностей основы;

- задание списком ряда неправильных глаголов.

Поисковый массив основ формируется из единиц, содержащихся в зоне словарных статей глаголов. Каждой основе приписан порядковый номер, а также номер словоизменительного класса соответствующего глагола.

Поисковый массив окончаний включает все значения переменной t из формул $N + t$, встречающиеся в парадигмах словоизменительных классов французских глаголов. Поисковый массив окончаний, в принципе, является закрытым множеством. Лишь пополнение набора словоизменительных классов может повлечь за собой те или иные изменения в данном массиве. Каждому окончанию в поисковом массиве приписана информация о том, в каких словоизменительных классах и в сочетании с какими номерами основ оно зафиксировано.

Процесс лемматизации словоформ цельнооформленных лексических единиц французского языка осуществляется следующим образом. На первом этапе запрошенная ЛЕ сравнивается с поисковым массивом словоформ. При полном ее совпадении с одним из элементов этого массива по специальной ссылке находится нужная словарная статья. Если найденная таким образом заглавная лексическая единица является существительным или прилагательным, а запрошенная словоформа оканчивается на $-e$, $-ê$, $-i$, $-s$, $-t$ или $-u$, то необходимо прове-

рить наличие глагольной формы - омографа, для чего запрошенная словоформа подвергается дополнительной обработке блоком лемматизации, осуществляющим членение словоформы на основу и окончание. Если же сравнение запроса с поисковым массивом словоформ дало отрицательные результаты, запрошенная словоформа сразу поступает на обработку в блок лемматизации.

В целом принцип действия блока лемматизации для французского языка аналогичен порядку функционирования соответствующего блока для русского языка. Специальная программа производит членение запрошенной ЛЕ на гипотетическую основу и окончание, проводя последовательно отсечение букв с конца словоформы, а затем ищет соответствующие элементы в поисковом массиве основ и поисковом массиве окончаний и проверяет их на совместимость. Поскольку практически исключается ситуация, когда совпадали бы отдельные формы различных глаголов, для каждой глагольной словоформы возможен лишь один вариант разбиения, и, следовательно, поиск может быть прекращен после нахождения первой же правильной гипотезы. Если в результате применения блока лемматизации запрошенную единицу не удалось найти, она направляется на словообразовательный анализ.

Первым шагом при лемматизации французских словосочетаний является обращение к поисковому массиву, в который заносятся все канонические формы словосочетаний, а для именных и атрибутивных сочетаний также все числовые и родовые варианты. Если в результате такого обращения получен отказ, то словосочетание разбивается на отдельные слова и каждое слово подвергается лемматизации. В этом случае каждому отдельному французскому слову дается переводной эквивалент. Если же первый компонент словосочетания является глаголом, что выясняется после его лемматизации, то в запросе производится замена: вместо первого компонента, в качестве которого могла выступать личная или причастная форма глагола, подставляется его каноническая форма и происходит повторное обращение к поисковому массиву.

2.4.4. Алгоритм обработки "некорректных" запросов

2.4.4.1. Типы "некорректных" запросов

Опыт работы с автоматическими словарями показал, что многие отказы, возникающие при обработке запросов,

объясняются не только отсутствием запрошенных единиц в АПС или ошибками на лингвистическом и программном уровнях, но и искажениями в тексте запроса или их "некорректным" составлением. "Некорректными" считаются запросы, содержащие такие посторонние элементы контекста, как служебные слова, определения, цифры и т.п., или запросы, состоящие из двух или более самостоятельных лексических единиц. Причем посторонние элементы могут быть в начале, в конце или в середине запроса, например /посторонние элементы подчеркнуты/: with chip calculators, control unit MA-2400, intellectual /smart/ system, thin common sense algorithm /CSA/ data base system has been...

Причиной появления "некорректных" запросов может быть неаккуратность пользователя при составлении запросов методом подчеркивания, когда пользователь только подчеркивает в тексте оригинала интересующую его лексику, а вводят в ЭВМ его запросы операторы, которые могут не знать языка оригинала. Второй причиной может быть неумение правильно выделять в тексте лексические единицы из-за слабого знания языка оригинала.

С целью устранения таких ошибок разработан единый для всех рабочих языков АПС алгоритм обработки "некорректных" запросов, суть работы которого заключается в отсечении посторонних элементов контекста, разбиении запрошенного набора слов на потенциальные терминологические единства и рассмотрении различных гипотез, из которых путем сравнения со словарной базой данных выбирается верная.

2.4.5. Алгоритм словообразовательного анализа

К алгоритму обработки "некорректных" запросов примыкает алгоритм словообразовательного анализа, который решает такую же задачу, но уже не на уровне словосочетания, а в пределах цельнооформленной ЛЕ /Здорных, 1984/.

Деривационный потенциал научного языка в процессе развития самой науки непрерывно возрастает, так как в терминологических системах вследствие интенсивности протекающего отбора словообразовательных средств возникают собственные, характерные для данной области науки терминологические элементы, с участием ко-

торых формируются определенные словообразовательные структуры. Более того, в силу интеграции и дифференциации наук и их терминологий число производных и сложных терминов, образованных путем перекрестного заимствования стандартных терминоэлементов, постоянно увеличивается. В качестве примера можно привести такие широко используемые в различных терминосистемах форманты, как рус.: авто-, энерго-, -скоп", англ.: poly-, graphy; нем.: hoch-, neu-; франц.: de-, -eur.

Включение в АПС всех производных и сложных терминов трудновыполнимо и нецелесообразно. Гораздо удобнее иметь для каждого входящего в состав АПС языка набор наиболее продуктивных терминоэлементов с указанием их значений и/или способов передачи на других языках, сопряженный с механизмом их вычленения в составе запрашиваемых ЛЕ. Исследователи языковых деривационных систем указывают, что при образовании производных терминов используется только часть деривационного потенциала языка. Так, при образовании производных военных терминов в английском языке используется только 45% всего аффиксального инвентаря английского языка. Этот факт с учетом высокой частоты употребления производных терминов в технических текстах существенно повышает роль блока словообразовательного анализа в АПС. Вместе с тем следует предостеречь от не критического восприятия результатов работы блока словообразовательного анализа. Значение выделяемой морфемы не должно противоречить общему смыслу термина и должно обязательно сверяться с контекстом /Шевчук, 1983/.

Блок словообразовательного анализа в нашем АПС является факультативным элементом в системе лингвистических алгоритмов. Это означает, что не все запрошенные словоформы подвергаются максимально дробному морфологическому членению, а лишь те из них, которые не удалось найти на этапе лемматизации в поисковых массивах, т.е. прежде всего различные потенциальные лексические единицы и неологизмы. Таким образом, алгоритм словообразовательного анализа помогает решить две взаимосвязанные задачи: с одной стороны, сократить объем лексических массивов АПС за счет дериватов, с другой – повысить эффективность его работы за счет уменьшения количества отказов, вызванных отсутствием запрошенного слова в АПС.

Лингвистическое обеспечение алгоритма словообразовательного анализа включает:

- поисковый массив производящих основ;
- словарь словообразовательных элементов.

Для русского языка поисковым массивом служит словарь основ. Для английского языка в качестве поискового используется массив словоформ. Во французском языке поиск производящей основы ведется по поисковому массиву словоформ для существительных, прилагательных и неизменяемых частей речи и по поисковому массиву основ для глаголов.

Массив словообразовательных элементов для каждого языка включает как традиционно выделяемые словообразовательные элементы, так и наиболее характерные для научных терминосистем аффиксы и полуаффиксы. В категорию полуаффиксов попадают, с одной стороны, исконные единицы языка, прежде всего предлоги /например, русские "сверх", "около-"/, с другой - элементы, происходящие от греческих и латинских корней /"фото-", "гео-" и т.п./. Эти единицы широко используются для образования терминов. На каждый словообразовательный элемент создается словарная статья, которая по своей структуре напоминает словарные статьи самостоятельных лексических единиц. Различия заключаются в наполнении отдельных зон. Кроме того, словарные статьи словообразовательных элементов необратимы. "Переводные эквиваленты", приводимые в словарных статьях иноязычных словообразовательных формантов, носят иллюстративный характер /см. Приложение, пример № 17/.

Суть работы алгоритма словообразовательного анализа заключается в выделении в составе не найденной в лексических массивах словаря ЛЕ словообразовательных элементов и сравнении остатка с массивами словоформ и основ языка запроса. В случае совпадения пользователю выдаются переводные эквиваленты слова входного языка, с которым совпадает или в состав которого входит найденная производящая основа, и словарные статьи выделенных словообразовательных элементов. В принципе в составе слова может быть более одного словообразовательного элемента. В этом случае членение запроса ведется по методу непосредственно составляющих с последовательным сравнением остатка с массивами словоформ или основ. В общем случае в первую очередь отделяются префиксы. Если данное отсечение не

дало положительного результата, отсекается суффикс. Если и в этом случае не произошло отождествления остатка с поисковым массивом производящих основ, отсекаются префикс и суффикс и поиск повторяется снова. Если же все разбиения оказались неудовлетворительными, пользователю выдаются словарные статьи выделенных словообразовательных элементов и остаток с пометой "нет в словаре".

В качестве иллюстрации работы алгоритма словообразовательного анализа можно привести анализ английского существительного "irregularity" /нерегулярность/. Предположим, что этого слова в английском лексическом массиве не оказалось. Тогда запрос поступает в блок словообразовательного анализа. Отсекается префикс *ir-*. Предположим, что остаток "regularity" в поисковом массиве английских ЛЕ не обнаружен. Тогда следует разбиение *irregular - ity*. Предположим, что остаток "irregular-" также не найден, тогда следует разбиение *ir - regular - ity*. Остаток "regular" в английском лексическом массиве найден. Пользователю выдаются следующие словарные статьи:

REGULAR (A)

ОБЩ. *Б

1. НОРМАЛЬНЫЙ /ПРИЛ./

ОБЩ. *Б

1. РЕГУЛЯРНЫЙ /ПРИЛ./

ОБЩ. *Б

1. ПРАВИЛЬНЫЙ /ПРИЛ./

IR (PREF)

ОБЩ. *Б

1. НЕ/ПРЕФ/

(IR+N/A=N/A)

НП: IRRATIONAL/НЕРАЦИОНАЛЬНЫЙ

ТО: ОТСУТСТВИЕ ПРИЗНАКА, ВЫРАЖЕННОГО ПРОИЗВОДЯЩЕЙ ОСНОВОЙ

ITY (SUF)

ОБЩ. *Б

1. ОСТ/СУФ/

/A+ITY=N/

НП: RATIONALITY/РАЦИОНАЛЬНОСТЬ; SARABILITY/СПОСОБНОСТЬ

ТО: ПРИЗНАК, КАЧЕСТВО, УСЛОВИЕ; СОСТОЯНИЕ, СВОЙСТВО

Особым видом алгоритма словообразовательного анализа фактически является алгоритм обработки русских причастий. Для всех причастий русского языка, не включенных в состав словаря, предусмотрена возможность найти глаголы, от которых они образованы. При этом в качестве ответа выдается исходный глагол и грамматическая информация для словообразовательного суффикса.

Несколько особняком стоят так называемые интернациональные приставки, которые в нашем словаре выделены в отдельные списки. Эти списки составляются отдельно для каждого языка и свободны для пополнения. В силу своей интернациональности такие словообразовательные элементы не нуждаются в полных словарных статьях с перечнем всех переводных элементов. Для пользователя вполне достаточно простого разбиения запрашиваемого слова на приставку и производящую основу, для которой будет дана полная словарная статья. Например, при запросе слова *photosynthesis* оно будет разбито на *photo* - и *synthesis*. Интернациональный элемент *photo* будет выдан пользователю без переводного эквивалента, а слово *synthesis* получит стандартную словарную статью.

Как мы уже говорили выше, блок словообразовательного анализа является факультативным компонентом в лингвистическом обеспечении АПС и вступает в действие тогда, когда все другие средства анализа неопознанной лексической единицы исчерпаны. Данный блок увеличивает эффективность АПС особенно для пользователей с недостаточно глубоким знанием языка оригинала. Алгоритм словообразовательного анализа совместно с алгоритмами лемматизации будут особенно эффективны при обработке лексики агглютинативных языков, например венгерского /Узони, 1983/. Мы также полагаем, что блок словообразовательного анализа может представлять определенный интерес для исследователей словообразовательных систем русского, английского, немецкого и французского языков, поскольку с его помощью можно автоматическим образом получить на больших массивах текстов данные по частоте производных слов и по продуктивности отдельных словообразовательных элементов и моделей. Этот же блок может применяться при использовании нашего АПС в качестве автоматизированного тренажера при изучении названных выше языков как иностранных, хотя такой зада-

чи мы перед собой в настоящее время специально не ставили.

Возможности гибкого обращения к АПС наряду со структурой словарной статьи являются, на наш взгляд, основными компонентами его лингвистического обеспечения. Наш словарь позволяет благодаря алгоритмам лемматизации осуществлять поиск лексической единицы в словаре по любой ее словоформе. Алгоритмы обработки "некорректных" запросов и словообразовательного анализа способствуют повышению надежности работы АПС. Возможность обращения к словарю в любой текстовой форме резко снижает уровень требований к знанию входного языка, что позволяет работать с АПС практически всем категориям пользователей, а также использовать его как компонент автоматизированной лексикографической системы для решения некоторых задач одноязычной и переводной лексикографии.

2.5. Ответ

2.5.1. Лингвистическая форма ответа и способы ее воспроизведения

Проблема выбора лингвистической формы ответа в АПС эквивалентна проблеме выбора лингвистической формы переводного эквивалента в традиционном переводном словаре, которая в языках с богатыми лексикографическими традициями давно и однозначно решена. Переводной эквивалент, как и заглавное слово в традиционных переводных словарях, представлен в виде словарной /канонической/ формы слова, под которой выступает одна из словоформ этого слова, традиционно избранная в качестве представителя данной лексемы в языке.

Разработчики автоматических переводных словарей и терминологических банков проблему лингвистической формы ответа даже не дискутируют и полностью принимают решения традиционной лексикографии. Во всех АПС и ТБД в качестве лингвистической формы ответа выбрана словарная форма слова. В нашем многоязычном автоматическом словаре ответ для всех языков также дается в словарной форме в соответствии с традициями лексикографии каждого конкретного языка. Помимо традиций такое решение объясняется еще тем, что в АПС, в отличие от систем МП, не предусмотрен полный цикл морфологического и синтаксического анализа. Лемматизация,

которая решает задачу сведения текстовых форм слова к его словарной форме, в некоторых случаях дает достаточную информацию для выдачи ответа в виде соответствующей словоформы вместо традиционной словарной формы. Однако поскольку последовательно решить эту проблему для всех запросов средствами только блока лемматизации невозможно, то в ответах будут чередоваться словарные и текстовые формы слов, что, по нашему мнению, скорее будет затруднять пользователю восприятие ответа, чем помогать ему.

Конкретные способы воспроизведения словарных форм слов для разных языков в нашем АПС зависят от стратегического направления, выбранного для решения проблемы лемматизации для каждого языка, и реализуются на основе тех же лингвистических алгоритмов и лексических массивов.

В нашем АПС реализованы два способа воспроизведения словарной формы слова, а именно:

- включение словарной формы наравне со словоформами данного слова в поисковый массив и вызов ее при формировании ответа:

- синтез словарной формы из основы и флексии, которые хранятся в массиве основ и таблице флексий данного языка.

Выбор первого или второго способа из этих двух зависит от конкретного языка и версии АПС. В версиях АПС на базе ЭВМ ЕС и ЕС 1832 используется способ включения словарной формы в поисковые массивы словоформ английского, немецкого и французского языков и создание специального массива словарных форм отдельных слов и словосочетаний для русского языка наряду с массивом основ русских слов. При этом каждая русская основа имеет адресную отсылку к соответствующей словарной форме. В модельной версии АПС на базе ЭВМ СМ в целях экономии объема памяти ЭВМ используется метод синтеза словарных форм слов из соответствующих основ и окончаний.

2.5.2. Форма представления ответа

Размещение ответа на дисплейном терминале или на распечатке АЦПУ должно максимально облегчать пользователю восприятие информации. Сам ответ не должен затемняться не нужной пользователю служебной инфор-

мацией. Ответы, получаемые при пакетной обработке запросов, особенно при большом количестве запросов, фактически являются текстоориентированными микрословарями, которые направлены на удовлетворение потребностей конкретного пользователя, с учетом уровня его языковой компетенции, так как составлены по его запросам. В то же время эти микрословари отражают специфику лексического состава данного текста, так как составлены на основе этого текста.

Ответы в таком текстоориентированном микрословаре могут быть по желанию пользователя размещены либо в алфавитном порядке, либо в порядке поступления запросов. Алфавитный способ размещения ответов является более экономным и весьма удобен в тех случаях, когда перевод текста большого объема осуществляется по частям одновременно несколькими переводчиками. При этом легче достигается единообразие перевода терминологической и другой лексики. Второй способ размещения ответов является менее экономным, однако он уменьшает затраты времени при работе с текстоориентированным микрословарем. Этот способ применяется также при автоматическом переводе списочных частей технической документации /перечней деталей, закладных листов, накладных и т.п./, в которых возможны многократные повторы одних и тех же лексических единиц, но не допустимо их слияние. АПС ВЦП позволяет получать оба варианта текстоориентированных словарей.

Текстоориентированные словари довольно неожиданно для разработчиков автоматических словарей нашли свое применение не только при переводе, но также и среди технических специалистов со средним и слабым знанием иностранного языка при работе с иноязычными текстами. Это объясняется тем, что именно технические специалисты теряют больше всего времени при чтении иностранной литературы в оригинале из-за недостаточного знания иностранного языка. Текстоориентированный словарь является также хорошим стимулятором создания обратной связи между переводчиком и АПС, заставляя переводчика постоянно оценивать надежность и полноту лексических массивов автоматического словаря на основе оригинальных текстов, с которыми он работает /Krollmann, 1971, p. 121/.

2.5.3. Объем и состав информации в ответе

Разные задачи и разные уровни языковой подготовки пользователя предъявляют разные требования к составу и объему информации в словарной статье словаря, что в условиях традиционной лексикографии требует создания словарей разных типов. Так, в переводных словарях интенсивного типа словарная статья может содержать помимо переводных эквивалентов различные сведения семантического и стилистического характера о заглавном слове и его эквивалентах в выходном языке, примеры употребления, типовые и идиоматические словосочетания, синонимы и т.п. Как правило, значительная часть этой информации для читателя с глубоким знанием языка оригинала, каковым является, например, профессиональный переводчик, избыточна. С другой стороны, в переводных технических словарях, которые в большинстве являются словарями экстенсивного типа, словарная статья значительно беднее и в большинстве случаев не содержит практически никакой информации, кроме переводных эквивалентов, что для читателя со слабым или средним уровнем знания языка оригинала явно недостаточно. В условиях традиционной лексикографии это противоречие между уровнем языковой подготовки читателя и разнообразием задач и целей словаря, с одной стороны, и объемом и характером информации в словарной статье этого словаря – с другой, не разрешимо в рамках одного словаря из-за его статичности.

АПС имеют принципиальную возможность преодолеть противоречие между словарями интенсивного и экстенсивного типов, представляя возможность самому пользователю определять состав и объем необходимой ему информации. АПС ВЦП предлагает пользователю девять вариантов ответа, начиная с самого простого, содержащего только один /первый/ переводной эквивалент без какой-либо дополнительной информации, до полной словарной статьи. Расширение состава информации в ответе идет в соответствии с логикой структуры словарной статьи. К первому переводному эквиваленту могут быть добавлены зона его лексикограмматического класса, затем зона пояснительных синтаксических и семантических помет, затем зона контекстных примеров и зона толкования. Следующая группа ответов включает все переводные эквиваленты с нарастанием состава ин-

формации по изложенному выше принципу. Подробно состав всех вариантов ответа представлен в соответствующей графе матрицы параметрического анализа нашего АПС /см. Приложение № 2/. На все варианты ответа могут налагаться дополнительные ограничения по тематике и индексу надежности, т.е. пользователь может ограничить ответ лексикой только одной определенной тематической области, например вычислительной техникой, или одним индексом надежности, например стандартом.

Возможность варьировать по желанию пользователя состав и объем информации в ответе является принципиально важным свойством АПС ВЦП, которое позволяет до некоторой степени преодолеть противоречия между словарями интенсивного и экстенсивного типов, более точно отвечать требованиям разных категорий пользователей /переводчиков, технических специалистов, лексикографов, педагогов и издателей/, более экономно решать различные задачи, стоящие перед этими категориями пользователей. Девять вариантов ответа, которые может представить на выбор наш АПС, совместно с другими его возможностями позволяют рассматривать его как автоматизированную систему для решения некоторых лингвистических, педагогических и издательских задач. Например, первый вариант ответа предназначен в основном для автоматического перевода списочной документации. Вместе с тем этот же режим при незначительной модификации программного обеспечения позволяет использовать наш АПС как автоматический тренажер в педагогических целях для активизации и контроля за усвоением системы словоизменения родного и иностранного языков и лексического запаса иностранных переводных словарей отражает одно из перспективных направлений развития многоязычных автоматизированных систем /Rondeau, 1980/. Другие варианты ответа совместно с ограничениями по тематике и индексу надежности дают возможность использовать АПС ВЦП как средство частичной автоматизации лексикографической и издательской деятельности /см. разд. 2.7./.

2.6. Параметрический анализ автоматического переводного словаря Всесоюзного центра переводов

Лингвистическое обеспечение АПС ВЦП предполагает использование весьма разнообразной лингвистической информации, которая распределяется по разным лингвистическим алгоритмам и словарным статьям, причем ни в одной из частей лингвистического обеспечения эта информация не используется в полном объеме. Многоязычие словаря естественно требует учета специфики грамматического описания каждого конкретного языка. Все это привело нас к убеждению о целесообразности создания специального раздела, в котором следует в компактной форме перечислить все типы лингвистических сведений, используемых в данном АПС.

Состав лингвистических сведений, используемых в нашем словаре, предопределен теми задачами, которые мы поставили перед словарем. Однако формулировки типов этих сведений и сама форма их представления построены нами с учетом опыта работ Ю.Н. Караулова, который называет подобные лексикографические сведения параметрами /Караулов, 1981, с. 74-108/. Всего нами использовано для описания четырех языков 47 параметров. Многоязычие АПС ВЦП, а также ограниченные возможности этого пособия не позволяют представить параметры для всех языков, поэтому мы включили в пособие только параметрический анализ английской части АПС /см. Приложение № 2/.

Следует отметить, что лексикографируемые параметры представлены в нашем АПС не в одинаковой степени. Для одних параметров выделяются специальные зоны, которые обязательно заполняются для всех лексических единиц всех языков. К таким параметрам, например, относятся часть речи, тематическая группа, нормативный параметр. Для других также выделяются специальные зоны, однако они не являются обязательными для заполнения для всех ЛЕ, например толкование и экзemplарно-иллюстративный параметр. Третья группа параметров не имеет своих собственных зон, но информация данного типа может фиксироваться в зонах других параметров. Так, ареальный параметр не имеет своей зоны, но информация об ареальной принадлежности заглавной лексической единицы может включаться в зону стилистических помет, например "американизм" или "IBM", т.е. терминология фирмы

IBM, что, по нашему мнению, можно рассматривать как специфическое проявление ареальности терминологических единиц на современном уровне развития терминологических систем некоторых отраслей науки и техники.

Перечень лексикографических параметров, взятый нами за основу, был разработан для описания традиционных словарей, поэтому он оказался недостаточным для описания автоматического словаря. Для описания специфических свойств автоматических словарей автор предлагает десять дополнительных квазипараметров, которые приведены в Приложении № 2 под номерами 32-41. Строго говоря, они не являются лексикографическими параметрами в точном понимании этого термина, ибо предназначены не для описания фактов языка, а для того, чтобы отразить специфические особенности данного АПС как представителя особого класса автоматизированных лексикографических систем.

2.7. Использование автоматического переводного словаря как автоматизированной системы лексикографического обслуживания

Изменения по желанию пользователя состава и объема информации в ответе обуславливается в первую очередь тем, что словарная статья АПС ВЦП имеет четкую логическую структуру. Каждый вид информации занимает в словарной статье строго фиксированное место — так называемую зону. Разноязычные лексические единицы имеют обратимые связи между собой напрямую или через их русские ПЭ. Подобная организация словарной базы данных открывает широкие возможности для оптимизации труда лексикографа и издателя, например, при редактировании словарных массивов или при подготовке новых специализированных изданий.

Возможность изолированного обращения к любой информационной зоне позволяет осуществлять автоматический поиск словарных статей практически по любому отдельно взятому признаку или их комбинации, например наличию или отсутствию той или иной зоны в словарной статье, а также по конкретному варианту ее заполнения. Такой селективный подход существенно упрощает задачу лексикографа при проведении коррекции одновременно на больших массивах словаря. Например, он может проинформировать замену различных комбинаций символов, что

оказывается необходимым при изменении норм правописания или системы условных обозначений, принятых в словаре; может получить списки лексических единиц, представляющих собой стандартизованные или новые термины конкретной предметной области /в качестве поисковых признаков в данном случае выступают тематическая рубрика и индекс надежности/; может выбрать словарные статьи, в которых отсутствуют толкования, с тем чтобы впоследствии дополнить их в этом отношении, и т.п. При этом существенно, что автоматический словарь, в принципе, в состоянии вести учет числа обращений к словарной статье и вносимых в нее изменений, что позволяет следить за ходом редактирования лексических массивов /От автоматического словаря к..., 1987/.

На основе одного и того же исходного материала - разнородного множества словарных статей - могут быть автоматически получены различные конечные продукты. Например, можно генерировать двуязычные словари или одноязычные словники по определенным тематикам, по индексу надежности. При этом словарные статьи в таких словарях могут иметь по указанию пользователя различный состав информации. С помощью специальных сервисных программ можно вычленять из лексических массивов АПС словосочетания, содержащие ту или иную словоформу, с целью создания словарей словосочетаний или сортировки единиц по алфавитно-гнездовому принципу.

По мнению ряда лексикографов, основной резерв времени при создании двуязычных словарей лежит в возможности автоматически формировать "перевертыши" на основе созданных человеком исходных двуязычных массивов, т.е. менять местами входной и выходной языки словарных статей /Borsdorf, Gross, 1981/. Свойство обратимости нашего АПС позволяет автоматически получать также перевертыши" для любых имеющихся в его составе словарных статей. Правда, в результате мы получаем лишь своего рода полуфабрикат, который должен быть подвергнут тщательному редактированию. АПС выдает в "перевертыше" неупорядочный набор элементов для будущей словарной статьи, который человеку предстоит структурировать: объединить переводные эквиваленты в соответствии с выражаемыми ими значениями и ранжировать их определенным образом, убрать лишнюю или добавить любую необходимую информацию /например, пометы, примеры/. Автоматически полученные "перевер-

тыши" сами по себе представляют весьма любопытный объект лексикографических исследований, давая подчас примеры довольно неожиданных переводных соответствий. Полученные автоматически "перевертыши", на наш взгляд, могут послужить материалом для переработки исходных словарных статей как с целью устранения окказиональных переводных эквивалентов, так и, наоборот, с целью расширения состава исходной словарной статьи.

Динамичность АПС позволяет наблюдать за процессом его использования и устанавливать, таким образом, обратную связь между пользователями и лексикографами /Neubert, Kukuczka et.al., 1981/. АПС ВЦП позволяет осуществлять сбор разнообразной статистической информации в ходе его эксплуатации. Данные о частоте обращения к отдельным ЛЕ или целым классам их, например, по тематике несут довольно большую информационную нагрузку. Они могут свидетельствовать об употребительности данной ЛЕ и служить для лексикографа критерием для включения или невключения ее в очередное издание словаря. На их основе можно делать определенные выводы о тематике обрабатываемых текстов или об особенностях отдельных групп пользователей.

Важным побочным продуктом при эксплуатации АПС является массив так называемых неопознанных лексем, в который автоматически заносятся все запрошенные, но не найденные в словарной базе лексические единицы. Сюда кроме ошибочных вариантов написания и неверно выделенных словосочетаний попадут новые, еще не зафиксированные в лексикографической практике ЛЕ. Таким образом накапливается материал для пополнения лексических массивов самого АПС. Эти же лексические единицы могут быть изданы в виде Тетрадей новых терминов в традиционной книжной форме.

В настоящем разделе мы показали некоторые из основных направлений возможного использования АПС для автоматизации лексикографических работ. Следует отметить, что эта задача не является основной для АПС ВЦП и возникла по мере его создания. По нашему мнению, здесь проявляется современное понимание автоматизации в сфере интеллектуального труда - человек освобождается от рутинного, механического труда в пользу более содержательного, творческого, по-

лучая к тому же новую, ранее недоступную ему информацию об исследуемом объекте.

2.8. Реализация лингвистического обеспечения АПС ВЦП в автоматизированном рабочем месте переводчика

АПС ВЦП разрабатывается как типовая автоматизированная лексикографическая система, способная адаптироваться к разнообразным условиям его реализации. Этот фактор делает необходимым возможность его эксплуатации на ЭВМ разных типов. АПС ВЦП в настоящее время реализуется на ЭВМ трех типов: ЕС /не ниже 1033/, СМ /1420/ и персональной ЭВМ /ЕС-1832/. При этом естественно, что при реализации АПС на больших ЭВМ, на мини-ЭВМ и на персональных ЭВМ /ПЭВМ/ между ними происходит перераспределение функций. Образуется структура автоматических словарей. Полученную структуру можно рассматривать как распределенный лингвистический банк данных, реализуемый в виде иерархической структуры аппаратно-программных средств /Воржев, Ежов и др., 1985/.

Высший уровень такой структуры - многоязычный автоматический словарь, который фактически является многоязычным банком лингвистических данных. Принципы построения лингвистического обеспечения этого банка данных изложены в гл. II данного пособия. Весь многоязычный словарь в полном объеме реализуется на ЕС ЭВМ. Он позволяет формировать и корректировать терминологическо-лексикографическую базу, генерировать двуязычные словари на бумажных и машиночитаемых носителях, обеспечивать рабочие места лексикографов-терминологов на базе дисплеев ЕС ЭВМ. Второй уровень структуры - банк унификации терминов на базе мини-ЭВМ /СМ-1420/. Двуязычные словари этого банка можно либо формировать и корректировать автономно на мини-ЭВМ, либо генерировать на центральной ЕС ЭВМ. Третий уровень /персональный терминологический банк - основа автоматизированного рабочего места /АРМ/ переводчика реализуется на ПЭВМ с объемами оперативной памяти от 300 Кб и внешней памяти от 5 Мб.

Второй и третий уровни структуры позволяют вести персональные отраслевые словари и использовать их в переводческой деятельности, а также выполнять не-

которые дополнительные функции. При использовании ПЭВМ дисплей выступает в качестве пишущей машинки с памятью на несколько страниц текста, возможностью обращения в словарь для поиска незнакомых лексических единиц /слов и словосочетаний/, а также с полным набором редакторских функций.

Общая логика работы иерархической структуры сохраняется и при наличии только двух уровней: первого и второго или первого и третьего. Основная стратегия, принятая в настоящее время в ВЦП, ориентируется на использование больших ЕС ЭВМ и персональных ЭВМ.

2.8.1. Сфера и цели автоматизации перевода с помощью АРМ переводчика

В значительной степени эффективность АРМ переводчика определяется существенным повышением производительности труда переводчика за счет снятия или оптимизации некоторых трудоемких, но непроизводительных действий.

Важнейшими действиями переводчика являются словарные работы, выполнение перевода и его редактирование, учет и координация выполняемых переводов и - для некоторых категорий переводчиков - совершенствование языковой компетенции.

Словарные работы переводчика состоят из следующих действий:

- поиск в словарях переводных эквивалентов незнакомой ему лексики,
- ведение переводческих карточек новых терминов, сокращений и т.д.
- дополнение имеющихся словарей.

Выполнение перевода и его редактирование включают:

- изготовление первичного текста перевода,
- редактирование текста перевода,
- перепечатка отредактированного текста,

Для учета и координации выполненных переводов требуется:

- ведение картотеки выполненных переводов,
- поиск по картотеке выполненных переводов.

К совершенствованию языковой компетенции относятся:

- возможность контроля и самоконтроля знания лексики, орфографии и правил словоизменения рабочего языка.

Применение АРМ устраняет необходимость перепечатки перевода после редактирования и значительно уменьшает затраты на поиск по словарю. Отсутствие необходимости перепечатки переводов сокращает общее время выполнения перевода примерно на 10% /Рябцева, 1986, с. 74/. Кроме того, известно, что при использовании автоматического словаря для поиска незнакомых слов и словосочетаний производительность труда переводчика возрастает на 20 - 50% /Cogen, 1980; Bachrach, 1971; Язык и машины, 1968/. Таким образом, с использованием АРМ повышение эффективности переводческой деятельности может достигнуть 40 - 70%.

2.8.2. Функции АРМ переводчика и возможности его использования

Основное предназначение АРМ - оптимизация выполнения перевода, что и определяет его применение в качестве рабочего инструмента переводчика /как индивидуальных, так и объединенных в коллектив/ /Wright, 1984/.

Главные функции АРМ переводчика:

- ввод текста перевода с клавиатуры ПЭВМ,
- получение переводных эквивалентов запрашиваемых лексических единиц,
- обработка запроса в текстовой форме,
- редактирование текста,
- работа с полиэкраном, позволяющим одновременно видеть текст оригинала, переводные эквиваленты, текст перевода и другую дополнительную информацию /Brodsky, 1986/,
- варьирование объема и содержания словарных статей,
- полиграфическое оформление текста,
- учет и координация выполняемых переводов, в том числе и автоматическое определение объема выполненного перевода.

В АРМе переводчика предусмотрена также возможность ведения индивидуальных терминологических или иных словарей, которые затем можно слить как с используемым базовым словарем, так и с индивидуальными словарями других переводчиков-пользователей /Cogen,

1980; Melby, 1984; Казакевич, 1987/. При этом решаются следующие основные задачи:

- создание массива лексических единиц, отсутствующих в словарных массивах АРМ,
- редактирование и коррекция лексических массивов АРМ.

Поскольку АРМ является переводческой, а не лексикографической системой, в словарную статью АРМ, созданного в ВЦП, были отобраны следующие зоны из словарной статьи основного АПС:

- заглавная лексическая единица,
- лексико-грамматический класс заглавной ЛЕ,
- индекс надежности информации в словарной

статье,

- переводной эквивалент заглавной ЛЕ,
- толкование,
- контекстный пример,
- перевод примера,
- код тематики /предметной области/,
- сокращения заглавной ЛЕ и переводного эквивалента.

Грубая оценка средних длин указанных зон показывает, что на 20 тыс. таких словарных статей следует отвести около 6 Мб внешней памяти. Экономии объема внешней памяти можно получить сжатием текстовой части словарных статей /Воржев, Коростелев, 1983; Hamilton, 1980/, что позволит сократить объем памяти на внешнем носителе на 30%, т.е. на 20 тыс. словарных статей потребуется около 4 Мб внешней памяти.

АРМ переводчика, создаваемые в ВЦП на базе основного многоязычного АПС, являются двуязычными и в полной мере обладают свойствами гибкости, динамичности и обратимости. Важным параметром при оценке эффективности работы АРМ переводчика является среднее время обработки одного запроса. В настоящее время скорость обработки запроса составляет около двух секунд.

Лингвистическая информация, содержащаяся в словарном компоненте АРМ переводчика, позволяет обрабатывать запрос пользователя, заданный в текстовой и в словарной формах. Включение алгоритма лемматизации в структуру лингвистического обеспечения АРМ объясняется двумя причинами. Во-первых, пользователи автоматических словарей имеют различные уровни языковой компетентности. Определенная часть из них не об-

ладает достаточными знаниями иностранного языка, поэтому для них может представлять значительную трудность необходимость поставить незнакомое слово или словосочетание в словарную форму. Во-вторых, если переводчик-пользователь работает с текстом оригинала, уже имеющимся в памяти ЭВМ, он может только пометить незнакомую ЛЕ в этом тексте, т.е. запрос в текстовой форме в данном случае обусловлен самой технологией работы с системной АРМ. Наличие средств морфологического анализа делает также возможным использование АРМ как учебного средства для контроля знаний лексики, орфографии и морфологии рабочего языка.

Указанные выше функции АРМ переводчика представляют собой некоторый минимальный набор. Если же указывать набор перспективных функций, то в него следует включить еще и такие, как:

- возможность пользоваться несколькими наборами шрифтов на экране дисплея,
- возможность использования и корректировки отраслевого тезауруса,
- средства контроля орфографической правильности создаваемого текста,
- автоматизированный ввод текста оригинала в ЭВМ,
- вывод текста перевода на машиночитаемый носитель для непосредственной передачи его на фотонабор.

Эти дополнительные функции можно считать ориентирами для будущего развития систем типа АРМ переводчика.

2.8.3. Лингвистический блок в АРМ переводчика

Поскольку АРМ этого типа представляет собой переводческую автоматизированную систему, особое внимание было уделено оптимизации такого параметра, как разрешающая способность его словарного компонента. Ее повышению способствуют специальные лингвистические алгоритмы, в том числе: алгоритм словообразовательного анализа /позволяет обрабатывать словообразовательные неологизмы и окказионализмы; результатом его работы является выделение в производном слове одной или нескольких производящих основ и аффиксальных элементов и выдача семантической и переводческой информации для каждого из найденных компонентов/ и алгоритм об-

работки "некорректных" запросов /дает возможность разбивать запрошенное, но не найденное полностью словосочетание на составляющие; искать лексические единицы, частично пересекающиеся с запрошенной последовательностью, и т.п./ /Коростелев, Соколова и др., 1988/.

Лингвистическое обеспечение и программное обеспечение АРМ переводчика делают эту систему сравнительно простой и удобной для пользователя: пополнение и коррекция словарных массивов осуществляются в режиме диалога, причем большая часть операций автоматизирована, а человеку лишь приходится отвечать на отдельные вопросы, не предполагающие специальной лингвистической подготовки.

В заключение еще раз подчеркнем, что использование ПЭВМ для автоматизации переводческой деятельности представляет собой актуальную задачу, решение которой позволит повысить оперативность выполнения научно-технического перевода в нашей стране. При этом в идеальном варианте ЭВМ, на которой реализуется АРМ переводчика, будет соединяться с другими АРМами /Воржев, Игнатов и др. 1988/, системами машинного перевода /Коростелев, 1987/, информационными системами и сетями, библиотеками, базами данных и терминологическими банками для получения необходимой лексикографической информации.

ЗАКЛЮЧЕНИЕ

Основное отличие словаря автоматического от словаря традиционного заключается не в том, что автоматический словарь описывает языковые факты глубже и шире, чем традиционный словарь. В этом традиционные словари в настоящее время превосходят автоматические словари. Главное отличие и преимущество автоматического словаря заключается в том, что он дает пользователю такую степень свободы владения лексикографированным материалом, которая недоступна традиционным словарям, а это в первую очередь предопределяется его лингвистическим и программным обеспечением.

Лингвистическое обеспечение, по нашему мнению, является доминирующим компонентом автоматического словаря. Ведущая роль лингвистического компонента в автоматическом словаре объясняется тем, что лингвистические концепции, заложенные в АПС, глубина их проработки, конкретные пути решения лингвистических проблем предопределяют то, насколько создателям АПС удастся решить поставленные перед собой и своим словарем задачи. Разрабатывая лингвистическое обеспечение нашего АПС, мы исходили из того, чтобы он достаточно надежно решал поставленные перед ним задачи, был демократичен по отношению к пользователю, т.е. мог удовлетворять потребности широкого круга пользователей, имел возможности дальнейшего роста и был технологичен в разработке и эксплуатации.

Многоязычный АПС ВЦП фактически является автоматизированным банком лингвистических данных, включающим грамматические описания нескольких языков, массивы разноязычных лексических единиц, связанных между собой отношениями переводной эквивалентности, лингвистические алгоритмы. На основе многоязычного АПС можно генерировать двуязычные словари на традиционных и машиночитаемых носителях, создавать более компактные /с ограничениями по языкам, тематике, степени стандартизации лексики и т.п./ АПС с реализацией их на ЭВМ разных типов.

Растущая компьютеризация современного общества ведет к проникновению вычислительной техники во все сферы жизни общества и в лексикографию в том числе. ЭВМ как материальная база автоматического словаря расширяет его возможности и дает ему ряд серьезных преимуществ по сравнению с традиционными словарями.

В то же время ЭВМ не снимает ни одной из проблем, стоящих перед лексикографией, но заставляет взглянуть на них несколько с другой стороны и искать более оптимальные в новых условиях решения.

ОБРАЗЦЫ СЛОВАРНЫХ СТАТЕЙ
АВТОМАТИЧЕСКОГО ПЕРЕВОДНОГО СЛОВАРЯ
ВСЕСОЮЗНОГО ЦЕНТРА ПЕРЕВОДОВ

ПРИМЕР 1

FUNCTION(N)

F, FN, FUN, FUNC, FUNCTIONS

ОБЩ.*Б

1. ФУНКЦИЯ (Ж)

- - - 0035, 0041

ОБЩ.*Б

1. НАЗНАЧЕНИЕ (С)

НП: THE MAIN FUNCTION OF THE CONTROL INPUT IS TO
TELL THE PROCESSOR WHAT TRANSFORM TO PERFORM
ON THE INPUT IMAGE\ГЛАВНАЯ ФУНКЦИЯ
УПРАВЛЯЮЩЕГО ВХОДА - ДАТЬ УКАЗАНИЕ
ПРОЦЕССОРУ, КАКОЕ ПРЕОБРАЗОВАНИЕ СЛЕДУЕТ
ВЫПОЛНИТЬ НАД ВХОДНЫМ ИЗОБРАЖЕНИЕМ

- - - 0035, 0037

ОБЩ.*Б

1. ЦЕЛЬ (Ж)

НП: THE FUNCTION OF EDUCATION IS TO DEVELOP
MIND\ЦЕЛЬЮ ОБРАЗОВАНИЯ ЯВЛЯЕТСЯ РАЗВИТИЕ
УМСТВЕННЫХ СПОСОБНОСТЕЙ

- - - 0035

ОБЩ.*Б

2. РАБОТА (Ж, ЕД)

(ОБЫКН. РЛ)

- - - 0041

ОБЩ.*Б

2. ДЕЯТЕЛЬНОСТЬ (Ж, ЕД)

(ОБЫКН. РЛ)

- - - 0041

ОБЩ.*Б

2. ФУНКЦИОНИРОВАНИЕ (С, ЕД)

(ОБЫКН. РЛ)

НП: FUNCTIONS OF THE NERVES\ФУНКЦИОНИРОВАНИЕ
НЕРВНОЙ СИСТЕМЫ

- - - 0037

ОБЩ.*Б

2. ФУНКЦИИ (МН)

(ОБЫКН. РЛ)

ТО: ХАРАКТЕРИСТИКА
ПРЕДНАЗНАЧЕННОСТИ ЛИЦА

ФУНКЦИОНАЛЬНОЙ

- - - 0037

ОБЩ.*Б

2. ДОЛЖНОСТНЫЕ ОБЯЗАННОСТИ (ИМ. СОЧ. - МН)

(ОБЫКН. РЛ)

- - - 0029

ВЫЧ. СЕТИ, КОМПЛЕКСЫ*А

3. ФУНКЦИЯ (Ж)

ТО: ОБЪЕКТ, В РЕЗУЛЬТАТЕ ВЫЧИСЛЕНИЯ КОТОРОГО
ПОЛУЧАЕТСЯ ЗНАЧЕНИЕ ЗАДАННОГО ТИПА

- - - 0005

ВЧТ*Б

3. ОПЕРАЦИЯ (Ж)

НП: ADDITIONAL LOGIC CAN BE USED AT EACH TERMINAL
FOR SUCH FUNCTIONS AS ACKNOWLEDGING
RECEIVED DATA\НА КАЖДОМ КОНЦЕ СОЕДИНИТЕЛЬНОЙ
ЛИНИИ МОЖЕТ БЫТЬ ИСПОЛЬЗОВАНА ДОПОЛНИТЕЛЬНАЯ
ЛОГИКА, ПРЕДНАЗНАЧЕННАЯ ДЛЯ РЕАЛИЗАЦИИ ТАКИХ
ОПЕРАЦИЙ, КАК ВЫДАЧА СИГНАЛА ПОДТВЕРЖДЕНИЯ

- - - 0011

МАТЕМ.*Б

4. ФУНКЦИЯ (Ж)

НП: LINEAR FUNCTION\ЛИНЕЙНАЯ ФУНКЦИЯ

- - - 0020

МАТЕМ.*Б

4. ЗАВИСИМОСТЬ (Ж)

- - - 0020

ВЫЧ. СЕТИ, КОМПЛЕКСЫ*А

5. ФУНКЦИЯ УРОВНЯ (ИМ. СОЧ. - Ж)

ТО: В АРХИТЕКТУРЕ ОТКРЫТЫХ СИСТЕМ - ПЕРЕЧЕНЬ
ДЕЙСТВИЙ, ВЫПОЛНЯЕМЫХ В ПРЕДЕЛАХ УРОВНЯ С
ЦЕЛЮ ПРЕДОСТАВЛЕНИЯ ОПРЕДЕЛЕННЫХ УСЛУГ ПО
ПЕРЕДАЧЕ ДАННЫХ ОБЪЕКТОМ БОЛЕЕ ВЫСОКОГО
УРОВНЯ; ПРИМЕРАМИ ФУНКЦИЙ ЯВЛЯЮТСЯ:
АДРЕСАЦИЯ, МУЛЬТИПЛЕКСИРОВАНИЕ, ПЕРЕДАЧА
ДАННЫХ, СБОРКА И РАЗБОРКА БЛОКОВ И Т.Д.

- - - 0007

ВЧТ*Б

5. ФУНКЦИОНАЛЬНАЯ СХЕМА (ИМ. СОЧ. - Ж)

- - - 0011

ВЧТ*Б

5. ФУНКЦИОНАЛЬНЫЙ БЛОК (ИМ. СОЧ. -М)

НП: FAILURE RATES ARE CALCULATED BY CONVENTIONAL METHODS FOR EACH FUNCTION\ИНТЕНСИВНОСТЬ ОТКАЗОВ ВЫЧИСЛЯЕТСЯ ОБЫЧНЫМИ МЕТОДАМИ ДЛЯ ВСЕХ ФУНКЦИОНАЛЬНЫХ БЛОКОВ УСТРОЙСТВА

- - - 0011

 ПРИМЕР 2

АНГЛ. ЯЗ.: #COMPUTER (N) COMPUTERS
 ВЧТ*А
 1. ЭЛЕКТРОННО-ВЫЧИСЛИТЕЛЬНАЯ МАШИНА
 (ИМ. СОЧ. -Ж) ЭВМ&...#
 НЕМ. ЯЗ.: #FREI PROGRAMMIERTE RECHENANLAGE (NP-F)
 ВЧТ*Б
 1. ВМ С ГИБКОЙ ПРОГРАММОЙ (ИМ. СОЧ. -АББ)
 - - - 0007#
 АНГЛ. ЯЗ.: #WHISKER (N) WHISKERS
 ЭЛЕКТРОНИКА*В
 1. ЭЛЕКТРОННЫЙ ВЫВОД (ИМ. СОЧ. -М)
 - - - 0213
 ЭЛЕКТРОНИКА*В
 1. "УСИК" (М)
 - - - 0213
 ЭЛЕКТРОНИКА*В
 2. СПИРАЛЬНАЯ СТРУЖКА (ИМ. СОЧ. -Ж)
 - - - 0213
 ЭЛЕКТРОНИКА*В
 3. НИТЕВИДНЫЙ КРИСТАЛЛ (ИМ. СОЧ. -М)
 - - - 0213
 ЭЛЕКТРОНИКА*В
 4. УС (АББ)
 - - - 0213#

 ПРИМЕР 3

БАРЬЕР (0002)
 010000Б*
 ФРАН: .BARRIE>RE (0003)

БАЗА (0003)

010000Б*010000Б*

ФРАН: .BASE (0003)

АНГЛ: 1.BASE (0001)

ДЕЙСТВИЕ (0004)

000000Б*000000Б*010000Б*010000Б*

АНГЛ: .OPERATION (0001)

АНГЛ: 1.WORK (0001)

ФРАН: 2.MANOEUVRE (0003)

000000Б*000000Б*000000Б*000000Б*

НЕМЦ: .BETRIEB (0002)

НЕМЦ: 2.FUNKTION (0003)

ФРАН: 2.MARCHE (0003)

(1.УСТРОЙСТВО; 3.SG)

НП: METTRE EN MARCHE \ ВВЕСТИ В
ДЕЙСТВИЕ

ФРАН: 1.FONCTIONNEMENT (0011)

(1.ОРГАНИЗАЦИЯ)

ПРИМЕР 4

BAR 1 (N)

ОБЩ.*Б

1.БРУСОК (M)

- - - 0009

ОБЩ.*Б

2.КУСОК (M)

- - - 0009

BAR 2 (N)

ОБЩ.*Б

1.БАРЬЕР, ОТДЕЛЯЮЩИЙ СУДЕЙ (ИМ.СОЧ.-M)

- - - 0009

BAR 3 (N)

ОБЩ.*Б

1.БАР (M)

- - - 0009

1.БУФЕТ (M)

- - - 0009

ПРИМЕР 5

АНГЛ.ЯЗ.: LIGHT(N)
 LIGHT(VT\VI,IR)
 LIGHT(ADJ)
 ФРАНЦ.ЯЗ.: ANTI-E<BLOUISSANT(M)
 ANTI-E<BLOUISSANT(ADJ)
 НЕМ.ЯЗ.: ABKOMMEN(N)
 ABKOMMEN(VI,IR)

ПРИМЕР 6

АНГЛ.ЯЗ.: #FUNCTION(N)FUNCTIONS...#
 #FUNCTION(VI)FUNCTIONED,FUNCTIONING,
 FUNCTIONS...#
 ФРАНЦ.ЯЗ.: #ACCE<LE<ROCOMPTEUR(M)
 ACCE<LE<ROCOMPTEURS...#
 НЕМ.ЯЗ.: #BETRIEB(M)BETRIEBE,BETRIEBEN,
 BETRIEBS,BETRIEBES...#

ПРИМЕР 7

АНГЛ.ЯЗ.: #MUG(VI)MUGGING,MUGGED,MUGS
 ОБЩ.*Б
 1.<РАЗГ.>ГРИМАСНИЧАТЬ(НЕПЕР,НСВ)
 - - - 0009
 ОБЩ.*Б
 2.<ТЕАТР.>ПЕРЕИГРЫВАТЬ(НЕПЕР,НСВ)
 - - - 0009
 ОБЩ.*Б
 3.<АМЕР.>ФОТОГРАФИРОВАТЬ(ПЕР,НСВ)
 (ОСОБЕННО ПРЕСТУПНИКОВ для
 ПОЛИЦЕЙСКОГО АРХИВА)
 - - - 0009#

ПРИМЕР 8

НЕМ.ЯЗ.: #CHIP(M)CHIPS
 ВЧТ*Б

1.<АНГЛ.>ИНТЕГРАЛЬНАЯ СХЕМА(ИМ.СОЧ.-
Ж)
- - - 0025
ВЧТ*Б
1.<АНГЛ.>МИКРОСХЕМА(Ж)
- - - 0025...#

ПРИМЕР 9

ФРАНЦ.ЯЗ.: #AME<L IORER(VT)AME<L IOR 1 01
ОБЩ.*Б
1.УЛУЧШАТЬ(ПЕР,НСВ)
- - - 0004
ОБЩ.*Б
1.УЛУЧШИТЬ(ПЕР,СВ)
- - - 0004#

ПРИМЕР 10

АНГЛ.ЯЗ.: #FRAME(N)
ОБЩ.*Б
1.КОРПУС(М)
- - - 0009
ОБЩ.*Б
1.КАРКАС(М)
- - - 0009
,
.
.
ВЧТ*Б
5.(ПЕРЕВОДНОЙ ЭКВИВАЛЕНТ НЕ
ОПРЕДЕЛЕН)
ТО:ПРЕДСТАВЛЕНИЕ В НАГЛЯДНОЙ
ФОРМЕ ТЕКУЩИХ ВЫБОРОЧНЫХ
ДАНЫХ
- - - 0037
ВЧТ*Б
6.(ПЕРЕВОДНОЙ ЭКВИВАЛЕНТ НЕ
ОПРЕДЕЛЕН)
ТО:ОДНА ИЛИ НЕСКОЛЬКО СТРОК БИТ,
РАСПОЛОЖЕННЫЕ ПОПЕРЕК ЛЕНТЫ
- - 0042 ...#

ПРИМЕР 11

АНГЛ. ЯЗ.: #ADJUST (VT)
 ВЧТ*Б
 1. ПОДСТРАИВАТЬ (ПЕР, НСВ)
 (УСТРОЙСТВО)
 - - - 0014
 ВЧТ*Б
 1. ПОДСТРОИТЬ (ПЕР, СВ)
 (УСТРОЙСТВО)
 - - - 0011 ...#

ПРИМЕР 12

АНГЛ. ЯЗ.: #LISTING (N)
 ВЧТ*Б
 1. РАСПЕЧАТКА (Ж, ЕД)
 (ДЕЙСТВИЕ)
 - - - 0012
 ВЧТ*Б
 2. РАСПЕЧАТКА (Ж)
 (РЕЗУЛЬТАТ ДЕЙСТВИЯ)
 - - - 0012 ...#

ПРИМЕР 13

АНГЛ. ЯЗ.: #RUN (N)
 :
 ВЧТ*Б
 2. ПРОГОН (М)
 (ПРОГРАММЫ)
 - - - 0012
 ВЧТ*Б
 3. ЭКСПЛУАТАЦИЯ (Ж, ЕД)
 (МАШИНЫ)
 - - - 0012
 :
 ВЧТ*Б
 5. ОТРЕЗОК (М)
 (ВРЕМЕНИ)
 - - - 0011#

ПРИМЕР 14

АНГЛ. ЯЗ.: #ACCESS (N)
 ВЧТ*Б
 1. ДОСТУП (М)
 (К ЧЕМУ-ЛИБО)
 - - - 0011
 ;
 ОБР. ДАННЫХ*Б
 2. ВЫБОРКА (Ж)
 (ИЗ ЧЕГО-ЛИБО)
 - - - 0099 ...#

АНГЛ. ЯЗ. #FUNCTION (N)
 ;
 ОБЩ.*Б
 2. РАБОТА (Ж)
 (ОБЫКН. PL)
 - - - 0009
 ОБЩ.*Б
 2. ДЕЯТЕЛЬНОСТЬ (Ж, ЕД)
 (ОБЫКН. PL)
 - - - 0009 ...#

ПРИМЕР 15

АНГЛ. ЯЗ.: #MAIN STORAGE (NP-SB)
 ВЧТ*Б
 1. ОСНОВНАЯ ПАМЯТЬ (ИМ. СОЧ.-Ж)
 ТО: УНИВЕРСАЛЬНАЯ ПАМЯТЬ
 КОМПЬЮТЕРА. РЕГИСТРЫ,
 УЧАСТВУЮЩИЕ В ВЫПОЛНЕНИИ
 ОПЕРАЦИЙ, ОБЫЧНО ИМЕЮТ ПРЯМОЙ
 ДОСТУП К ОСНОВНОЙ ПАМЯТИ
 - - - 0012
 ВЧТ*Б
 2. ОСНОВНАЯ ПАМЯТЬ (ИМ. СОЧ.-Ж)
 ТО: ВСЯ ПРОГРАММНО-АДРЕСУЕМАЯ
 ПАМЯТЬ, НАХОДЯЩИЕСЯ В КОТОРОЙ
 ИНСТРУКЦИИ МОГУТ ВЫПОЛНЯТЬСЯ,
 А ДАННЫЕ ПОСТУПАТЬ
 НЕПОСРЕДСТВЕННО В РЕГИСТРЫ
 - - - 0012#

ПРИМЕР 16

АНГЛ. ЯЗ. #ACCESS(N)

:
ФЗ: ARBITRARY ACCESS
DIRECT ACCESS
DISPLAY ACCESS
IMMEDIATE ACCESS
INSTANTANEOUS ACCESS
MAGNETIC DRUM ACCESS
MAGNETIC TAPE ACCESS
MEMORY ACCESS
MULTIPLE ACCESS ...#

ПРИМЕР 17

РУССКИЙ ЯЗЫК

ПРИ(ПРЕФ)

ОБЩ.*Б

1. (ПЕРЕВОДНОЙ ЭКВИВАЛЕНТ НЕ ОПРЕДЕЛЕН)

(ПРИ+S/V=V)

НЕ: ПРИГОТОВИТЬ, ПРИДУМАТЬ, ПРИОБРЕСТИ

ТО: ЗНАЧЕНИЕ ДОВЕДЕНИЯ ДЕЙСТВИЯ ДО КОНЕЧНОЙ
ЦЕЛИ; ПОЛНОТЫ, ИСЧЕРПАННОСТИ ДЕЙСТВИЯ

ОБЩ.*Б

2. UP(ADV)

(ПРИ+S/V=V)

НП: ПРИЧАЛИТЬ\SAIL UP; ПРИПЛЫТЬ\SWIM UP

ОБЩ.*Б

2. DOWN(ADV)

(ПРИ+S/V=V)

НП: ПРИЗЕМЛИТЬСЯ\LAND DOWN

ТО: ЗНАЧЕНИЕ ДОСТИЖЕНИЯ КАКОГО-НИБУДЬ МЕСТА,
ПРИБЫТИЕ, СОЕДИНЕНИЕ

ОБЩ.*Б

3. EN(PREF)

(ПРИ+S/V=V)

НП: ПРИВИВАТЬ\ENGRAFT

ОБЩ.*Б

3. IM (PREF)
(ПРИ+S/V=V)
НП: ПРИВИВАТЬ\IMPLANT
ОБЩ.*Б
3. IN (PREF)
(ПРИ+S/V=V)
НП: ПРИВИВАТЬ\INOCULATE
ОБЩ.*Б
3. TO (PREF)
(ПРИ+S/V=V)
НП: ПРИСТАВАТЬ\ADHERE TO; ПРИЛИПАТЬ\STICK TO
ТО: ДОБАВЛЕНИЕ, СБЛИЖЕНИЕ, СКРЕПЛЕНИЕ
ОБЩ.*Б
4. A LITTLE (ADV)
(ПРИ+S/V=V)
НП: ПРИПОДНЯТЬ\LIFT A LITTLE
ОБЩ.*Б
4. SLIGHTLY (ADV)
(ПРИ+S/V=V)
НП: ПРИПОДНЯТЬ\LIFT SLIGHTLY
ОБЩ.*Б
4. MORE (ADV)
(ПРИ+S/V=V)
НП: ПРИСЫПАТЬ\ADD (SOME) MORE
ТО: ЗНАЧЕНИЕ НЕПОЛНОТЫ ДЕЙСТВИЯ
ОБЩ.*Б
5. IN (PREF)
(ПРИ+S/V=V)
НП: ПРИУМНОЖАТЬ\INCREASE
ОБЩ.*Б
5. EM (PREF)
(ПРИ+S/V=V)
НП: ПРИУКРАШИВАТЬ\EMBELLISH
ТО: ЗНАЧЕНИЕ ИНТЕНСИВНОСТИ ДЕЙСТВИЯ
ОБЩ.*Б
6. (ПЕРЕВОДНОЙ ЭКВИВАЛЕНТ НЕ ОПРЕДЕЛЕН)
(ПРИ+S/N=N; ПРИ+S/A=A)
НП: ПРИГОРОД, ПРИСТАНЬ, ПРИСТАНЦИОННЫЙ
ПРИОЗЕРНЫЙ, ПРИГОРОК

АНГЛИЙСКИЙ ЯЗЫК

OVER (PREF)
ОБЩ.*Б

1. ОБ (ПРОФ)

(OVER+N/A/V=N/A/V)

НП: OVERCAST\ОБШИВКА

ОБЩ. *Б

1. НАД (ПРЕД)

(OVER+N/A/V=N/A/V)

НП: OVERGROUND\НАДЗЕМНЫЙ

ОБЩ. *Б

1. ПЕРЕ (ПРЕФ)

(OVER+N/A/V=N/A/V)

НП: OVERLAP\ЧАСТИЧНО ПЕРЕКРЫВАТЬ

ТО: ПОЛОЖЕНИЕ НАД ИЛИ СВЕРХУ НЕКОТОРОГО
ОРИЕНТИРА

ОБЩ. *Б

2. ПЕРЕ (ПРЕФ)

(OVER+N/A/V=N/A/V)

НП: OVERBURDEN\ПЕРЕГРУЖАТЬ

ОБЩ. *Б

2. СВЕРХ (ПРЕФ)

(OVER+N/A/V=N/A/V)

НП: OVERTIME\СВЕРХУРОЧНЫЙ

ОБЩ. *Б

2. ПРЕ (ПРЕФ)

(OVER+N/A/V=N/A/V)

НП: OVERCOME\ПРЕОДОЛЕВАТЬ

ОБЩ. *Б

2. ПРО (ПРЕФ)

(OVER+N/A/V=N/A/V)

НП: OVERDUE\ПРОСРОЧЕННЫЙ

ТО: ПРЕВЫШЕНИЕ НОРМЫ ПО НЕКОТОРОМУ ПАРАМЕТРУ

НЕМЕЦКИЙ ЯЗЫК

SCHAFT (СУФ)

ОБЩ. *Б

1. СТВО (СУФ)

(S/N+SCHAFT=N; S/A+SCHAFT=N; S/V+SCHAFT=N)

НП: EIGENSCHAFT\СВОЙСТВО

ОБЩ. *Б

1. ОСТ (СУФ)

(S/N+SCHAFT=N; S/A+SCHAFT=N; S/V+SCHAFT=N)

НП: FEINDSCHAFT\ВРАЖДЕБНОСТЬ

ОБЩ. *Б

1. ЕНИ (СУФ)

(S/N+SCHAFT=N; S/A+SCHAFT=N; S/V+SCHAFT=N)

НП: ERRUNGENSCHAFT\ДОСТИЖЕНИЕ

ТО: ЗНАЧЕНИЕ СОСТОЯНИЯ, ВИДА ДЕЯТЕЛЬНОСТИ,
ВЫРАЖЕННОГО ПРОИЗВОДЯЩЕЙ ОСНОВОЙ

ОБЩ. *Б

2. ОСТ (СУФ)

(S/N+SCHAFT=N; S/A+SCHAFT=N)

НП: GEMEINSCHAFT\ОБЩНОСТЬ

ОБЩ. *Б

2. СТВО (СУФ)

(S/N+SCHAFT=N; S/A+SCHAFT=N)

НП: BAUERNSCHAFT\КРЕСТЬЯНСТВО

ТО: ЗНАЧЕНИЕ СОБИРАТЕЛЬНОСТИ

ФРАНЦУЗСКИЙ ЯЗЫК

EUR (СУФ)

ОБЩ. *Б

1. ТЕЛЬ (СУФ)

(S/V+EUR=N)

НП: SAVEUR\СПАСАТЕЛЬ

ОБЩ. *Б

1. ЧИК (СУФ)

(S/V+EUR=N)

НП: MITRAILLEUR\ПУЛЕТЕЧИК

ОБЩ. *Б

1. ЩИК (СУФ)

(S/V+EUR=N)

НП: REMOULEUR\ТОЧИЛЬЩИК

ОБЩ. *Б

1. ИСТ (СУФ)

(S/V+EUR=N)

НП: COMPTEUR\СЧЕТЧИК

ПАРАМЕТРИЧЕСКИЙ АНАЛИЗ АНГЛИЙСКОЙ ЧАСТИ
МНОГОЯЗЫЧНОГО АВТОМАТИЧЕСКОГО ПЕРЕВОДНОГО
СЛОВАРЯ ВСЕСОЮЗНОГО ЦЕНТРА ПЕРЕВОДОВ

ПАРАМЕТРЫ

Названия параметров пронумерованы. Способы задания и глубина лексикографирования конкретного параметра приведены непосредственно под его названием.

1. Языки

Английский

2. Входы в словарь: количество и характер

Прямой алфавитный. Заглавные единицы могут запрашиваться в любой текстовой форме. Сокращения с их расшифровкой. Далее поиск лексик может осуществляться от следующих параметров /зон словарной статьи/:

1. тематическая рубрика и подрубрика;
2. стиль;
3. индекс надежности;
4. лексико-грамматический класс;
5. толкование;
6. иллюстративный пример;
7. источник.

АПС позволяет осуществлять одиночное обращение /запрашивается одна ЛЕ/, а также групповое обращение /запрашивается сразу несколько ЛЕ/.

Технически обращение к АПС может осуществляться:

1. в режиме диалога "человек-ЭВМ";
2. в пакетном режиме, когда обращение к АПС /запрос/, поиск по АПС и выдача ответа разорваны по времени.

3. Хронологический параметр

Хронологические рамки специально не устанавливаются. Терминологическая лексика поступает в АПС из источников издания не ранее 1970 г. Косвенно хронологический параметр может проявляться в зоне индекса надежности /группа источников типа Г - новые термины/, а также в зоне стилистических помет при использовании пометы уст. /устаревшее слово/. Указывается дата ввода словарной статьи в АПС.

4. Количественный параметр /число словарных статей в словаре/

84 тыс. англо-русских словарных статей.

5. Орфографический параметр

Побуквенно. Описывается в зоне морфограмматической информации, в которой для английских лексических

единиц указываются все их словоформы и все варианты написания, например слитно-раздельный, американский.

6. Графическая длина слова /количество букв/
Неограниченная

7. Часть речи

Указывается в зоне лексико-грамматического класса; A - adjective; AP - adjective phrase; ABB - abbreviation; ADV - adverb; ADVP - adverb phrase; CJ - conjunction; N - noun; NP - noun phrase; PL - noun pluralia tantum; Sg - noun singularia tantum; NP-ABB - noun phrase with an abbreviation as the head word; CONST. - for unchangeable words e.g. "if-then"; PART - particle; PREP - preposition; PRON - pronoun; PREF - prefix; PREF N - noun prefix; PREF A - adjective prefix; PREF V - verbal prefix; PREF ADV - adverbial prefix; SUFF - suffix; SUFF N - noun suffix; SUFF A - adjective suffix; SUFF V - verbal suffix; SUFF ADV - adverbial suffix; VT - verb transitive regular; VT, IR - verb transitive irregular; VI - verb intransitive regular; VI, IR - verb intransitive irregular; VT/VI - verb transitive/intransitive regular; VT/VI, IR - verb transitive/intransitive irregular; VP, T - verbal phrase transitive; VP, I - verbal phrase irregular

8. Число

Pluralia, singularia tantum указываются в зоне лексико-грамматической информации

9. Степени сравнения прилагательных

В зоне лексико-грамматической информации указываются синтетические формы как словоформы заглавной ЛЕ.

10. Переходность

Указывается в зоне лексико-грамматической информации.

11. Управление глагола

Может указываться в зоне пояснительных помет.

12. Инфинитив

Используется в качестве словарной формы глагола,

13. Спряжение глагола /в наст. времени/

В зоне лексико-грамматической информации указываются формы 3-го лица настоящего времени и причастия настоящего времени /Participle I/.

14. Причастие /или основа прош. времени/

Указывается в зоне лексикограмматической информации: worked, run, gone

15. Прошедшее время

Указывается в зоне лексикограмматической информации: worked, ran, went

16. Морфологическое членение слова

Английские ЛЕ, включенные в АПС, морфологическому членению не подвергаются. Однако такому членению может подвергаться лексика, отсутствующая в АПС. Для этой цели предусмотрен алгоритм словообразовательного анализа. Словообразовательные элементы могут выступать в качестве заглавных единиц самостоятельных словарных статей.

17. Словообразовательный параметр

Дериваты могут включаться в АПС в качестве самостоятельных словарных статей либо обрабатываться с помощью алгоритма словообразовательного анализа.

18. Многозначность - однозначность

Отдельные значения обозначаются арабскими цифрами 1, 2, 3... сплошной нумерацией. Переводной эквивалент однозначной лексической единицы получает номер 1. Разные переводные эквиваленты в пределах одного и того же значения получают один и тот же номер.

19. Семантема /семантический эквивалент, дефиниция/

Для толкования выделена особая зона, которая не является обязательной для заполнения. Само толкование обычно заимствуется из традиционных лексикографических источников.

20. Ареальный параметр

Может эксплицироваться в зоне стилистических помет, например АМ. /американизм/, либо в зоне тематической рубрики указанием на принадлежность к терминологии определенной фирмы, например IBM - /терминология фирмы IBM/.

21. Синтагматический параметр /свободная сочетаемость/

Может указываться в зоне пояснительных помет для уточнения отдельных значений многозначных слов. Частично отражается в контекстных примерах. По запросу пользователя из массива АПС собираются все словосоче-

тания, в которые входит интересующее пользователя слово и которые оформлены как самостоятельные словарные статьи.

22. Фразеологический параметр /несвободная сочетаемость/

В данной версии АПС не лексикографируется. Фразеологические единицы могут фиксироваться в зоне сочетаемости.

23. Экземплярно-иллюстративный параметр /примеры употребления/

Для иллюстраций /контекстов/ выделена специальная зона в словарной статье, которая не является обязательной для заполнения. Контексты представлены краткими английскими предложениями или их фрагментами и могут иметь русский перевод. Контексты взяты из научно-технических текстов и даны не всем словам.

24. Стилистический параметр

Каждое значение потенциально может иметь стилистическую помету: проф. /профессиональная лексика/, разг. /разговорная лексика/ и др.

25. Нормативный параметр

В зоне индекса надежности указывается тип источника, откуда взята вся словарная статья или данный переводной эквивалент: А. Стандарт; Б. Авторитетный словарь; В. Тетрадь или любой другой лексикографический источник новых терминов; Г. Прочие источники.

26. Статистический /частотный/ параметр

Словарная статья не несет никакой статистической информации, однако с помощью сервисных программ можно получить следующие данные: 1. количество лексических единиц по грамматическим классам, т.е. сколько существительных, прилагательных, глаголов и т.д.; 2. количество цельнооформленных ЛЕ; 3. количество словосочетаний - двусловных, трехсловных и остальных.

27. Синонимы

Предусмотрена возможность фиксации синонимов в зоне лексической сочетаемости. Зона не является обязательной для заполнения.

28. Антонимы

Предусмотрена возможность фиксации антонимов в зоне лексической сочетаемости. Зона не является обязательной для заполнения.

29. Омонимы

Омонимы подаются в качестве самостоятельных словарных статей.

30. Терминологический параметр

В словарной статье выделена обязательная для заполнения специальная зона, в которой фиксируется принадлежность каждого переводного эквивалента к терминологии определенной области науки и техники: ВТ – вычислительная техника, АВ – авиация. Общеупотребительная лексика получает помету ОБЩ.

31. /Лингво-/ исторический параметр

Данный параметр отдельно не рассматривается. Может быть зафиксирован пометой, например уст. /устаревшее слово/ в зоне стилистических помет

32. Лексикографический параметр

Предусмотрена специальная зона для указания конкретного источника поступления данной словарной статьи или указания фамилии ее составителя. Зона обязательна для заполнения.

33. Аббревиатуры

Аббревиатуры фиксируются в зоне морфограмматической информации. Если составителям не известна расшифровка сокращения, то оно может выступать в качестве заголовка самостоятельной словарной статьи.

34. Коррекция ошибок пользователя при обращении к АПС

АПС имеет блок обработки "некорректных" запросов, который позволяет исправлять наиболее типичные ошибки пользователя при обращении к АПС: устранять необязательные элементы /предлоги, артикли и др./ в начале, середине и конце запроса; расчленять многословные запросы на отдельные слова.

35. Обратимость АПС /возможность произвольно выбирать язык запроса и язык ответа/

На сегодняшний день АПС обладает обратимостью при эксплуатации его в двуязычных вариантах с обязательным присутствием русского языка, т.е. реализуются следующие варианты АПС: англо-русский, русско-английский, французско-русский, русско-французский, немецко-русский, русско-немецкий.

36. Динамичность словника АПС

Имеется возможность оперативной коррекции словника АПС.

37. Динамичность состава информации в словарной статье

АПС имеет возможность осуществлять любую коррекцию по любой зоне словарной статьи.

38. Обработка лексики, отсутствующей в АПС

В АПС предусмотрена возможность анализа лексики, отсутствующей в АПС: а/ на уровне словосочетаний и фрагментов текста с помощью алгоритма обработки "некорректных" запросов; б/ на уровне отдельных слов с помощью блока словообразовательного анализа.

39. Количество лексических единиц, поиск которых может осуществляться в АПС одновременно

Неограниченное. Пользователь может запрашивать лексические единицы по одной /индивидуальный запрос/ и одновременно несколько /групповой запрос/.

40. Анализ обращений к АПС

АПС имеет возможность накапливать данные по количеству обращений к каждой лексической единице, а также собирать "отказы", т.е. лексику, запрошенную пользователем, но отсутствующую в АПС, с указанием числа запросов по каждой отсутствующей лексической единице.

41. Ответ /объем и состав информации в словарной статье, выдаваемой пользователю/

Объем ответа, т.е. объем информации в словарной статье запрошенной лексической единицы, по желанию пользователя может изменяться. Пользователю предлагаются следующие варианты ответа:

1. Только один /первый/ переводной эквивалент /ПЭ/;
2. Первый ПЭ + индекс его лексико-грамматического класса /ЛГК/;
3. Первый ПЭ+ЛГК+зона помет+зона контекстных примеров;
4. Первый ПЭ+ЛГК+зона помет+зона контекстных примеров+зона толкования;
5. Все переводные эквиваленты;
6. Все ПЭ + их ЛГК;

7. Все ПЭ + их ЛГК + их пометы + их контекстные примеры

8. Все ПЭ + их ЛГК + их пометы + их контекстные примеры + их толкования;

9. Все ПЭ + их ЛГК + их пометы + их контекстные примеры + их толкования + зона лексической сочетаемости.

На эти варианты ответа могут налагаться дополнительные ограничения по тематике и индексу надежности, т.е. пользователь может ограничить ответ лексикой только одной определенной области, например вычислительной техникой, и одним индексом надежности, например стандартом.

42. Форма ответа на групповой запрос

Ответ на групповой запрос может выдаваться в порядке поступления запрашиваемых лексических единиц или в алфавитном порядке.

43. Форма представления лексических массивов АПС

Лексические массивы многоязычного АПС могут быть представлены в виде распечатки на АЦПУ или на магнитном носителе в форме:

1. Двужычных иностранно-русских словарей;

2. Двужычных русско-иностраннх словарей.

На эти словари могут налагаться дополнительные ограничения по тематике и индексу надежности лексики входного языка.

3. В виде библиотек двужычных словарных статей, в которых содержится вся, в том числе и служебная информация в кодовой записи.

Общее число параметров, лексикографируемых в АПС для английского языка

Полностью - 25

Частично, косвенно или потенциально - 7

Всего - 32

Число квазипараметров - 10

ЛИТЕРАТУРА

Абрамова Н. Н., Бевзенко Е. А., Зеленков Ю. Г. Алгоритм нормализации именных словосочетаний // Вопросы информационной теории и практики. - М., 1985. - № 53. - С. 39-43.

Автоматизированная словарная служба ВИНТИ/ Белоногов Г. Г., Панова Н. С., Рыжова Е. Ю. и др. // Вопросы информационной теории и практики. - М., 1985. - № 53. - с. 5-9.

Автоматизированная словарно-терминологическая служба по системам связи/ Ахмеджанов М. Ф., Гельфман Г. С., Королев Э. И., Мальцев Е. П. // НТИ. Сер. 2 - 1984. - № 12. - С. 23-33.

Автоматическое обнаружение ошибок в словах и отождествление наименований понятий/ Котов Р. Г., Тимонин С. И., Седаков А. В., Юдин В. Г. // Лингвистические вопросы алгоритмической обработки сообщений. - М., 1983. - С. 212-232.

Алексеев П. М. Частотный англо-русский словарь-минимум по электронике. - М.: Военное издательство, 1971. - 301 с.

Анализ и обобщение результатов исследования тематических потоков научно-технических переводов, выполненных во Всесоюзном центре переводов научно-технической литературы и документации за период с 1У кв. 1984 г. по 1У кв. 1985 г.: Отчет о НИР/ВЦП; Ночкова Л. М. - ВЦП; № ГР 01.86.0 039080. - М., 1986. - 32 с.

Англо-русский частотный словарь по электронике/ Под ред. Убина И. И. - М.: ВЦП, 1977. - 207 с.

Андрющенко В. М. Вычислительная лексикография, ее возможности и перспективы // Вопросы языкознания. - 1986. - № 3. - С. 42-53.

Ануфриева И. А. Особенности французского языка // Рабочий проект многоязычного автоматического словаря на 60 тыс. словарных статей. Т. 1: Лингвистическое обеспечение. № ГР 81026333, Инв. № 0285.0052201. - М., 1984. - С. 76-82.

Архитектура лингвистических массивов АРМ переводчика / Коростелев Л. Ю., Соколова Е. Н., Убин И. И., Чаусова И. А. // Всесоюз. конф. "Совершенствование перевода научно-технической литературы и документов": Тез. докл. и сообщ. - М., 1988. - С. 231-233.

Бартков Б. И. Количественный дериваторий английского языка /300 аффиксов научного стиля и литературной нормы/. - Владивосток: ДВНЦ АН СССР, 1984. - 63 с.

Бектаев К. Б. Промышленный МП в режиме диалога человек-ЭВМ // Междунар. семинар по машинному переводу: Тез. докл. - М., 1983. - С. 35-36.

Берков В. П. Вопросы двуязычной лексикографии: Дис. д-ра филол. наук. - Л., 1971. - 637 с.

Берков В. П. Вопросы двуязычной лексикографии. - Л.: Изд-во ЛГУ, 1973. - 192 с.

Берков В. П. Слово в двуязычном словаре. Таллин: Валгус, 1977. - 140 с.

Воржев А. В., Коростелев Л. Ю. Методы кодирования текстовой информации для автоматизированных систем обработки естественного текста // Междунар. семинар по машинному переводу. - М.: ВЦП, 1983. - С. 61-62.

Герд А. С. Терминология - унификация или стандартизация? // НТИ. Сер. 2. - 1978. - № 4. - с. 1-4.

Герд А. С. Основы научно-технической лексикографии. - Л.: Изд-во ЛГУ, 1986. - 72 с.

Горбачевич К. С. Словарь и цитата /о рационализации иллюстрирования слов и значений во втором издании семнадцатитомного словаря/ // Вопросы языкознания. - 1978. - № 5. - С. 14-24.

Городецкий Б. Ю. Лексико-статистическая инвентаризация комплекса подязыков // Проблемы теоретической и экспериментальной лингвистики. - М., 1977. - С. 21-42.

Городецкий Б. Ю. Прикладное моделирование морфосемантического уровня ограниченного подъ-

языка // Особенности словообразования в терминосистемах и литературной норме. - Владивосток, 1983. - С. 15-31.

Действующие системы машинного перевода и автоматические словари / Под ред. В. Н. Герасимова. - М.: ВЦП, 1979. - 70 с.

Денисов П. Н. Об универсальной структуре словарной статьи // Актуальные проблемы учебной лексикографии. - М., 1977. - С. 205-225.

Здорных Т. А. Алгоритм обработки "некорректных" запросов. Алгоритм словообразовательного анализа // Рабочий проект многоязычного автоматического словаря на 60 тыс. словарных статей. Т. 1: Лингвистическое обеспечение. № ГР 81026333, Инв. № 0285. 0052201. - М., 1984. - С. 80-94.

Зубов А. В., Нехай О. А., Трибис Л. И. Экстралингвистические и лингвистические компоненты банка данных микрокомпьютера-переводчика // Проблемы внутренней динамики речевых норм. - Минск, 1982. - С. 191-198.

К вопросу о структуре лексических полей словарей по химии и химической технологии / Ибрагимова М. Б., Мельникова М. М., Потапов И. И., Соковикова Н. К., Ратникова М. Г. // Подготовка и использование научно-технических словарей в системе информационного обеспечения. - М., 1986. - С. 83-85.

Кад О. Проблемы перевода в свете теории коммуникации // Вопросы теории перевода в зарубежной лингвистике. - М., 1978. - С. 69-90.

Казакевич О. Д. Автоматизация лексикографических работ. Автоматические словари // Научно-техническая информация. Сер. 2: Информационные процессы и системы. - 1985, № 9. - с. 25-29.

Караулов Ю. Н. Общая и русская идеография. - М.: Наука, 1976. - 355 с.

Караулов Ю. Н. Лингвистическое конструирование и тезаурус литературного языка, - М.: Наука, 1981. - 366 с.

К о ж и н а М. Н. О речевой системности научного стиля сравнительно с некоторыми другими. - Пермь.: ПГУ, 1972. - 395 с.

К о м и с с а р о в В. Н. Слово о переводе. - М.: Междунар. отношения, 1973. - 215 с.

К о р о с т е л е в Л. Ю. Автоматизированное рабочее место переводчика и редактора: взаимодействие с системой машинного перевода // Перевод и автоматическая обработка текста. - М.: ИЯ АН СССР, 1987.- с. 69-73.

Л е в ы й И. Искусство перевода. - М.: Прогресс, 1974. - 397 с.

Лексическая основа русского языка: Комплексный учебный словарь / Под ред. В. В. М о р к о в к и н а. - М.: Рус. яз., 1984. - 1166 с.

М а р ч у к Ю. Н. Проблемы машинного перевода. - М.: Наука, 1983. - 232 с.

М а р ч у к Ю. Н. Методы моделирования перевода. - М.: Наука, 1985. - 200 с.

М о р к о в к и н В. В. О базовом лексикографическом знании // Учебники и словари в системе средств обучения русскому языку как иностранному. - М.: Рус. яз., 1986. - С. 102-117.

М к р т ч я н Г. А. О словарно-терминологическом обеспечении перевода // Междунар. семинар по машинному переводу: Тез. докл. - М., 1979. - С. 141-143.

М к р т ч я н Г. А., Н е л ю б и н Л. Л. О некоторых особенностях научно-технической терминологии / Междунар. семинар по машинному переводу: Тез. докл. - М., 1979. - С. 143-144.

Н е л ю б и н Л. Л. Перевод и прикладная лингвистика. - М.: Высш. школа, 1983. - 207 с.

Н о в о с е л о в А. П., Х о р о ш и л о в А. А. Алгоритмы автоматической нормализации слов // Вопросы информационной теории и практики. - М., 1985. - № 53. - С. 26-30.

О многоцелевом автоматическом словаре русского языка / Б е л я е в а Л. Н., К р и с е в и ч В. С.

Л и п н и ц к и й С. Ф., П и о т р о в с к и й Р. Г.
// Вопросы общей и прикладной лингвистики. - Минск,
1975. - С. 149-164.

От автоматического словаря - к автоматизированной системе лексикографического обслуживания / А н у ф р и е в а И. А., Е ж о в А. С., З д о р - н ы х Т. А., П о д б е л ь с к а я Р. А., С о - к о л о в а Е. Н., У б и н И. И. // Текст в языке и речевой деятельности / состав, перевод, автоматическая обработка/. - М.: ИЯ АН СССР: ВЦП: КСХИ, 1987. - С. 13-20.

П е р е р в а В. М. О принципах и проблемах отбора терминов и составления словника терминологических словарей // Проблематика определений терминов в словарях разных типов. - Л., 1976. - С. 190-204.

П и о т р о в с к и й Р. Г. Текст, машина, человек. - Л.: Наука, 1975. - 327 с.

П и о т р о в с к и й Р. Г. Инженерная лингвистика и теория языка. - Л.: Наука, 1979. - 111 с.

П и о т р о в с к и й Р. Г. Лингвистические автоматы и машинный фонд русского языка // Вопросы языкознания. - 1987. - № 4. - С. 69-73.

Применение локальной сети ПЭВМ для автоматизации перевода / В о р ж е в А. В., И г н а т о в С. Г., К о р о с т е л е в Л. Ю., О р е ш к о в П. А. // Всесоюз. конф. "Совершенствование перевода научно-технической литературы и документов". - М.: ВЦП, 1988. - с. 219-222.

П о п о в Э. В. Общение с ЭВМ на естественном языке. - М.: Наука, 1982. - 360 с.

Р е ц к е р Я. И. Теория перевода и переводческая практика. - М.: Междунар. отношения, 1974. - 216 с.

Р у с и н о в а Л. Н. О "хорошем" термине и "хорошей" терминологии // Научно-техническая терминология" Тез. докл. Всесоюз. конф. - М., 1983. - С. 77-79.

Русско-английский частотный словарь по электронике / Под общ. ред. И. И. У б и н а. - М.: ВЦП, 1977. - 235 с.

Рябцева Н. К. Информационные процессы и машинный перевод /лингвистический подход/. - М.: Наука, 1986, 165 с.

Скороходько Э. Ф., Стогний А. А. Некоторые вопросы создания банка терминов: лингвистический аспект // НТИ. Сер. 2. - 1986. - № 8. - С. 29-33.

Словарь словообразовательных элементов немецкого языка / Под рук. М. Д. Степановой. - М.: Рус. яз., 1979. - 536 с.

Слокум Дж. Перспективы развития машинного перевода - мнение американских специалистов // ТИЭР. - 1986. - Т. 74, вып. 7. - С. 86-98.

Смирницкий А. И. К вопросу о слове /проблема "отдельности слова"/ // История советского языкознания. Хрестоматия. - М.: Высш. школа, 1981. - С. 202-209.

Смирнов И. П. Основы научно-технического перевода. Организационные аспекты // Методическое пособие. 1987. - М.: ВЦП. - 83 с.

Сokolova Е. Н. Алгоритмы лемматизации для русского языка // Рабочий проект многоязычного автоматического словаря на 60 тыс. словарных статей. Т. 1: Лингвистическое обеспечение. № ГР 81026333, Инв. № 0285.0052201. - М., 1984. - С. 45-62.

Сokolova Е. Н., Убин И. И. Алгоритмы распознавания словоформ в автоматическом словаре // Теория и практика учебной лексикографии. - М., 1988. - С. 191-216.

Степанов Г. М. Проблема лексикографической обработки новых слов в специальных словарях // Общее и сопоставительное языкознание. - М., 1986. - С. 64-73.

Тихомиров Б. Д. Промышленные системы машинного перевода // Научно-технический перевод. - М., 1987. - С. 92-105.

Туовинен С. Пример организации терминологической работы и построение банка терминов в Финляндии / Междунар. конф. "Теория и практика научно-технического перевода": Тез. докл. - М., 1985. - С. 218-221.

У б и н И. И. Критерии выбора формы представления лексики в терминологических банках данных и автоматических словарях // Системный анализ научного текста. - Владивосток, 1984. - С. 167-172.

У б и н И. И. Перевод с помощью ЭВМ /автоматические словари и автоматические терминологические банки данных/: Обзорная информация. - М.: ВЦП, 1986. - 78 с.

У б и н И. И. Основные типологические свойства автоматических словарей // Переводът и научно-техническият прогрес. - София, 1988. - С. 14-23.

У з о н и П. О машинном словообразовательном анализе венгерских производных слов // Междунар. семинар по машинному переводу: Тез. докл. - М., 1983. - С. 233-236.

Ф е л ь д м а н Н. И. Об анализе смысловой структуры слова в двуязычных словарях // Лексикографический сборник. - М., 1957а. - Вып. 1. - С. 9-57.

Ф е л ь д м а н Н. И. О границах перевода в иноязычно-русских словарях // Лексикографический сборник. - М., 1957б. - Вып. 2. - С. 81-109.

Цели и задачи автоматического словаря ВЦП и ориентация на типы ЭВМ / В о р ж е в А. В., Е ж о в А. С., К и к о т ь А. И., П о д б е л ь - с к а я Р.А. // Междунар. конф. "Теория и практика научно-технического перевода". - М.: ВЦП, 1985. - С. 134-136.

Ш а й к е в и ч А. Я., У б и н И. И. Взгляд переводчика и научного сотрудника на переводной терминологический словарь // Теория и практика научно-технической лексикографии. - М. /В печати/.

Ш а й к е в и ч А. Я. Проблемы терминологической лексикографии: Обзорная информация. - М.: ВЦП, 1983. - 67 с.

Ш а й к е в и ч А. Я. Современное состояние научно-технической лексикографии: Обзорная информация. - М.: ВЦП, 1986. - 94 с.

Ш в е й ц е р А. Д. Перевод и лингвистика. - М.: Воен. изд-во Мин. обороны, 1973. - 280 с.

Шевчук В. Н. Производные военные термины в английском языке. - М.: Воен. изд-во, 1983. - 231 с.

Ширяев А. Ф. // Синхронный перевод. - М.: Изд-во МО СССР, 1979. - 183 с.

Язык и машины. Применение электронных вычислительных машин в переводе и в исследовании языка: Отчет наблюдательного комитета по автоматической обработке текстов // НТИ. Сер. 2. - 1968. - № 8. - С. 25-37.

Bachrach J. A. L'ordinateur au Service du Terminologie: Naitre ou Exclave // META. - 1971. - Vol. 16, N 1-2. - P. 115-115.

Barnes J. User perspective on computer-assisted translation for minority languages. // Computer and Translation-1987. - Vol. 2/3, P. 131-134.

Baudot J., Clas A., Gross M. Un Modèle de Mini-banque de Terminologie Bilingue. // META. - 1981. - Vol. 26, N 4. - P. 315-331.

Boitet Ch. Pros and Cons the pivot and Transfer Approach in Multilingual Machine Translation: New Directions in Machine Translation // Conference Proceedings. Budapest, 18-19, August, 1988. - Foris Publishers, Dordrecht/Providence, 1988, p. 78-93.

Borsdorf W., Gross H. Zur Zeitökonomie bei der Erarbeitung zweisprachiger Fachwörterbücher / Neubert C /Hrg./ // Rechnerunterstützung bei der Bearbeitung fachlexikalischer Probleme. - Leipzig, 1981. - S. 94-114.

Brodsky F. Computer aids to literary translation: description of an ideal system // The ATA Chronicle. - 1986. - Vol. X., N 7. - P. 1-4.

Cogen P. Le text-système 5823 au bureau Cogen // Siemens data contact. 1980. - N 8. - P. 5-7.

Dubuc R. TERMIUM System Description // META. - 1972. - Vol. 17, N 4. - P. 203-219.

Dubuc R., Gregoire J. F. Banque de terminologie et traduction // Babel. - 1974. - Vol. 20, N 3. - P. 180-184.

Felber H. Computerized Terminography in TernNet: the Role of Terminological Data Banks / Snell B. (Ed.) // Term Banks for Tomorrow's World. Translating and the computer: Proceedings of a Conference, 11-12 November 1982, London. - London, 1983. - P. 8-20.

G a l i n s k i Ch. Advanced Terminology Banks Supporting Knowledge-Based MT: New Directions in Machine Translation // Conference Proceedings, Budapest, 18-19 August, 1988. - Foris Publishers. Dordrecht/ Providence, 1988. - P. 150-164.

G o e t s c h a l c k x J. Terminological Activities in the European Institutions, with Special Reference to EURODICATOM // Third European Congress on Information Systems and Networks: Overcoming the language barrier, Luxemburg, 3-6 May, 1977. - London, etc., 1978 - Vol. 1. - P. 123-152.

G o l d D. L. Problems in Interlingual Lexicography // Babel. - 1978. - Vol. 24, N 3-4. - P.161-168.

H a m i l t o n O. A., H e r r o l d P. P., O s s e f o r t M. J. Digrammatic text compression // IBM Technical disclosure bulletin. - 1980. V. 23, N.2. - P. 448-452.

H ö p p n e r W. Begleitinformationen als notwendige Elemente der Terminologiedokumentation // Междунар. конф. "Теория и практика научно-технического перевода": Тез. докл. - М., 1985. - С. 183-184.

H u t c h i n s W. J. Recent Development in Machine Translation: New Directions in Machine Translation // Conference Proceedings. Budapest, 18-19 August. 1988. - Foris Publishers. Dordrecht / Providence, 1988. - P. 1-57.

K a k i z a k i N. Research and development at an electronic dictionary: current status and future plan // Machine Translation Summit (manuscripts program). - Hakone, Japan, 1987. - P. 67-64.

K r o l l m a n n F. Linguistic Data Banks and the Technical Translator // META. - 1971. - Vol. 16, N 1-2. - P. 117-124.

K r o l l m a n n F. User Aspects of an Automatic Aid to Translation as Employed in a Large Translation Service // Third European Congress on Information Systems and Networks: Overcoming the Language Barrier. Luxemburg, 3-6 May, 1977. - London, etc., 1978. - P. 243-257.

K r o l l m a n n F. Computer-Aids to Translation // META. - 1981. - Vol. XXVI, N 1. - P. 85-94.

K r o l l m a n n F., S c h u c k H. J., W i n k e r U. Herstellung textbezogener Fachwortlisten mit einem Digitalrechnerin Verfahren der automatischen Übersetzungshilfe // Beiträge zur Sprachkunde und Informationsverarbeitung. - 1965. - N 5. - S. 7-30.

L i p p m a n n E. O. On-Line Generation of Terminological Digests in Language Translation. An Aid in Terminology Processing // IEEE: Transactions on Professional Communication. - 1975. - Vol. 18, N 4. - P. 309-319.

M e l b y A. K. On-line terminology: options for communications // Silver tongues. - N.V., 1984.- P. 139-141.

N e u b e r t G. Zum Wiederauffinden von Informationen in automatischen Wörterbüchern - COLING-82 // Proceedings at the Ninth International Conference on Computational Linguistics. Prague, 1982. - Prague, 1982. - P. 259-264.

N e u b e r t G. Bildungselemente deutscher Fachwörter der Technik und ihre Verwendung in einem Analysealgorithmus // Междунар. семинар по машинному переводу: Тез. докл. - М., 1983. - С. 164-168.

N e u b e r t G., K u k u c z k a H., M e y e r E. Das Datenverwaltungssystem für Fachwortschätze EWE / N e u b e r t G. (Hrg.) // Rechnerunterstützung bei der Bearbeitung fachlexikalischer Probleme. - Leipzig, 1981. - S. 9-70.

New Directions in Machine Translation // Conference proceedings, Budapest 18-19 August, 1988. - Foris Publishers. Dordrecht / Providence, 1988.- 224 p.

O u b i n e I. I., T i k h o m i r o v B. D. Machine Translation System and Computer Dictionaries in the Information Service, Ways of their Development and Operation: Coling - 82 // Proc. of the 9th Int. Conf. on Computational Linguistics, Prague, July 5-10. 1982. - Prague, 1982. - P. 289-294.

R e i c h l i n g A. EURODICATOM. La banque de données terminologiques de la Commission des Communautés Européennes // Bulletin terminologique.- 1977. - N. 30. - P. 7-18.

R o l l i n g L. N. Computer-Aided Translation Today and Tomorrow // New Trends in Documentation and Information Processing 39 FID. Congress, Edinburg, 1978. - London, 1980. - P. 164-169.

R o n d e a u G. Typologie des banques de terminologie // InfoTerm series 5. Terminological data banks: Proceedings of the First International Conference. - München; New York; London; Paris, 1980. - P. 22-33.

R o u s s e a u L. J. La normalisation terminologique a l'office de la langue française // dans Colloque international "Problèmes théoriques et méthodologiques de la terminologie". - München, 1981. - P. 320-333.

S a g e r J. C., M c N a u g h t J. Selective Survey of Existing Linguistic Data Banks in Europe// British Library Research and Development Department Report. - 1980. - N 5643. - 75 p.

S c h u l z J. Le système TEAM, une aide à la traduction. // META. - 1971. - Vol. 1, N 1-2. - P. 125-131.

S c h u l z J. A terminology data bank for translators: Methods of interrogation in the TEAM system // Third European Congress on Information Systems and Networks: Overcoming the Language Barrier, Luxembourg, 3-6 May. 1977. - London. etc., 1978. - P. 153-188.

S c h u l z J. A terminology Data Bank for Translators (TEAM) // META - 1980. - Vol. XXV, N 2. - P. 211-299.

S c h u l z J., G ö r i c k e H. The Dictionary in the Computer. - Babel. - 1977, N 23. - P. 29-33.

S c h w a b W. Traduction et informatique: perspectives pour les années 80 // META. - 1981. - Vol. XXVI, N. 1. - P. 48-55.

S h a i k e v i c h A., O u b i n e I. Translators and researchers look at bilingual terminological dictionaries // Babel. - 1988. - Vol. 34, No. 1. - P. 10-16.

S l o c u m J. A Survey of Machine Translation: its History, Current Status, and Future Prospects // Computational Linguistics. - 1985. - Vol. 11, N 1. - P. 1-17.

S t r a z d s A., M c C r a c k e n D. TARGET: Research in Computer aids for Translators // Babel. - 1980. - Vol. 26, N 2. - P. 83-92.

T a n k e E. Intracompany and Intercompany Networking in Terminology TEAM // Abstracts Second InfoTerm Symposium. - Vienna: InfoTerm, 1985. - P. 25.

V a n S l y p e G., G u i n e t J. F., S e i t z F., B e n e j a n E. Better translation for better communication. - Luxembourg: Pergamon Press, 1983. - 230 p.

W e l l s W. The TERMIUM Cooperative // Abstracts Second Infoterm Symposium. - Vienna: InfoTerm. 1985. - P. 27.

W r i g h t L. D. Guidelines for standardizing the compilation of personal on-line glossaries // Silver tonques. - N.Y., 1984. - P. 143-148.

Z e m b J. M. Introduction by the Chairman of session 5 // Third European Congress on Information Systems and Networks: Overcoming the Language Barrier. Luxemburg, 3-6 May, 1977. - London, etc., 1978.

СО Д Е Р Ж А Н И Е

	стр.
ВВЕДЕНИЕ	3
Глава I. ОПРЕДЕЛЕНИЕ ОСНОВНЫХ ЭКСПЛУАТАЦИОННЫХ ХАРАКТЕРИСТИК АВТОМАТИЧЕСКОГО ПЕРЕВОД- НОГО СЛОВАРЯ НА ОСНОВЕ АНАЛИЗА ПРОИЗ- ВОДСТВЕННОЙ СФЕРЫ НАУЧНО-ТЕХНИЧЕСКОГО ПЕРЕВОДА	5
1.1. Место автоматического переводно- го словаря в процессе перевода .	5
1.2. Краткая характеристика Всесоюз- ного центра переводов	8
1.2.1. Анализ потока переводов по языкам и тематикам ...	8
1.2.2. Типы материалов, посту- павших на перевод	11
1.3. Необходимость автоматизации пе- ревода в специализированной пе- реводческой организации	12
1.3.1. Необходимость создания автоматического словаря для переводческой орга- низации	12
1.3.2. Модели основных категорий пользователей автомати- ческого переводного слова- ря	14
1.3.3. Цели и задачи автоматичес- кого переводного словаря .	19
1.4. Режимы эксплуатации автоматичес- кого переводного словаря	21
1.5. Многоязычие автоматического пе- реводного словаря	23
1.6. Обратимость автоматического пере- водного словаря	25
1.7. Динамичность и гибкость АПС	30
1.8. Выводы	33
Глава II. ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ АВТОМАТИЧЕ- СКОГО ПЕРЕВОДНОГО СЛОВАРЯ ВЦП	35
2.1. Состав и объем лексических единиц АПС ВЦП	35

2.1.1.	Типы самостоятельных единиц состава АПС	35
2.1.2.	Распределение лексического состава по частям речи ...	39
2.1.3.	Тематическая принадлежность лексических единиц .	41
2.1.4.	Новая лексика	42
2.1.5.	Индекс надежности	46
2.1.6.	Требование эквивалентности разноязычных лексических массивов в многоязычном автоматическом словаре	50
2.1.7.	Динамичность состава словаря	54
2.2.	Структура словарной статьи АПС ВЦП	54
2.3.	Описание многоязычных лексических единиц	63
2.3.1.	Принципы разграничения многоязычности	63
2.3.2.	Средства разграничения многозначности и омонимии в автоматическом словаре ВЦП	65
2.3.3.	Ранжирование значений в словарной статье многоязычного слова	72
2.4.	Запрос. Обращение к АПС	74
2.4.1.	Технология осуществления запроса	74
2.4.2.	Лингвистическая форма запроса	76
2.4.3.	Алгоритмы лемматизации ...	77
2.4.4.	Алгоритм обработки "некорректных" запросов	87
2.4.5.	Алгоритм словообразовательного анализа	88
2.5.	Ответ	93
2.5.1.	Лингвистическая форма ответа и способы его воспроизведения	93
2.5.2.	Форма представления ответа	94
2.5.3.	Объем и состав информации в ответе	96

2.6.	Параметрический анализ автоматического переводного словаря Всесоюзного центра переводов	98
2.7.	Использование автоматического переводного словаря как автоматизированной системы лексикографического обслуживания	99
2.8.	Реализация лингвистического обеспечения АПС ВЦП в автоматизированном рабочем месте переводчика	102
2.8.1.	Сфера и цели автоматизации перевода с помощью АРМ переводчика	103
2.8.2.	Функции АРМ переводчика и возможности его использования	104
2.8.3.	Лингвистический блок в АРМ переводчика	106
ЗАКЛЮЧЕНИЕ		108
ПРИЛОЖЕНИЕ 1.	Образцы словарных статей автоматического переводного словаря Всесоюзного центра переводов ...	110
ПРИЛОЖЕНИЕ 2.	Параметрический анализ английской части многоязычного автоматического переводного словаря Всесоюзного центра переводов	122
ЛИТЕРАТУРА		129

канд. филол. наук Иван Иванович УБИН

АВТОМАТИЧЕСКИЙ ПЕРЕВОДНОЙ СЛОВАРЬ.

ПРИНЦИПЫ ПОСТРОЕНИЯ

/Лингвистические концепции/

Ответственный редактор

к.т.н. Иванов В.А.

Рецензент

д.ф.н. Шайкевич А.Я.

Технический редактор

Н.К. Дудова

Корректор Г.Н. Игнатьева

Подп. в печать 24.01.89.	Формат 60x84/16.
Бум. офс. № 2. Печать офсетная.	Усл.печ.л. 8,37.
Усл.кр.-отт. 8,56. Уч.-изд.л. 6,92.	Зак. № 616
Тираж 700 экз.	Цена 1 р. 40 к.

Всесоюзный центр переводов научно-технической
литературы и документации
117218, Москва, В-218, ул.Кржижановского, д.14, корп.1

ПИК ВИНТИ, 140010, Люберцы-10, Моск. обл.,
Октябрьский пр., 403

