

УДК 801.1+801.7

А. В. РАССАДНИКОВА

## ПРИНЦИПЫ ПОСТРОЕНИЯ ДВУЯЗЫЧНОГО УНИВЕРСАЛЬНОГО ЭЛЕКТРОННОГО СЛОВАРЯ

Создание электронных словарей рассматривается как одна из основных проблем компьютерной лингвистики. С другой стороны, одной из технологически удобных и коммерчески перспективных областей внедрения компьютерной техники неизменно остается лингвистика, особенно двуязычная лексикография. На Западе во впечатляющей прогрессии растет количество электронных словарных массивов; весьма многочисленны лексические базы данных на компакт-дисках, заметная часть которых успешно реализуется на рынке. Ожидается, что в развитых странах к началу следующего тысячелетия около 1/4 книг будет электронной. По мнению авторитетных издателей США, справочники, словари, энциклопедии в самом скором времени будут храниться только на безбумажных носителях.

Большим преимуществом электронной лексикографии по сравнению с традиционной является оптимальное сочетание огромных объемов лексической информации с высокими актуализационными и издательскими возможностями. Трудности заключаются в основном в необходимости усовершенствования системы обработки источниковой лексической информации для оперативного отслеживания неологизмов и актуализации машинных лексикографических массивов.

Поскольку отечественная лингвистическая практика столкнулась с проблемой формализации лексической системы украинского языка для ее сопряжения с развитыми западными аналогами и обеспечения эффективного межъязыкового перекодирования текстовой информации, необходимость внедрения языковой лексикографии в Украине ощущается особенно остро. Создание новых двуязычных электронных словарей с участием украиноязычного звена позволило бы организовать оперативный выпуск столь необходимых полноценных двуязычных словарей для широкой рыночной реализации и существенно облегчить межнациональные языковые контакты. В отечественной лексикографии уже идет поиск концепции решения указанной проблемы. В этом отношении весьма эффективным представляется использование опыта разработки и практического внедрения двуязычных электронных словарей, накопленного в развитых западных странах, а также в России.

В частности, в российской лексикографической практике выделяется две интересные разработки: создание машинного фонда русского языка и создание системы электронных словарей для полумашинного перевода.

Принципы построения машинного фонда русского языка, формируемого под руководством специалистов Института русского языка Российской Академии наук, могут быть использованы в качестве прототипной базы для машинного упорядочения лексической системы украинского языка в целях ее последующего эффективного вывода на билингвистический уровень. Публикации разработчиков машинного фонда русского языка уже довольно многочисленны [1].

Концептуальный анализ лингвистических и машинных аспектов российской системы автоматических словарей для полумашинного перевода научно-технических текстов [2] также может оказаться полезным для отечественной электронной лексикографии, хотя следует подчеркнуть что уровень западных разработок в этой области намного выше. Заметим, что в новейшей российской лексикографии преобладают успехи в создании крупных двуязычных словарей на бумажных носителях, в то время как западная лексикографическая практика развивается преимущественно в направлении безбумажной информатики.

Динамика объемов крупнейших традиционных словарей с англоязычным звеном, изданных в последние годы в России, представлена в таблице.

Библиографические сведения	Англо-русский словарь		Русско-английский Политехнический словарь	
	70—80-е гг.	90-е гг.	70—80-е гг.	90-е гг.
Год первого издания (начала издания)	1972	1993	1980	(1996)
Число томов	2	3	1	4
Количество слов (терминов), тыс.	150	250	(90)	(>500)

Несмотря на внушительный рост объемов и несомненное повышение общего лексикографического уровня, новейшие традиционные двуязычные словари, изданные в России, содержат лексикографические лакуны, имеют недостаточный спектр проработки многозначных слов и, естественно, не поддаются текущей актуализации в период межпубликационного периода, а

также не позволяют получать удобных форматных распечаток в сетевых режимах работы, что заметно снижает неоспоримую информационную ценность данных лексикографических произведений.

Таким образом, как в России, так и в Украине существует проблема создания двуязычных универсальных электронных словарей. Здесь необходимо подчеркнуть два существенных, концептуальных обстоятельства. Во-первых, универсальность проектируемого двуязычного электронного словаря предполагает слияние лексикографических полей общей и специальной лексики языка в единое семантически формализованное поле, приближающееся к реальному совокупному лексическому узусу охватываемой пары языков. Такое монолитное лексикографическое произведение будет обладать повышенной лексической наполняемостью, принципиально отличаясь от отдельных и распыленных по отраслевому признаку двуязычных словарей прежних поколений. Это позволит значительно оперативнее решать задачи лингвистического перекодирования информации. Во-вторых, данный гипотетический двуязычный универсальный электронный словарь будет ориентирован как непосредственно на полумашинный перевод, так и на создание мощной лексикографической базы для совершенствования систем машинного перевода. Известно, что даже при наличии в развитых странах многочисленных разработок в технологии машинного перевода последний остается лишь подспорьем в межъязыковых системах перекодирования. Реальной базой для межнациональных контактов на лингвистическом уровне продолжает устойчиво оставаться живой (человеческий) перевод. Однако его процесс в цивилизованных странах интенсивно перестраивается на основе использования электронных словарей, приобретая отчетливые черты полумашинной технологии.

Перспективу создания систем двуязычных универсальных электронных словарей нового поколения в развитых и развивающихся странах с обеспечением сетевого межнационального взаимодействия этих систем на унифицированной форматной основе можно с уверенностью рассматривать как кардинальный лексикографический прорыв начала будущего тысячелетия. Активное участие основных славяноязычных звеньев (особенно русско- и украиноязычного) в разработке этих уникальных систем является вопросом национального престижа.

Попытаемся кратко проанализировать практические цели, преимущества, источники формирования, лингвистические и программные принципы построения, например, славяноязычно-английских (русско-английского украинско-английского и др.) универсальных электронных словарей. Практические цели заключаются в создании мощной формализованной лексикографической базы для оперативного перекодирования текстовой информа-

дии с русского (украинского) языка на английский и в коммерческом использовании базы как на основе информационного обслуживания потребителей, так и путем рыночной реализации целевых распечаток отдельных блоков информационного массива (например, в виде отраслевых словарей) на электронных или бумажных носителях. Преимущества разрабатываемого словаря состоят в лексической универсальности информации, в увеличении ее объема (в идеале с приближением к реальному лексическому узусу), оптимальном режиме актуализации лексической информации, возможности сетевого сопряжения с ее зарубежными электронными источниками, структурной простоте формата и поэлементном использовании отраслевых помет, ориентирующем потребителя на фактографические источники детализации узкоспециальных слов. К источникам формирования словаря относятся отраслевые электронные словари, произведения традиционной двуязычной лексикографии, фактографические источники, содержащие элементы перевода, толкования и номинации понятий, продукты переводческого процесса и дедуктивное расширение лексической информации.

Основными лингвистическими принципами построения данного электронного словаря являются: двуязычность, универсальность, семантическая структура, маркирование и актуализация. Попытаемся кратко их охарактеризовать.

Очевидно, лексикографическая стратегия должна включать в себя естественное движение от простого к сложному, которое принципиально сводится к следующему генерационному ряду: одноязычность → двуязычность → многоязычность. В данном случае это принципиальное положение предусматривает предварительную тщательную подготовку лексикографических баз славяно- и англоязычного звеньев. На практике это означает: наличие солидной машинной базы входного языка (в России, как указывалось, она именуется машинным фондом русского языка), так как формализованная англоязычная лингвистическая база развита достаточно хорошо; последующее сопряжение лексикографических массивов обеих баз. Заметим попутно, что формирование многоязычных универсальных электронных словарей на указанной основе, очевидно, представляет собой еще более сложную задачу.

Универсальность словаря обуславливается оригинальной разработкой входного славяноязычного словника, объединяющего общую (общепотребительную, общелитературную) лексику и специальную: преимущественно термины, а также значительную часть номенклатурных названий, торговых (фирменных) названий, специальных сокращений и других знаковых единиц, употребляемых в специальных текстах и имеющих лексическую или частично лексическую природу (лексические ктематонимы и т.п.).

Гнездовая описательная интерпретация терминов, характерная для традиционных специальных словарей, не предусмотрена, однако отсутствие таких обычно семантически прозрачных лексикографических описаний компенсируется несвойственной прежним словарям повышенной полиэквивалентностью лексикографической интерпретации входных слов.

Семантической структуре данного словаря будут присущи две взаимосвязанные особенности: наличие повышенной полиэквивалентности описания входных слов и диктуемое этим обстоятельством применение семантических ключей многозначных слов. В результате формируемая общая структура словаря будет отличаться существенной новизной.

Повышенная лексическая насыщенность словаря будет достигнута путем разработки развитой системы лексикографического маркирования в виде многочисленных специальных отраслевых помет и обычных лингвистических помет, которые могут быть сведены в общий вводный алфавитный список.

Машинная природа словаря обеспечивает возможность его оперативной текущей актуализации по источникам формирования. При этом особую генеративную ценность представляют выборки эквивалентов новой специальной лексики, формируемые в результате переводческого процесса. Они характеризуются исследователями в области лексикографии и переводоведения как "билингвистическая неологическая специальная лексическая информация" [3].

Уточненная интерпретация программных принципов построения двуязычного универсального электронного словаря в значительной мере будет зависеть от реального процесса увязывания его формата с форматом развитых сетевых лексикографических систем, функционирующих по каналам Интернет и Термнет или иным специальным каналам. Естественно, этот процесс может возникнуть лишь после окончания разработки словаря. Здесь можно указать лишь на общие положения, касающиеся программного обеспечения словаря. Идеальный словарь должен быть большим по объему, но давать возможность достаточно быстрого поиска слова. Следует обеспечить наличие простого и понятного интерфейса, но в то же время допускать широкие возможности варьирования форм представления запросов. Из различных требований к организации словаря с помощью ЭВМ особое значение приобретают защита данных от несанкционированного доступа, т.е. от возможностей разрушения и искажения записей. Необходимо также позаботиться о разработке механизма расширения словаря за счет новых терминов, причем новые сведения следует извлекать из источников, заслуживающих доверия, в связи с чем можно ввести индекс качества вводимой информации. Особое внимание нужно обратить на качество подачи словар-

ной информации, которая должна отражать связи между словами и их значениями. Требуется наличие развитой справочной системы для пользователя словаря. Чтобы облегчить разработку программного обеспечения для универсального электронного словаря, можно воспользоваться анализом структуры уже существующих словарей — таких, как Контекст 3.0, Мультилекс 1.0, RusLan 1.0, POLYGLOSSUM, Lingvo.

Список литературы: 1. *Караулов Ю.Н.* Машинный фонд русского языка: идеи и суждения. М.: Наука, 1986. 240 с. 2. *Убин И.И.* Перевод с помощью ЭВМ. М.: Всесоюз. центр пер. (ВЦП). 1986. 80 с (Обзор информ. / ВЦП; Вып. 11). 3. *Михайлов В.И.* О системе текущего отслеживания эквивалентов неологизмов специальной лексики из неопубликованных научно-технических переводов. Х., 1985. 51с. Деп. в ИНИОН РАН 19.05.86, № 25281 52/86000 2587 Деп. 25281.

*Поступила в редколлегию 29.12.97*