

ПРИКЛАДНОЕ ЯЗЫКОЗНАНИЕ

И. К. БЕЛЬСКАЯ

О ПРИНЦИПАХ ПОСТРОЕНИЯ СЛОВАРЯ ДЛЯ МАШИННОГО ПЕРЕВОДА

Словарь для машинного перевода (МП) отличается от обычного словаря рядом особенностей¹.

Первая особенность состоит в разделении всего словаря на некоторое число независимых словарей — по числу языков, привлеченных для МП. Таким образом, отдельно записывается словарь языка, с которого происходит перевод, и языка, на который перевод осуществляется.

В словарях группы А (т. е. тех языков, с которых делается перевод) каждое слово снабжено указанием на номер его эквивалента (или эквивалентов) в словаре группы Б (например, русском, словаре). Для того чтобы сделать этот последний пригодным не только для МП на русский язык, но и с русского, каждое русское слово снабжается индексами, указывающими номера его эквивалентов в словарях группы А.

Вторая особенность состоит в разделении словаря каждого языка на две секции — секцию однозначных слов и секцию многозначных слов. Это необходимо для того, чтобы обеспечить каждое многозначное слово в словаре МП правилами выбора нужного значения этого слова. Секция однозначных слов в свою очередь делится на подсекцию терминов, объединенных в несколько групп соответственно принадлежности к той или иной области науки (термины математические, физические и т. д.), и подсекцию общеупотребительных однозначных слов. Такое строение словаря обеспечивает его четкую и быструю работу при МП, а также создает возможность расширения отдельных частей словаря без нарушения общей его структуры.

В лингвистическом отношении новой является секция многозначных слов. Это опыт полной формализации семантического анализа многозначных слов. До сих пор наблюдения в этой области фиксировались как разнообразные, иногда случайные заметки «о трудных случаях» многозначности слов². Переводчику обычно предлагалось «просто запомнить» эти «трудные случаи», причем сколько-нибудь конкретные указания о выборе нужного варианта из названных не сообщались.

Третья особенность заключается в том, что словарь МП содержит не только перечень лексических единиц (слов), но и определенный комплекс грамматических характеристик каждого слова. В существующих

¹ О технических вопросах реализации МП см. следующие работы: Д. Ю. П а н о в, Автоматический перевод, 2-е изд., М., 1958; И. С. М у х и н, Опыты автоматического перевода на электронной вычислительной машине БЭСМ, М., 1956; И. К. Бельская, Л. Н. Королев, И. С. Мухини др., Некоторые вопросы автоматизации перевода, «Вестник АН СССР», 1956, № 12; С. Н. Р а з у м о в с к и й. К вопросу об автоматизации программирования задач перевода с одного языка на другой, «Докл. АН СССР», т. 113, № 4, 1957.

² См., например, С. С. Т о л с т о й, Основы перевода с английского языка на русский, М., 1957, стр. 20—21.

словарях обычного типа грамматические сведения о словах не носят сколько-нибудь систематического характера и сообщаются преимущественно как предупреждение о некоторой нерегулярности в грамматическом оформлении слова. Словарь МП содержит систематизированное грамматическое описание каждого слова, называемое «инвариантной (или словарной) характеристикой слова».

Четвертая особенность состоит в том, что словарь МП дает систему релятивных значений слов. Система значений слов языка-источника представлена здесь не абсолютно, а в ее отношении к лексической системе другого языка, с которым сопоставляется данный язык в целях осуществления перевода.

Пятая особенность заключается в том, что словарь МП фиксирует «нулевые значения» слов, т. е. случаи, когда слово не следует переводить на другой язык как отдельную лексическую единицу.

Поясним подробнее последние три особенности. Определенная грамматическая характеристика слова желательна во всяком словаре. Говоря о задачах составителей обычных (толковых) словарей, С. И. Ожегов замечает: «...показ грамматического... оформления слова и синтаксических возможностей его приобретает огромное значение для словаря. Чем точнее указания словаря в этом отношении, тем лучше он выполняет свою роль»¹. В словаре, составленном для машинного перевода, грамматическая характеристика слов обязательна. Здесь не предполагается участия человека, который мог бы «досказать» некоторую исходную грамматическую информацию, необходимую при грамматическом анализе переводимого предложения. Существующие словари обычного типа, как правило, в описании грамматической стороны слова ограничиваются общим указанием на принадлежность слова к определенной части речи и, в лучшем случае, фиксированием некоторых «неправильностей» в его формообразовании. Предполагается, что «правильные» грамматические сведения систематизированы в грамматике. Переводчику предлагается мысленно соединить эти сведения со словарными. При машинном переводе отсутствует не только переводчик, но и систематический грамматический очерк отдельного языка где-либо вне словаря. Есть словарь и правила сравнительно-грамматического анализа предложения. Чтобы эти правила работали, из словаря должны быть получены сведения не только о лексическом значении слова, но и некоторая исходная грамматическая информация.

К мысли о необходимости так или иначе сообщать в словаре некоторую грамматическую информацию о слове пришли почти все исследователи проблем МП. Так, американские и английские исследователи, отказавшись от первоначальной идеи сведения всех различий между языками к лексическим различиям², были вынуждены признать весьма желательным выделять в словаре некоторые «значимые для предложения» категории слов. Однако их высказывания на эту тему долгое время были очень осторожны и в ряде случаев непоследовательны. В частности, ими была сделана попытка, поместив в словаре отдельно основы слов и окончания, предоставить весь грамматический анализ делать словарю на основе анализа окончаний слов.

Выяснение этого вопроса имеет принципиальное значение для МП. Для того чтобы обеспечить перспективность системы языкового анализа для МП, необходимо правильно разграничить лексические и грамматические категории в языке. «Грамматическим является выражение обобщенного отношения посредством лишь изменения слов и их соединения как

¹ С. И. Ожегов, О трех типах толковых словарей современного русского языка, ВЯ, 1952, № 2, стр. 102.

² См. статьи: A. G. Oettinger, The design of an automatic Russian-English technical dictionary; Y. Bar-Hillel, Idioms (стр. 52), сб. «Machine translation of languages», New York - London, 1955.

такового, отвлеченного от их конкретности; лексическим же является выражение словами как таковыми, т. е. как частями словарного состава»¹.

В словаре МП должно храниться все то, что выражается в языке лексически. Принципом, определяющим включение в словарь той или иной лексической единицы, может служить следующий тезис, выдвинутый Л. В. Щербой: «Все, что происходит по правилам, будет явлением грамматическим, а все то, что является индивидуальной принадлежностью того или иного слова, будет явлением лексическим»². Этот принцип оказывается весьма полезным и при решении в условиях МП вопроса о «неправильных», индивидуальных формах того или иного слова, которые при таком подходе утверждаются в своем праве быть занесенными в словарь МП. Явления, отнесенные Л. В. Щербой к грамматическим, могут быть вынесены за рамки словаря и выделены в систему грамматических правил, но в словаре МП должны быть фиксированы указания на все эти грамматические правила в виде инвариантных признаков слов.

Грамматические категории, присущие слову, могут быть двух типов. К первому из них относятся те грамматические категории, которые определяют тип изменения слова, сохраняя при этом постоянное значение независимо от употребления слова в предложении. Значение же категорий второго типа варьируется в зависимости от функций слова в предложении. В каждом языке категории первого типа составляют некоторую определенную и постоянную систему грамматических категорий, которую одновременно с переводом лексического значения слова следует сообщить в словаре МП. Устанавливаемые таким образом системы инвариантных (или словарных) признаков противопоставляются системам вариантных (или контекстных) грамматических признаков слова, значение которых выясняется из анализа слова в предложении.

Так, для имен существительных в русском языке, в отличие от вариантных признаков числа и падежа, инвариантными признаками являются тип склонения, тип основы, категории грамматического рода, одушевленности, нарицательности и местоименности, а также особенности в формообразовании и принадлежность к какой-либо лексической группе. Для русского прилагательного инвариантными признаками являются только тип основы, категории местоименности, особенности в формообразовании и принадлежность к лексической группе. Грамматические категории имени прилагательного — род, число, падеж, степени сравнения и т. д. — составляют его вариантные признаки. Доминирующим в системе грамматических признаков слова является признак части речи, относящийся к числу инвариантных признаков.

Естественно, что системы вариантных и инвариантных признаков различны для разных языков. Важно отметить, что их состав определяется для целей МП с учетом их роли (точнее, необходимости в них) при переводе на заданный второй язык (в нашем случае — на русский). Поэтому в нашем словаре не фиксированы многие инвариантные грамматические признаки, о которых сообщается в традиционной грамматике. Напротив, ряд грамматических признаков этого типа впервые зафиксирован в словаре МП. Системы инвариантных признаков, фиксированные в словаре МП, содержат все те исходные данные, которые необходимы для последующего грамматического анализа слова в предложении.

Последние две из названных выше особенностей словаря МП обусловлены специфическим характером лексических значений слов, зафиксированных в словаре МП. Уже давно было замечено, что фактически при переводе переводчик почти никогда не может ограничиться выбором из

¹ А. И. Смирницкий, Аналитические формы, ВЯ, 1956, № 2, стр. 50.

² Л. В. Щерба, Очередные проблемы языковедения, ИАН ОЛЯ, 1945, вып. 5, стр. 129.

числа тех значений, которые зафиксированы в словаре. Примеры того, к каким недоразумениям и искажениям нередко приводит некритическое пользование существующими словарями, цитируются во всех статьях и книгах, обобщающих наблюдения известных переводчиков¹. По нашему мнению, всякий двуязычный словарь, поскольку он является переводческим словарем, должен отличаться от любого одноязычного (нормативного или толкового) словаря тем, что он не ставит своей задачей дать системы значений слов в каком-либо одном языке. Двуязычный словарь должен фиксировать не систему значений слов какого-либо языка, а перевод этой системы на другой язык. В этой связи нами введены понятия собственных и релятивных значений слов.

Под собственными значениями слов мы понимаем те значения, которые слова имеют в пределах языка, к которому принадлежат. Релятивными значениями слов мы называем те значения (или систему значений), которые получают слова при переводе их на другой язык. Собственное и релятивное значения слова совпадают не всегда. Так, английское прилагательное *good* в сочетании с существительным *chance(s)* приобретает при переводе на русский язык релятивное значение, содержащее количественную характеристику слова *шанс(ы)*, тогда как в английском языке представлена качественная характеристика: *to have good chances* «иметь много шансов» (а не «хорошие шансы»).

Пригодность двуязычного словаря для перевода определяется тем, насколько он отражает именно релятивные значения слов. Отсутствие установки на передачу релятивных значений существенно снижает практическую ценность двуязычных словарей. В англо-русском словаре Мюллера для слова *amount* (существит.) указываются три значения: 1) сумма, итог; 2) количество; 3) (перен.) значение, смысл. Из них только первое отчасти отражает релятивное значение «сумма» слова *amount*. Два других значения при переводе не могут быть использованы. Ср.: *if we fix a certain amount (величину) as the allowed error...*; *In many cases the amount (объем) of mathematics... is quite adequate*. При переводе с помощью обычного словаря со стороны переводчика требуется «додумывание», которое по существу сводится к замене собственных значений слов релятивными. В нашем примере значения, указанные в словаре Мюллера, потребовалось заменить релятивными значениями «величина» и «объем». Важно подчеркнуть, что ни одно из них не является ни «окказиональным» значением, ни особым оттенком какого-либо другого значения. Это вполне устойчивые, закрепленные постоянным употреблением значения слова. Отсутствие релятивных значений в словаре МП может практически свести на нет всю работу словаря. Очевидно, необходимое «додумывание» должно быть произведено предварительно и зафиксировано в словаре.

Частным случаем релятивного значения является «нулевое значение», т. е. тот случай, когда слово не следует переводить на другой язык как отдельную лексическую единицу. Так, при переводе на русский язык английского предложения *It is interesting to note, how closely these findings parallel statistical studies* первые пять слов будут переданы двумя русскими словами «интересно отметить», а три слова *it, is, to* получают нулевое значение. Слово с нулевым релятивным значением остается самостоятельным структурным элементом в переводимом предложении, хотя лексически его роль сводится к модификации значения слова или слов, с ним связанных. Поэтому слово, получающее в словаре нулевое значение, сопровождается инвариантными признаками наравне с остальными словами.

Наиболее актуально нулевое значение для служебных слов, однако его могут иметь и полнозначные слова. Ср.: англ. *take* «делать»; *take + +sure* «убедиться» +0; *take + clear* «объяснить» +0; *take + ready* «приготовить» +0. Ср. также различные идиоматические выражения, эквива-

¹ См., например, А. В. Федоров, Введение в теорию перевода, М., 1953; С. С. Толстой, указ. соч.

лентные в языке, на который делается перевод, одному слову. В этом случае все слова, кроме «ключевого», получают нулевое значение.

Проблема автоматического перевода идиоматических выражений на другой язык многим исследователям казалась неразрешимой. Этот вопрос был специально рассмотрен Бар-Хиллелем в указанной статье «Идиомы». Решение проблемы автор статьи видит в том, чтобы иметь словарь идиом в дополнение к словарю обычных слов. Теоретически, полагает Бар-Хиллел, вполне очевидно, что достаточно распространенный общий словарь может выполнить задачу обоих словарей, практически, однако, при таком варианте словарь МП будет слишком перегружен дополнительными переводами и выбор нужного значения будет затруднителен без специальных указаний по каждому случаю. По нашему мнению, словарь идиом может быть включен в общий словарь и это не будет обременительным, если предварительно в общем словаре выделить в отдельную часть с л о в а р ь м н о г о з н а ч н ы х с л о в, чтобы сюда же влить и словарь идиоматических выражений. В таком случае идиоматическое использование слова найдет отражение в общей системе указаний, обеспечивающих правильный перевод слов, неоднозначно переводимых на другой язык.

В заключение заметим, что проблема словаря МП, бывшая той первой проблемой, с которой собственно началось обсуждение вопроса о машинном переводе, оказалась едва ли не самой трудной. На первых этапах наибольшее сомнение вызывали вопросы относительно объема словаря МП; приводились астрономические цифры слов (до 30 млн.)¹, которые необходимо «запомнить» в словаре. В последнее время такой неразрешимой проблемой считается проблема многозначности слов. По нашему мнению, вопрос о многозначных словах получает удовлетворительное решение для всех языков, если используются следующие два метода:

1. Разделение словаря МП на серию «специальных словарей» соответственно различным сферам человеческой деятельности; в нашем случае — соответственно различным отраслям науки (математический словарь, физический словарь, электротехнический и т. д.). Полисемия слов в пределах такого специального словаря существенно сужается. Так, английское прилагательное *high* в англо-русском словаре Мюллера имеет 10 различных значений. Между тем в специальном математическом словаре система его значений может быть сокращена до трех релятивных значений: «высокий», «большой», «высшего порядка». Типичными математическими контекстами *high* являются следующие: 1) *The difference equation is of too high an order to permit easy numerical solution* «Дифференциальное уравнение имеет слишком высокий порядок, чтобы допускать простое численное решение»; 2) *Hence, there is a need for numerical methods which enables one to calculate particular solutions of any differential equation... with a higher accuracy than...* «Следовательно, есть необходимость в численных методах, которые позволяют вычислять частные решения любого дифференциального уравнения... с большей точностью, чем...».

В китайском, японском и других языках полисемия слов в специальном (математическом) словаре сокращается не менее существенно.

2. Контекстный анализ слова. Понимая под многозначностью способность слова выступать в неоднотипном контексте, можно регламентировать значения слова путем установления (в той или иной области) типовых контекстов данного слова. Типовыми мы называем такие контексты слова, из которых каждый ассоциируется с некоторым новым семантическим вариантом последнего. Определение типовых контекстов осуществляется путем классификации возможных контекстов слова, полученных статистически.

¹ Y. Bar-Hillel, Can translation be mechanized? «American scientist», № 42, 1954.

Приведем пример контекстного анализа двух английских слов:¹

available:

171. (а, 172): Проверить анализируемое слово на *available*.
 а (I, б): Проверить вправо ближайшее слово на *commercially* (*)¹.
 б (II, в): Проверить вправо ближайшие 4 слова на сочетание *at + the + starting + point*;
 в (III, IV): Проверить влево ближайшее существительное на признак «название прибора» или на группу *method, theorem, formula...*
 I (о): Перевод: «имеющийся» (прилагательное, причастное).
 Перевод найденного слова: «в продаже» (наречн. группа).
 II (о): Перевод: «заданный» (прилагательное).
 III (о): Перевод: «применимый» »
 IV (о): Перевод: «доступный» »

after:

56. (а, 57): Проверить анализируемое слово на *after* и отсутствие признака «наречие»².
 а (I, б): Проверить вправо следующее слово (**) на существительное с признаком «подлежащее»³.
 б (I, в): Проверить вправо следующее слово (**) на существительное, за которым следует глагол с признаком «сказуемое» или с показателями сказуемого⁴.
 в (II, г): Проверить предшествующее слово на *one*, а следующее на *another*.
 г (III, д): Проверить влево ближайший предшествующий глагол на принадлежность группе «*pattern...*».
 д (IV, V): Проверить вправо следующее слово (**) на существительные группы «*method... rule...*».
 I (о): Перевод: «после того, как» (союз неоднородный).
 II (о): Перевод: «за» (предлог, управляет твор. пад.).
 III (о): Перевод: «с» (» » род. пад.).
 IV (о): Перевод: «по» (» » дат. пад.).
 V (о): Перевод: «после» (» » род. пад.).

¹ Индексы (*) и (**) указывают на необходимость применения определенных правил пропуска зависимых слов. Конкретные слова и сочетания слов, на которые опирается анализ полисемии, здесь приводятся для лучшего понимания характера анализа: в схеме они заменяются номерами или признаками групп, к которым эти слова принадлежат.

² Указанный признак слово может получить в схеме «Анализ омошимов».

³ Указанный признак получают в словаре, в частности, некоторые местоименные существительные.

⁴ Формальным показателем сказуемого дано единое описание.

Теоретически при контекстном анализе мы опираемся на положение о взаимосвязанности значений слов, составляющих словосочетания, о чем много писалось в последние годы в работах советских лингвистов. Ср. следующие высказывания А. И. Смирницкого и В. В. Виноградова: «...всякий язык... должен быть признан в основном аналитическим, так как в речи на этом языке большая часть значений будет выражаться словосочетаниями...»²; «иметь разные значения для слова чаще всего значит входить в разные виды семантически ограниченных фразеологических связей. Значение и оттенки значения слова большей частью обусловлены его фразовым окружением»³.

¹ В схемах принята следующая система линейной записи операций анализа: первая цифра (или буква) — номер команды; цифры (или буквы) в скобках: I-я — номер команды, к которой следует перейти в случае положительного ответа проверки, II-я — в случае отрицательного. (Даем вариант анализа, разработанный в ИТМ и ВТ.)

² А. И. Смирницкий, Аналитические формы, ВЯ, 1956, № 2, стр. 41.

³ В. В. Виноградов, Основные типы лексических значений слова, ВЯ, 1953, № 5, стр. 17.