

## ПРИКЛАДНОЕ И МАТЕМАТИЧЕСКОЕ ЯЗЫКОЗНАНИЕ

С. С. БЕЛОКРИНИЦКАЯ И ДР.\*

### РАЗЛИЧНЫЕ ТИПЫ ОМОНИМИИ И СПОСОБЫ ИХ РАЗЛИЧЕНИЯ ПРИ МАШИННОМ ПЕРЕВОДЕ

(На материале английского, немецкого, русского, китайского  
и японского языков)

Количество и качество лексико-грамматической омонимии при машинном переводе тесно и непосредственно связано с принятым в данной системе машинного перевода типом словаря. В рассматриваемых нами алгоритмах словари строятся следующим образом. Английский словарь включает существительные единственного числа в общем падеже, прилагательные и наречия в положительной степени (супплетивные степени сравнения фиксируются как самостоятельные слова), глаголы в основной форме, как и в словарях обычного типа. В китайский словарь включаются слова в их основной словарной форме. Немецкий, русский и японский словари представляют собой словари основ.

В специальных схемах различения омонимии, занимающих особое место в системе словарного анализа алгоритма машинного перевода, не рассматриваются ни случаи лексической омонимии, ни случаи грамматической омонимии, которые решаются в схемах грамматического анализа. Система правил различения омонимов, описанию которых и посвящена данная статья, рассчитана на анализ лексико-грамматической омонимии.

При описании типов лексико-грамматической омонимии мы будем классифицировать все омонимы следующим образом: 1) омонимы, распознаваемые вне контекста на основе одних окончаний, т. е. омоосновы, и 2) омонимы, для разграничения которых необходим анализ контекста, т. е. омоформы. Омонимы первого типа мы назовем **с л о в а р н ы м и**, омонимы второго типа — **к о н т е к с т н ы м и**.

#### Словарные омонимы

Внутри словарных омонимов мы различаем: а) омонимы, принадлежащие к разным частям речи, б) омонимы внутри одной части речи.

а) Среди омонимов разных частей речи большое место в русском и немецком языках занимают основы существительных и глаголов, как, например, в русском языке: *кос-(кос-)*, *косить — коса, замен-(замен-), заменить — замена*; в немецком языке: *Frag-(frag-), die Frage — fragen, band-(Band), binden — der Band*. Различать омонимы данного класса помогает проверка на определенный тип окончаний, характерный для одной части речи и не характерный для другой.

Подобным же образом различается словарная омонимия «прилагательное — существительное» в русском языке (*цел — целый/цель; част — частный/часть* и т. д.) и омонимия «прилагательное — глагол» в немецком языке (*gleich — gleichen*).

\* В написании статьи принимал участие коллектив авторов: С. С. Белокриницкая, А. А. Звонов, М. Б. Ефимов, Т. М. Николаева, Г. А. Тарасова.

При машинном переводе с японского языка на русский различаются следующие классы словарной омонимии: 1) «существительное — прилагательное»; 2) «существительное — глагол»; 3) «существительное — прилагательное — глагол»; 4) «прилагательное — наречие». Выделяемые нами классы омонимии не находят своего отражения в традиционной японской грамматике, которая предполагает наличие в японском языке двух основных частей речи: «тайгэн» и «ёгэн» — «имён» и «предикатов».

Особое место занимает разграничение омонимов, у которых совпадают основа первого слова и все второе слово целиком.

б) Внутри одной части речи омонимия основ различается также путем анализа конкретных окончаний, возможных для одного грамматического типа и невозможных для другого. Это относится, в частности, к омонимии существительных, принадлежащих к разным склонениям и совпадающих в словаре в одной словарной основе в русском языке (*ос/оса — ось*) или к видовой омонимии внутри одного глагола (*прида/придам — придаю*).

Конкретная работа по разграничению всех описанных выше типов омонимии проводится на основе особых словарных помет, которыми заранее снабжается каждая основа с потенциальной омонимией. В общей системе алгоритмов машинного перевода разграничение словарных омонимов производится непосредственно после нахождения данной основы в словаре, т. е. после обработки введенного текста схемой «Отбрасывания окончаний».

### Контекстная омонимия

Под контекстной омонимией мы понимаем прежде всего конверсионную омонимию. Выяснение и разграничение контекстной омонимии проводится в основном путем анализа структуры всего предложения или посредством анализа отдельных словосочетаний. Необходимо учитывать, что цель этого анализа — не получение перевода омонимичных форм, а выработка для анализируемого слова признака отнесенности к той или иной части речи. Так, в английском языке выделяются пять основных классов конверсионных омонимов: «глагол — существительное» (класс I), «существительное — прилагательное» (класс II), «глагол — прилагательное» (класс III), «прилагательное — наречие» (класс IV) и «предлог — наречие» (класс V). Бедность морфологического оформления слова в английском языке лишь в незначительной степени ориентирует нас на морфологический анализ при различении омонимов; поэтому первостепенное значение приобретает здесь синтаксический анализ при некотором использовании конкретно-лексического анализа контекста. Группа правил по различению омонимов построена на совокупности использования принципов морфологического и синтаксического анализа<sup>1</sup> с одновременным применением в ряде случаев контекстного лексического анализа, что дало довольно эффективные результаты при проверке схемы на большом количестве нового текстового материала.

Синтаксический анализ при различении омонимов опирается на выделение грамматических сочетаний, характерных для слов одного грамматического класса и нетипичных или невозможных для слов другого грамматического класса. Синтаксический анализ может быть очень простым. Например, чтобы отнести слово в разряд существительных при рассмотрении омонимов класса «глагол — существительное», достаточно установить, что ему предшествует +<sup>2</sup> или определенный артикль *the*, или не-

<sup>1</sup> Ср. также Т. Н. Мошная, Вопросы различения омонимов при машинном переводе с английского языка на русский, сб. «Проблемы кибернетики», вып. I, М., 1958.

<sup>2</sup> Здесь и далее знак + означает, что в данном случае возможно применение одного из правил пропуска, допускающих при проверке пропуск слов и сочетаний определенных грамматических типов (например, наречие, частица, вводное слово, однородный член и т. п.).

определенный артикль *a*, или предлог, или числительное, или прилагательное.

Если же в процессе анализа выясняется, что ни одно из вышеперечисленных слов не стоит перед данным, производится дополнительная проверка на слова, стоящие непосредственно после нашего слова: если позади нашего слова окажется формула, за которой не следует ни существительное, ни союз, соединяющий однородные члены предложения (так называемый союз «однородный»), то наше слово также получит признак существительного. Например: «...and hence again by integration a third approximation Fig. 1, curve C) is derived» [«...и отсюда снова при помощи интегрирования выводится третье приближение (рис. 1, кривая C)»].

В ряде случаев для установления части речи слова-омонима возникает необходимость тщательно обследовать не только слова, непосредственно примыкающие к данному, но и проанализировать грамматически более широкий контекст.

Сложные омонимы класса «существительное — глагол — прилагательное» (например, *check*, *end*, *group*, *limit* и др.) рассматриваются последовательно в два этапа на основе вышеизложенных принципов анализа. Такие омонимы имеют в словаре особую помету и направляются сначала в раздел, где помещаются омонимы класса «глагол — существительное». Если в процессе анализа выяснится, что данное слово не является ни глаголом, ни существительным, оно отсылается в раздел омонимов класса «глагол — прилагательное», где и получает соответствующую грамматическую информацию.

До сих пор мы говорили только о конверсионных омонимах, имеющих парадигму. Наша специальная группа омонимов, помимо уже рассмотренных случаев, содержит также 4 правила дифференциации омонимов класса «предлог — наречие».

Трудность разрешения проблемы конверсии при машинном переводе с китайского языка на русский усугубляется почти полным отсутствием в китайском языке морфологического оформления слова. Основными факторами для решения конверсии, таким образом, могут служить строгий порядок слов, синтаксическая сочетаемость определенных классов слов и служебные слова. На каждый класс конверсии составляется специальная группа правил, после обработки по которой слово получает признак отнесенности к определенной части речи.

Для разграничения омонимии «существительное — глагол» мы, помимо анализа возможного суффиксального оформления данного класса слов, учитываем способность китайского существительного и глагола сочетаться с формантами прошедшего времени и совершенного вида (*цэн*, *цэнцзин*, *и*, *ицзин*), формантом *со*, модальными глаголами, связочными глаголами, отрицаниями, наречиями, суффиксами единичности, прилагательными, предлогами, послелогами, служебными словами и т. д.

Проследим это на ряде примеров:

Цзинь таолунь гэн гуанфа ды цингуан

Динли юйши чжэнмин

Цы фанчэн ю вэй и цзе

Иньвэй чжэнмин ши шифэнь цзяньдэнь

«Теперь рассмотрим более общий случай»

«Тогда теорема доказана»

«Данное уравнение имеет единственное решение»

«Так как доказательство является три-  
виальным»

Конверсия класса «существительное — прилагательное» подразделяется нами на два подкласса: «существительное — прилагательное относительное» и «существительное — прилагательное качественное». Необходимость такого подразделения обуславливается невозможностью одновременного анализа разнородных по лексической и грамматической природе типов слов. Словам этого типа, находящимся в препозиции к существительному, вырабатывается признак «прилагательное относительное», в противном случае — «существительное».

Слова, относящиеся к классу конверсии «существительное — прилагательное качественное» (*тунгоу* — «изоморфизм — изоморфный»; *даньчуньтунгоу* — «гомоморфизм» — «гомоморфный» и т. д.), анализируются с учетом того, что, как показали наблюдения над сочетаемостью таких пар с другими частями речи, слова такого типа, выступая в функции сказуемого, стоят на конце предложения и имеют перед собой два или более существительных, соединенных союзами *юй*, *гэнь*, *хэ* или *тун*. Следует отметить, что в целях упрощения грамматического анализа для данного класса слов вырабатывается признак «прилагательное относительное» в тех случаях, когда они попадают в позицию перед существительным.

Решение конверсии класса «прилагательное — наречие» не представляет особых затруднений, так как для этого необходимо только установить, находится ли слово этого типа в позиции к глаголу (или оно оформлено на формант *ди*). Слова этого типа в позиции к глаголу (или оформленные на *ди*) являются наречиями.

В китайском языке все модальные глаголы, по сути дела, в определенных ситуациях выступают как модальные наречия (как и в предшествующих случаях, мы исходим при этом из потребностей перевода на русский язык). Основным критерий для выработки этому классу слов признака «глагол модальный» — наличие в предложении подлежащего, выраженного существительным одушевленным.

На каждый из указанных классов конверсии составлены схемы анализа. В тех случаях, когда наряду с конверсионной омонимией имеется многозначность, слову сообщается признак «многозначное», и оно анализируется по схемам полисемии.

Своеобразное разграничение омонимии данного типа в японском языке связано с методами членения японского предложения на слова при машинном переводе. Учитывая, что японское письмо представляет собой чередование иероглифов и азбучных знаков «каны», мы предварительно выделяем слова от точки до ближайшего иероглифа, от одного иероглифического сочетания до другого или соответственно до любого знака препинания или формулы. Таким образом, в одном случае предварительно в качестве одного условного слова выделяется *экибун зотэйсики* («интегральное уравнение»), а в другом *ёмиката* («чтение») — как два.

В немецком языке для разграничения омонимических рядов, в которые входят существительные, используется специфическая особенность немецкого правописания, выражающаяся в том, что в немецком языке существительное всегда пишется с большой буквы. Таким образом, мы можем сказать, что различение омонимов в немецком языке основано на графико-синтаксическом принципе. Анализ слов, входящих в классы омонимов «существительное — глагол» (*Beweis — beweisen, Gebrauch — gebrauchen* и др.) и «существительное — прилагательное» (*Komplex* «комплекс» или «комплексный», *Parallele* «параллель» или «параллельный» и др.), начинается с проверки на написание с большой буквы. Отрицательный результат проверки дает однозначное решение: данное слово является не существительным, а глаголом или, соответственно, прилагательным. При положительном результате производится проверка предшествующего слова на «точку», которая также в ряде случаев дает решение: если предшествующее слово не является точкой, то анализируемое слово — существительное.

Синтаксический анализ при различении омонимии «существительное — глагол» основан на выявлении различия в типах предложений, выражающегося в позиции глагола-сказуемого или глагольной части составного сказуемого, а также на относительно фиксированном порядке слов немецкого предложения вообще, особенно характерном для научного текста. Другой характерной особенностью различения омонимии «глагол — существительное» является то, что обработке соответствующей схемой

должны подвергнуться все без исключения глаголы, так как в немецком языке не только теоретически, но и на практике возможна субстантивация любого глагола в инфинитивной форме.

Различение омонимии неизменяемых слов («предлог — частица») основано на том, что этим двум категориям слов свойственна принадлежность к различным синтаксическим группам. Предлог служит для выражения синтаксических связей между словами в группе существительного, частицы же, как правило, входят в группу глагола. Поэтому для выявления предлога производится проверка следующего слова на существительное (или на формулу или цифру, которые могут выполнять синтаксические функции существительного). Так, легко определить, что в нижеследующем примере в первом случае *zu* является предлогом, а во втором — частицей: «*Um zu diesen Zahlenwerten zu gelangen, sind erhebliche konstruktive Arbeiten notwendig gewesen*» («Для того чтобы получить эти численные величины, понадобились значительные конструктивные работы»). Конверсионная омонимия в русском языке случайна и встречается очень редко.

Наиболее сложная и актуальная задача — разграничение омонимии существительных разных склонений. Рассмотрим, например, парадигмы склонения в единственном числе существительного 1-го склонения *математик* и существительного 2-го склонения *математика*:

|        |                                 |                              |
|--------|---------------------------------|------------------------------|
| Им.    | п. <i>м а т е м а т и к</i>     | <i>м а т е м а т и к а</i>   |
| Род.   | п. <i>м а т е м а т и к а</i>   | <i>м а т е м а т и к и</i>   |
| Дат.   | п. <i>м а т е м а т и к у</i>   | <i>м а т е м а т и к е</i>   |
| Вин.   | п. <i>м а т е м а т и к а</i>   | <i>м а т е м а т и к у</i>   |
| Твор.  | п. <i>м а т е м а т и к о м</i> | <i>м а т е м а т и к о й</i> |
| Предл. | п. <i>м а т е м а т и к е</i>   | <i>м а т е м а т и к е</i>   |

Как видно из этого ряда, неомонимичными являются лишь формы творительного падежа обоих склонений<sup>3</sup>. При разграничении омонимии такого типа постепенно исключаются сочетания, характерные для одного слова и невозможные для другого.

Особое место среди контекстной омонимии занимает омонимия «прилагательное — наречие — категория состояния», характерная для слов типа *хорошо*, *плохо* и т. п. Если в предыдущих случаях речь шла о совпадении словоформ двух р а з н ы х слов, то здесь, напротив, одно первоначально аморфное слово в зависимости от синтаксической функции может менять свою морфологическую принадлежность. Так, одно и то же слово *хорошо* может быть наречием (*хорошо сделано*), категорией состояния (*мне хорошо*) и именной частью сказуемого. Здесь имеет место случай превалирования синтаксических факторов над морфологическими, т. е. синтаксическая функция слова в конкретном предложении определяет его принадлежность к частям речи, а не наоборот. Широко распространенное во всех языках явление лексико-грамматической омонимии представляет одну из наиболее слабо изученных проблем лексикологии и грамматики.

В целом, на наш взгляд, классификация омонимии по типам и разграничение омонимов различных классов имеют значение не только для прикладного, но также и для общего языкознания, хотя авторы данной статьи своей основной задачей считали решение проблемы омонимии применительно к машинному переводу. В большинстве случаев методы разграничения омонимии, описанные в настоящей статье, дали положительный результат.

<sup>3</sup> Им. падеж слова *математик* соотносится с род. падежом множественного числа слова *математика*, а род. падеж слова *математика* совпадает с им. падежом множественного числа слова *математик*.