

УДК 811.93(075.8)

DOI: 10.18384/2310-712X-2016-1-26-32

## КОНТЕКСТНОЕ РАЗРЕШЕНИЕ ЛЕКСИЧЕСКОЙ МНОГОЗНАЧНОСТИ

**Марчук Ю.Н.**

*Московский государственный областной университет  
105005, г. Москва, ул. Радио, д.10А, Российская Федерация*

**Аннотация.** Разрешение лексической многозначности (неоднозначности) необходимо при любой автоматической обработке текстов. Выбор переводного эквивалента должен производиться с учётом формальных характеристик переводимого текста. Формальные признаки, находящиеся в контексте, помогают решить вопрос о правильном эквиваленте. Эффективным инструментом для выбора переводного эквивалента является так называемый контекстологический словарь. Это особого типа словарь, в котором каждому многозначному слову сопоставлен определённый контекст, включающий так называемые детерминанты – опорные слова, при наличии которых выбирается переводной эквивалент. Таким образом, для каждого многозначного слова приводится алгоритм выбора одного из возможных переводов. Контекстологический словарь составляется на основе конкорданса, который базируется на определённом массиве параллельных текстов – текстов на исходном языке и языке перевода.

**Ключевые слова:** лексическая многозначность, выбор переводного эквивалента, контекст, контекстологический словарь, алгоритм перевода.

## RESOLUTION OF POLYSEMY IN CONTEXT

**Yu. Marchuk**

*Moscow State Regional State University  
10A, Radio str., Moscow 105005, Russian Federation*

**Abstract.** The correct choice of equivalent for polysemantic words or resolution of ambiguity is a necessary step in automatic translation process. Context makes possible to resolve ambiguity of words and lexical units, thus using a special dictionary named “contextological”, which provides necessary equivalents, checking context determinants, proved effective. The contextological dictionary is produced on the basis of special dictionary named concordance, which consists of lists of contexts, in which the ambiguous word is used. These contexts are of certain lengths – four or three words to the left and to the right of the word in question – and are produced on the basis of parallel texts.

**Key words:** lexical ambiguity, choice of translation equivalent, context, contextological dictionary, translation algorithm.

Лексическая многозначность всегда возникает в переводе. Объём переводов в мире постепенно возрастает с развитием международного общения, прогрессом науки и техники, увеличением числа языков в мире. Универсальный язык,

как естественный (английский, китайский или испанский), так и искусственный, едва ли станет эффективным средством преодоления языковых барьеров, которые являются самыми труднопреодолимыми на пути распространения информации [6]. Каждый народ хочет общаться на том языке, который является для него родным, какой бы размах ни приобрели усилия по глобализации. Перевод, осуществляемый человеком, трудоёмок, требует много времени и дорог. Поэтому, как хорошо известно, в настоящее время в Интернете большое распространение получили системы машинного перевода, разработанные в разных странах и имеющие разные технические и научные характеристики. Общим для всех этих систем является низкое качество перевода и большое количество ошибок именно в переводе слов, как правило, не только терминов, но и общеупотребительных слов [2; 12]. Подробный анализ современных систем машинного перевода сделан П.Н. Хроменковым в работе «Современные системы машинного перевода» [11].

Машинный перевод позволяет уточнить само понятие «лексическая многозначность (неоднозначность)». Понятие «лексическая омонимия» уходит из употребления, поскольку компьютер не сможет разделить омонимию от лексической неоднозначности. «Стол» как 'предмет мебели' от «стола» в значении 'меню' отделить можно только в том случае, если в компьютер будет заложена чрезвычайно сложная картина мира или строго определены границы подъязыка. Но есть и более тонкие различия и сходства между лексической омонимией и лексической многозначностью. Действующие в Ин-

тернете системы машинного перевода дают иногда большое количество ошибок именно вследствие неправильного перевода слов – как правило, терминов.

В связи с этим проблема лексической многозначности в машинном переводе сводится к выбору одного из возможных переводов слова, и можно утверждать, что в машинном переводе, как и вообще в автоматической обработке, лексическая многозначность равна многопереводности, если можно употребить такой термин, вне зависимости от того, какое лексическое значение (словарное) имеют эти переводы, пусть даже они являются абсолютными синонимами и никакой многозначности в традиционном смысле не имеют. Разные переводы – разные значения, с точки зрения автоматической обработки.

Приведём пример неправильного перевода терминов в тексте, казалось бы, принадлежащем строго определённому языку компьютерных технологий. Речь идёт об описании типов компьютерной мыши. Перевод сделан одной из действующих систем.

#### **Гуртовщики мыши.**

**Microsoft компания получает много откликов после появления Окна 95. Мы выявили, что много пользователей встретили проблему мыши. В этом документе Служба Технического Упора Microsoft компании сводит вместе всю полезную информацию о возможных проблемах с мышами и гуртовщиками мыши и забота-стреляние. <...>**

**Для эффективной помощи технического упора, наш инженер должен знать торговую марку вашей мыши, тип (в-портовая мышь, без поводка**

**мышь, гениталий на гусеничном ходу и т. п.), версию гуртовщика, производителя компьютера (матери-доски), положение портов и рубильников на матери-доске (и расклад карт), а также содержимое досье Авто-### bat/config.Sys и Сапог-полено. txt.**

Морфологическая информация, т.е. согласование слов, падежи, перевод глаголов и многие другие трудности в данном тексте решены. Основные ошибки – перевод терминов. «Гуртовщики» явным образом появились из словаря общеупотребительных слов (driver). Непонятно, откуда появился «гениталий», да ещё «на гусеничном ходу». Можно предположить, что здесь сработал какой-то другой отраслевой словарь. Перевод Windows как «окна» тоже следует отнести к работе словаря общеупотребительных слов. «Служба Технического Упора» – буквальный перевод словосочетания, которое должно бы находиться также в отраслевом тематическом словаре.

Практически во всех моделях перевода затрагивается проблема разрешения лексической многозначности [5]. В ней есть несколько аспектов. Аспект 1 – взаимодействие предметных словарей. Безусловно, есть строго ограниченные предметные области, в которых каждое слово, каждый термин имеет только одно значение (перевод). В качестве примера можно привести микроподъязык водных коммерческих перевозок [1]. Но и в нём также есть общеупотребительные слова, значение которых определяется контекстом. Аспект 2 – взаимодействие словаря общеупотребительных слов и предметных словарей.

Взаимодействие словарей в машинном переводе есть проблема, весьма

трудно разрешимая, поскольку реально практически в каждом научном тексте могут быть описания, выходящие за пределы данной конкретной предметной области. Так, в медицинском тексте, безусловно обладающем своей собственной лексикой, может быть описание какого-то нового технического устройства, и в этом описании может появиться большое количество технических терминов, не входящих в традиционную медицинскую лексику.

Проблема разрешения омографии (лексической и морфологической неоднозначности) в системах машинного перевода возникла достаточно давно при конструировании и использовании первых в России систем практического машинного перевода. Так, в кандидатской диссертации А.Н. Киселева [3] была рассмотрена и детально разработана идея использования специальных словарей-конкордансов, составленных для определённых предметных полей, с целью выявить наиболее характерные наборы детерминант с последующим использованием этих наборов в алгоритмах разрешения многозначности по контексту. Однако и здесь было особо отмечено влияющее на правильное решение проблемы обстоятельство, заключающееся в том, что абсолютно изолированных, строго ограниченных предметных областей в реальных текстах практически не бывает. Взаимодействие разных предметных областей – объективный факт в существовании современных научно-технических текстов, поэтому и происходит переключивание и взаимовлияние семантики используемых в подобных текстах терминов. В этом проявляется новый этап семиозиса в массовой коммуникации, который по-

служил темой для формулировки академиком Ю.В. Рождественским новых принципов философии языка XX в. [8]. К этому же классу явлений относится взаимодействие языков, которое в настоящее время подробно изучается в новой лингвистической науке – лингвистической синергетике [7].

Особый интерес и особую сложность для автоматического разрешения лексической многозначности представляет соотношение слов общеупотребительного словарного фонда и лексики предметных областей. Приведённый выше образец машинного перевода убедительно доказывает это. Такие переводы, как «служба технического упора» (*technical support system*), и подобные им, являются прямым результатом такого взаимодействия. В некоторых случаях лексика вроде бы фиксированных предметных областей начинает взаимодействовать не только с нормативными словами общеупотребительного словарного фонда, но и с периферийными слоями. Так, представляет интерес «Англо-русский словарь жаргонных лексических единиц, употребляемых в сфере делового общения», составленный Н.В. Фоминой [10]. Этот словарь составлен не по каким-либо устным сообщениям информантов или по записям устных переговоров, а по материалам английских и американских периодических изданий *The Economist*, *The Times*, *The Financial Times*, *The Washington Post*, *The Business Week* и др. Иными словами, такое употребление лексики имеет, если можно так выразиться, некий закономерный статус и относится к печатному слову массовой коммуникации. Приведём некоторые примеры из этого словаря: *macaroni defence* – действия про-

тив поглощения компаниями; *locked market* – рынок, на котором сделки затруднены, поскольку цены продавцов равны ценам покупателей; *lemon* – 1) плохое, неудачное инвестирование; 2) перспективные акции, которые не оправдывают надежд; *pepper* – повышение привлекательности товара; *job depth* – сила влияния одного служащего компании на всех своих коллег; и т.п. Здесь мы видим не только новые словосочетания или словосочетания с новыми значениями, но и отдельные слова, которые приобретают новые значения (соответственно, новые переводы в автоматическом режиме).

Какими же средствами можно добиться удовлетворительного разрешения лексической многозначности слов, особенно слов-терминов, от правильного перевода которых зависит понимание смысла текста? Представляется, что шагами в этом направлении могли бы быть:

1) определение преобладающих и других предметных областей в подлежащих обработке текстах. В решении этого вопроса можно было бы использовать статистические характеристики распределения отдельных слов-терминов в пределах анализируемого предметного поля. Современные методики компьютерной обработки текстов дают возможность получить такого рода статистические характеристики со сравнительно небольшими затратами труда и машинного времени, по сравнению, например, с компьютерными технологиями прежних лет;

2) построение словаря-конкорданса по материалам обрабатываемых текстов. В такого рода словаре каждое слово представлено в контексте своего

употребления, и эти контексты употребляются. Методика составления и использования словаря-конкорданса описаны нами ранее [4]. По данным такого словаря можно, в частности, отделить однозначные слова (имеющие один перевод) от многозначных. Конкорданс может быть составлен на основе параллельных текстов, т.е. текстов оригинала и текстов перевода. В нём хорошо прослеживаются лексические, морфологические и синтаксические детерминанты, которые определяют конкретное значение (перевод) данного слова, в том числе и термина;

3) на основе конкорданса составляется контекстологический словарь, в котором каждое значение многозначного слова сопоставляется с набором контекстных детерминант, с помощью которого разрешается многозначность. Словарная статья контекстологического словаря представляет собой алгоритм последовательного запрашивания контекста на наличие тех или иных детерминант, определяющих значение слова. При отсутствии таковых слову придаётся значение, определяемое так называемым общим выхо-

дом алгоритма перевода.

Если составление конкорданса может быть максимально автоматизировано, то составление контекстологического словаря осуществляется человеком по материалам конкорданса. Это довольно трудоёмкое занятие, вследствие чего мы и имеем нынешнее положение с действующими системами машинного перевода, в которых, как правило, отсутствуют контекстологические словари. Между тем контекстологический словарь может также служить инструментом организации базы лексических данных, как это показал А.Л. Семёнов [9].

Представляется, что при достаточно точном решении этих задач неоднозначность, если и не будет полностью устранена, то в значительной мере уменьшится в обработанном тексте. При этом само собой разумеется, что чем более ограничена и специфична предметная область, чем в меньшей степени она будет взаимодействовать с общеупотребительной лексикой, тем выше будет степень разрешения неоднозначности.

#### ЛИТЕРАТУРА:

1. Зелко В.М. Проблемы разработки лингвистического обеспечения системы китайско-русского информационного машинного перевода: дис. ... канд. филол. наук. М.: Ин-т языкознания АН СССР, 1991. 165 с.
2. Исламов Р.С. Лексическая амбивалентность естественного языка в среде систем машинного перевода (на материале английской официально-деловой документации): автореф. дис. ... канд. филол. наук. Кемерово, 2015. 23 с.
3. Киселев А.Н. Исследование и алгоритмическое разрешение омографии при машинном переводе методом схемного анализа по структурным типам: дис. ... канд. филол. наук. М.: Военный Краснознаменный ин-т, 1984. 125 с.
4. Марчук Ю.Н. Компьютерная лингвистика: Учебное пособие. М.: АСТ, Восток-Запад, 2007. 317 с.
5. Марчук Ю.Н. Модели перевода. М.: Академия, 2010. 174 с.
6. Марчук Ю.Н. Лексика и новые информационные технологии. М.: Изд-во МГОУ, 2015. 207 с.
7. Пиотровский Р.Г. Лингвистическая синергетика: исходные положения, первые ре-

- зультаты, перспективы. СПб.: Филологический ф-т СПбГУ, 2006. 159 с.
8. Рождественский Ю.В. Философия языка. Культуроведение и дидактика. Современные проблемы науки о языке. М.: ГрантЪ, 2003. 239 с.
  9. Семёнов А.Л. Контекстологический словарь основных терминов маркетинга. М.: ВЦП, 1994. 122 с.
  10. Фомина Н.В. Англо-русский словарь жаргонных лексических единиц, употребляемых в сфере делового общения. Омск: Омский Государственный Технический университет, Омский Терминологический Центр, 2005. 87 с.
  11. Хроменков П.Н. Современные системы машинного перевода. М.: Изд-во МГОУ, 2005. 159 с.
  12. Challenges for Arabic Machine Translation / Ed. By Abdelhadi Souidi and others. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2012. 157 p.

#### REFERENCES

1. Zelko V.M. Problemy razrabotki lingvisticheskogo obespecheniya sistemy kitaisko-russkogo informatsionnogo mashinnogo perevoda: dis. ... kand. filol. nauk [The problem of developing a linguistic system of Chinese-Russian information machine translation: PhD thesis in Philology]. M., In-t yazykoznaniya AN SSSR, 1991. 165 p.
2. Islamov R.S. Leksicheskaya ambivalentnost' estestvennogo yazyka v srede sistem mashinnogo perevoda (na materiale angliiskoi ofitsial'no-delovoi dokumentatsii): avtoref. dis. ... kand. filol. nauk [The Lexical ambivalence of the natural language in the environment of machine translation (on the study of English official-business documents): abstract of PhD thesis in Philology]. Kemerovo, 2015. 23 p.
3. Kiselev A.N. Issledovanie i algoritmicheskoe razreshenie omografii pri mashinnom perevode metodom skhemnogo analiza po strukturnym tipam: dis. ... kand. filol. nauk [Research and the algorithmic solution of homography in machine translation by a method of scheme analysis of structural types: PhD thesis in Philology]. M., Voennyy Krasnoznamenny in-t, 1984. 125 p.
4. Marchuk YU.N. Komp'yuternaya lingvistika: Uchebnoe posobie [Computational linguistics: manual]. M., AST, Vostok-Zapad, 2007. 317 p.
5. Marchuk YU.N. Modeli perevoda [Translation models]. M., Akademiya, 2010. 174 p.
6. Marchuk YU.N. Leksika i novye informatsionnye tekhnologii [Vocabulary and new information technologies]. M., Izd-vo MGOU, 2015. 207 p.
7. Piotrovskii R.G. Lingvisticheskaya sinergetika: iskhodnye polozheniya, pervye rezul'taty, perspektivy [Linguistic synergetics: initial position, first results, prospects]. SPb., Filologicheskii f-t SPbGU, 2006. 159 p.
8. Rozhdestvenskii YU.V. Filosofiya yazyka. Kul'turovedenie i didaktika. Sovremennye problemy nauki o yazyke [Philosophy of language. Culture studies and didactics. Modern problems of the science of language]. M., Grant', 2003. 239 p.
9. Semenov A.L. Kontekstologicheskii slovar' osnovnykh terminov marketinga [Contextual dictionary of basic terms of marketing]. M., VTSP, 1994. 122 p.
10. Fomina N.V. Anglo-russkii slovar' zhargonnykh leksicheskikh edinit, upotrebyaemykh v sfere delovogo obshcheniya [English-Russian dictionary of slang lexical units used in the field of business communication]. Omsk, Omskii Gosudarstvennyi Tekhnicheskii universitet, Omskii Terminologicheskii Tsent, 2005. 87 p.
11. Khromenkov P.N. Sovremennye sistemy mashinnogo perevoda [Modern machine translation systems]. M., Izd-vo MGOU, 2005. 159 p.
12. Challenges for Arabic Machine Translation / Ed. by Abdelhadi Souidi and others. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2012. 157 p.

**ИНФОРМАЦИЯ ОБ АВТОРЕ**

*Марчук Юрий Николаевич* – доктор филологических наук, профессор, академик Международной академии информатизации, заведующий кафедрой теоретической и прикладной лингвистики Института лингвистики и межкультурной коммуникации Московского государственного областного университета;  
e-mail: marchuk43@migmail.ru

**INFORMATION ABOUT THE AUTHOR**

*Marchuk Yuri* – doctor of philological sciences, professor, academician at International Informatization Academy, head of Theoretical and Applied Linguistics department of Institute of Linguistics and Intercultural Communication at Moscow State Regional University;  
e-mail: marchuk43@migmail.ru

**БИБЛИОГРАФИЧЕСКАЯ ССЫЛКА**

*Марчук Ю.Н.* Контекстное разрешение лексической многозначности // Вестник Московского государственного областного университета. Серия: Лингвистика. 2016. № 1. С. 26–32. DOI: 10.18384/2310-712X-2016-1-26-32

**BIBLIOGRAPHIC REFERENCE**

*Yu. Marchuk.* Resolution of polysemy in context // Bulletin of Moscow State Regional University. Series: Linguistics. 2016. no 1. pp. 26–32. DOI: 10.18384/2310-712X-2016-1-26-32