

Исламов Роман Сергеевич

канд. филол. наук, старший преподаватель

ФГБОУ ВПО «Кемеровский государственный университет»

г. Кемерово, Кемеровская область

НЕОДНОЗНАЧНАЯ ЛЕКСИКА В СРЕДЕ СИСТЕМ МАШИННОГО ПЕРЕВОДА (НА МАТЕРИАЛЕ АНГЛО-РУССКИХ ПАРАЛЛЕЛЬНЫХ КОРПУСОВ ТЕКСТОВ ОФИЦИАЛЬНО-ДЕЛОВОЙ ДОКУМЕНТАЦИИ)

***Аннотация:** данное исследование посвящено особенностям функционирования неоднозначных лексических единиц, представленных такими явлениями естественного языка, как полисемия, омонимия и эврисемия, в среде систем машинного перевода с целью разработки критерия их классификации по степени неоднозначности. В работе был проведен эксперимент с использованием двух типов машинного перевода как инструмента для систематизации неоднозначной лексики, в ходе которого также давалась оценка переведенному машиной материалу с опорой на перевод, выполненный человеком.*

***Ключевые слова:** письменный перевод, машинный перевод, системы машинного перевода, неоднозначность, лексическая неоднозначность.*

Системам машинного перевода (далее СМП) на протяжении всего времени их создания, разработки и совершенствования приходится иметь дело с лексической неоднозначностью человеческого языка. Традиционно к категориям неоднозначной лексики относят такие явления как *омонимию* (равноименность), *полисемию* (многозначность) и *эврисемию* (широкозначность) [3–6; 10; 11].

В настоящем исследовании ставится цель использовать современные СМП для оценки качества перевода неоднозначной лексики, функционирующей в их среде, а также как инструмент для разработки критерия ее классификации. В качестве СМП были использованы следующие типы: SMT (статистический машинный перевод – statistical machine translation) и RBMT (перевод по правилам – rule-based machine translation) [2; 9]. Источником материала послужили англо-русские

параллельные корпуса текстов официально-деловой документации [7; 8]. Полученные результаты сравнивались с переводом выполненным человеком (НАТ – human-accomplished translation).

В ходе исследования было определено, что для *полисемии* особенности функционирования установлены через:

1. *Верный перевод* при разрешении многозначности – выборе верного эквивалента (пример 1 для SMT и RBMT).

Пример 1

Оригинальный текст:

but it is necessary to take into account other industries as they diversify the regional economy.

Машинный перевод (SMT):

но надо учитывать и другие отрасли, поскольку они диверсифицировать региональную экономику.

Машинный перевод (RBMT):

но необходимо принять во внимание другие отрасли промышленности.

НАТ:

однако учет и других видов деятельности необходим, поскольку они диверсифицируют экономику региона.

2. *Неверный перевод* при: а) неразрешении многозначности на языке-источнике в данном контексте (пример 2 для RBMT); б) отсутствии перевода для полисемичной единицы на переводящем языке (пример 3 для SMT).

Пример 2

Оригинальный текст:

in an educational institution or social service facility or in their immediate vicinity.

Машинный перевод (SMT):

или в учебном заведении или социального обслуживания объекта или в непосредственной близости.

Машинный перевод (RBMT):

в учебном заведении или средстве социального обеспечения или в их непосредственной близости.

НАТ:

в учебном заведении, или общественном учреждении, или в непосредственной близости от них.

Пример 3

Оригинальный текст:

but only if we break with business as usual.

Машинный перевод (SMT):

но только если мы откажемся от обычного.

Машинный перевод (RBMT):

но только если мы порываем с обычным бизнесом.

НАТ:

но только если мы откажемся от рутинного подхода.

Особенности функционирования омонимичной лексики в среде машинного перевода прослеживаются через:

1. *Верный перевод* моносемичных омонимов (пример 4 для SMT и RBMT) и *верный перевод* полисемичных омонимов со снятием омонимии и выбором верного варианта для полисемичной единицы (пример 5 для SMT и RBMT).

Пример 4

Оригинальный текст:

Because agriculture accounts for the bulk of the labour force.

Машинный перевод (SMT):

Поскольку сельское хозяйство приходится основная часть рабочей силы.

Машинный перевод (RBMT):

Поскольку сельское хозяйство составляет большую часть рабочей силы.

НАТ:

Поскольку в развивающихся странах основная часть трудовых ресурсов.

Пример 5

Оригинальный текст:

Use receptacles with perfectly flat, thick bases.

Машинный перевод (SMT):

Используйте сосуды с идеально ровной толстые баз.

Машинный перевод (RBMT):

Используйте сосуды с совершенно плоскими, толстыми основаниями.

НАТ:

Используйте посуду с ровным толстым дном.

2. Неверный перевод моносемичных (пример 6 для SMT и RBMT) и полисемичных омонимов при не снятой омонимии (пример 7 для SMT и RBMT), порождении новой неоднозначности при выборе варианта, являющегося омонимом на переводящем языке (данная ситуация прослеживается и в НАТ) (пример 8 для SMT).

Пример 6

Оригинальный текст:

there shall be safe means of access from the deck to the hold for their use.

Машинный перевод (SMT):

должны быть предусмотрены безопасные средства доступа с палубы в провести для их использования.

Машинный перевод (RBMT):

должны быть безопасные средства доступа от палубы до захвата для их использования.

НАТ:

им предоставляются безопасные средства доступа с палубы в трюм.

Пример 7

Оригинальный текст:

to use its worldwide pool of personnel.

Машинный перевод (SMT):

использовать свою всемирную пул персонала.

Машинный перевод (RBMT):

использовать свой международный бассейн персонала.

НАТ:

использовать свои имеющиеся во всем мире кадры.

Пример 8

Оригинальный текст:

so far as practicable, live steam to any crane or winch.

Машинный перевод (SMT):

насколько это практически возможно, острого пара для любого крана или лебедки.

Машинный перевод (RBMT):

насколько реальный, живой пар любому подъемному крану или лебедке.

НАТ:

по мере возможности рабочий пар от любой лебедки или крана.

Закономерности функционирования эврисемичной лексики прослеживаются через:

1. Верный перевод с разрешением широкозначности (пример 9 для SMT и RBMT).

Пример 9

Оригинальный текст:

for example, and they may fear that their state apparatus is too fragile to countenance devolution.

Машинный перевод (SMT):

например, и они могут опасаться, что их государственный аппарат слишком хрупки, чтобы одобрить передачи полномочий.

Машинный перевод (RBMT):

например, и они могут бояться, что их государственный аппарат слишком хрупок, чтобы одобрить передачу.

НАТ:

и они могут опасаться, что их слишком слабый государственный аппарат не устоит в случае делегирования полномочий на низовые уровни.

2. *Верный перевод* без снижения степени широкозначности (по сравнению с НАТ) (пример 10 для SMT и RBMT).

3. *Неверный перевод* (пример 11 для SMT и RBMT).

Пример 10

Оригинальный текст:

The provided information was then examined in detail.

Машинный перевод (SMT):

Представленная информация была затем подробно рассмотрены.

Машинный перевод (RBMT):

Предоставленная информация была тогда исследована подробно.

НАТ:

В дальнейшем было проведено детальное исследование представленных списков предприятий.

Пример 11

Оригинальный текст:

Criminal organizations are not purely creatures of the underworld.

Машинный перевод (SMT):

Преступные организации не являются чисто существа преисподней.

Машинный перевод (RBMT):

Преступные организации не просто существа преступного мира.

НАТ:

Преступные организации являются порождением не только уголовного мира.

Итогом проведенного исследования является разработка критерия классификации неоднозначной лексики – *степень неоднозначности* создаваемой объектом исследования, по шкале от минимальной до максимальной неоднозначности, где закономерности функционирования *полисемии* в СМП относят ее в категорию минимальной степени неоднозначности в языке, так как перевод машиной выполняется в 1 шаг с разрешением многозначности в том или ином контексте. *Омонимичная лексика* относится к категории средней степени неоднозначности, что прослеживается на примере полисемичных омонимов, когда перевод осуществляется в 2 шага: 1) разрешение машиной омонимии; 2) разрешение машиной полисемии. Закономерности функционирования *эврисемии* в СМП соотносят ее с максимальной степенью неоднозначности, поскольку широкозначность способна функционировать на уровне как полисемии, так и омонимии. Иными словами, оба указанных выше источника могут быть широкозначными. Из этого следует, что даже в тех случаях, когда перевод выполняется верно, семантическая неопределенность слова продолжает сохраняться в тексте, пока не будет разрешена через дихотомию «гиперонимия – гипонимия», то есть использование в тексте синонима для широкозначного слова с более узкой семантикой, например: matter → event → affairs → business → financial dealings и т. д. Кроме того, омонимичные единицы можно также разложить по шкале градации степени неоднозначности от минимальной (два моносемичных омонима) до максимальной (четыре и более полисемичных омонима). При этом омонимы являются полными (таблица 1).

Таблица 1

Степень лексической неоднозначности на основе данных СМП

Степень неоднозначности	Источник неоднозначности	Пример	Шаг 1	Шаг 2		Результат
М И Н И И	Полисемия	environment	окружающая обстановка	не требуется		Неоднозначность снимается за 1 шаг
			окружающая среда			
С Р Е Д Н	Омонимия (на примере полисемичных омонимов)	spring	весна	рассвет, время зарождения (чего-л.) (переносн.)	деформировать	Неоднозначность снимается за 2 шага
			пружина	упругость, эластичность (переносн.)		
			родник	происхождение (переносн.)		
			скакать			
М А К С И М А Л	Эврисемия (на примере широкозначных полисемичных омонимов)	plant	a living organism	растение	русские эквиваленты демонстрируют высокую степень семантической обобщенности	Сохраняется широкозначность => неоднозначность полностью не снимается
				урожай		
			a place where an industrial or manufacturing process takes place	завод, фабрика		
				оборудование		
				комплект машин, станков		

Список литературы

1. Англо-русский словарь общей лексики. АБВУ. 100 тыс. статей. – 2011.
2. Информация о ПО: PROMT Home 11 Multilingual. – 2016.
3. Малаховский Л.В. Теория лексической и грамматической омонимии [Текст]. – 2-е изд., доп. – М., 2009. – 248 с.
4. Пауль Г. Принципы истории языка [Текст]. – М.: Издательство иностранной литературы, 1960. – 501 с.
5. Покусаенко В.К. К вопросу о классификации омонимов в русском языке [Текст] // Научные труды. – 1968. – Вып. 81. – С. 185–192.
6. Пристайко Т.С. О семантике широкозначных слов-конкретизаторов в тексте терминологической дефиниции [Текст] // Вестник Челябинского государственного университета. – 2013. – №31 (322). – Филология. Искусствоведение. – Вып. 84. – С. 66–69.
7. [Электронный ресурс]. – Режим доступа: <http://termsearch.info>
8. [Электронный ресурс]. – Режим доступа: <http://www.ruscorpora.ru/search-para-en.html>
9. [Электронный ресурс]. – Режим доступа: <https://translate.google.ru>
10. Семина И.А. Широкозначность и семантическая неопределенность [Текст]. – Тамбов: Грамота, 2009. – №1. – С. 185–190.
11. Cohen A. Why ambiguity? // In H.-M. Gaertner, S. Beck, R. Eckardt, R. Munsan, & B. Stiebels (Eds.), Between 40 and 60 Puzzles for Manfred Krifka, 2006 [Электронный ресурс]. – Режим доступа: www.bgu.ac.il/~aric/Ambiguity.pdf
12. Oxford Dictionary of English, 3rd Edition. – Oxford University Press, 2010.
13. Sagarna D. Machine Translation Technologies / D. Sagarna, A. Alaña, N. Basterretxea, L. Gravina [Электронный ресурс]. – Режим доступа: http://paginaspersonales.deusto.es/abaitua/_outside/ikasle/hlt_04/group3-b.htm