

УДК 81'322.4

Чернова Ирина Валериевна

кандидат филологических наук,

доцент кафедры теории

и практики перевода

Кубанский государственный университет

г. Краснодар, Российская Федерация

Irina Chernova

PhD in Philology,

Associate Professor

of the Department of Translation

Kuban State University

Krasnodar, Russia

chernova.iv.elt@mail.ru

Барбитова Виолетта Сергеевна
студент
Кубанский государственный университет
г. Краснодар, Российская Федерация

Violetta Barbitova
Senior student
Kuban State University
Krasnodar, Russia
barbitova.violetta@mail.ru

ОЦЕНКА ЭФФЕКТИВНОСТИ РАСПОЗНАВАНИЯ СЕМАНТИКИ МНОГОЗНАЧНЫХ СЛОВ СИСТЕМАМИ АВТОМАТИЧЕСКОГО ПЕРЕВОДА

Данная статья посвящена анализу и оценке эффективности работы систем автоматического перевода при распознавании семантики многозначных слов. Актуальность настоящего исследования обусловлена несколькими факторами: необходимость оценки роли человеческого фактора в языке и переводе в эпоху стремительного научно-технического прогресса; требования современных государственных образовательных стандартов, предписывающих подготовку высокопрофессиональных конкурентоспособных специалистов в области лингвистики и перевода; важности использования современных информационно-коммуникационных технологий в профессиональной переводческой деятельности.

К л ю ч е в ы е с л о в а: информационно-коммуникационные технологии; автоматические системы перевода; машинный перевод; полисемия; многозначные слова.

EVALUATION OF THE EFFICIENCY OF SEMANTIC RECOGNITION OF POLYSEMANTIC WORDS BY AUTOMATIC TRANSLATION SYSTEMS

The current article is devoted to the analysis and evaluation of the effectiveness of automatic translation systems in recognizing the semantics of polysemantic words. The relevance of this research is determined by several factors: the need to evaluate the role of the human factor in language and translation in the era of rapid scientific and technological development; the requirements of modern state educational standards prescribing the training of highly professional competitive specialists in the field of linguistics and translation; the importance of using modern information and communication technologies in professional translation activities.

Key words: information and communication technologies; automatic translation systems; machine translation; polysemy; polysemantic words.

Государственные образовательные стандарты последнего поколения требуют включения в программу обучения лингвистов и переводчиков дисциплин, обучающих использованию современных информационно-коммуникационных технологий, в рамках которых содержится курс по ознакомлению с работой автоматических и автоматизированных систем перевода, а также искусственного интеллекта. Актуальность темы настоящего исследования обусловлена тем, что несмотря на популярность систем автоматического перевода и заметное улучшение качества их работы в последние годы, до сих пор остаются недостаточно изученными вопросы способности систем машинного перевода распознавать семантику многозначных слов и правильно интерпретировать их в контексте. Также не осуществлялся сравнительный анализ эффективности работы этих систем и не производилась оценка их

качества, что является, на наш взгляд, значительным упущением и делает данную тему важной и актуальной как для научных исследований, так и для обучения высокопрофессиональных специалистов – лингвистов и переводчиков, которым необходимо знать специфику устройства и особенности использования в своей профессиональной деятельности систем машинного перевода.

В качестве объекта исследования были выбраны такие системы машинного перевода, как DeepL, Bing Microsoft Translator и Reverso Context, поскольку их изучение представляет научную новизну.

Предметом исследования являются полисемантические слова русского и английского языков, взятые из литературы на бытовую, техническую и военную тематики, наиболее часто подвергающиеся переводу с использованием систем автоматического перевода. Всего в ходе исследования было изучено более 10 статей и 60 терминов.

Теоретическую базу данного исследования составили положения, идеи и выводы по лингвистике, семантике, компьютерной лингвистике, искусственному интеллекту и машинному обучению, содержащиеся в работах отечественных и зарубежных ученых, таких как З. А. Набиджонова [1], С. А. Песина [2], Б. А. Серебренников [3], А. А. Утробина [4], W. J. Hutchins [5] и др. Также важными постулатами теоретической базы настоящей работы являются исследования, посвященные изучению теории многозначности слов, анализу влияния полисемии на процесс перевода и выявлению методов разрешения проблемы многозначности слов при выполнении перевода с помощью автоматических систем.

Машинный перевод имеет более полувековую историю. Ключевой вехой его развития стало создание нейронных сетей и модели Transformer – архитектуры глубоких нейронных сетей, использующей механизмы внимания, которые позволяют модели запоминать различные части входных данных и учитывать их в процессе обработки [6]. Также в рамках обучения нейронных сетей DeepL использует специальные программы – краулеры, которые автоматически ищут переводы в интернете и проверяют их точность. Кроме того, согласно официальному сайту Reverso, системы задействуют в своей работе контекстуальные словари, двуязычные онлайн-словари перекрестных ссылок, инструменты проверки грамматики и орфографии, а также инструменты спряжения слов [7].

Эффективность работы систем автоматического перевода хорошо исследуется именно на том, как они справляются с переводом многозначных слов в различных контекстах употребления: от бытовой до военной.

Например, одним из слов, которое представляет сложность для интерпретации системами машинного перевода, является слово *stone*. Предложение *Could you stone the cherries for me?* системы перевели как ‘Можешь побить меня вишенками?’ и ‘Не могли бы вы нарисовать вишню для меня косточкой?’ И лишь одна система DeepL смогла установить связь между

словами *cherries* и *stone*, проанализировать порядок слов и части речи и подобрать соответствующее значение, выдав перевод: ‘Не могли бы вы очистить вишню от косточек?’

В предложении *Engineers must consider a number of criteria when selecting the proper pig for a pipeline*, взятом из статьи нефтегазовой сферы, две из трех систем успешно интерпретировали значение слова *pig* как ‘скребок’, получившем свое английское название из-за специфики применения – его используют для разрушения и удаления глинистой корки со стенки скважины. Reverso Context перевел *pig* как ‘свинья’, проигнорировав техническую направленность текста.

Машинный перевод многозначных слов с русского на английский в военно-технической сфере также вызывает особый интерес из-за важности соблюдения точности и правильности перевода терминов данной области. Неверная передача слов военной терминосистемы на иностранный язык может иметь серьезные последствия, начиная с недопонимания инструкций, распоряжений, приказов и заканчивая угрозой безопасности. В предложении *Использование табуна в военных действиях запрещено международными конвенциями* термин *табун* означает химическое вещество – фосфорорганическое соединение, этиловый эфир диметиламида цианофосфорной кислоты, которое фашисты использовали во время Великой Отечественной войны в целях уничтожения населения, и которое запрещено к применению в военных целях всеми конвенциями. Системе Reverso Context удалось верно интерпретировать слово *табун* как 'tabun' благодаря контексту и наличию специализированных лексических ресурсов и баз данных, учитывающих специфику военной терминологии. Недостаточное контекстное понимание и ограниченные лингвистические ресурсы привели к неверному переводу термина как ‘herd’ системами DeepL и Bing Microsoft Translator.

Проанализируем еще один пример, на этот раз взятый из бытовой сферы, в котором хорошо известно всем название фрукта употребляется в нетипичном значении: *На ‘банане’ доехал до первого после Биробиджана шиномонтажа при въезде в Читу* ‘On the ‘banana’ I reached the first tire repair shop/tire fitting/tire repair service after Birobidzhan at the entrance to Chita’. Все системы перевели слово *банан* как ‘banana’, в то время как в данном контексте оно означает ‘запаска’ или ‘докатка’. Ввиду ограниченности баз данных или алгоритмов, которые позволили бы распознать нестандартное значение слова *банан*, ни одна из систем машинного перевода не смогла определить его правильно, даже несмотря на присутствие в предложении явных маркеров сферы употребления лексемы, таких как слово *шиномонтаж*, в переводе которого системы как раз посоревновались, что говорит о хорошей натренированности их на перевод терминов, связанных с техникой; также при переводе предложения системы не учли факт того, что на банане можно было доехать. Присутствие или отсутствие кавычек, в которые было заключено слово *банан* в оригинальном предложении, на перевод никак не повлияло. Таким образом, для успешного перевода нестандартных или специфических значений слов системам машинного перевода требуется

обновление баз данных, а именно включение в них жаргонных, сленговых значений слов.

Для оценки качества работы систем автоматического перевода мы построили диаграммы, в которые включили данные, полученные по результатам перевода многозначных слов системами DeepL, Bing Microsoft Translator и Reverso Context (рис. 1, 2).

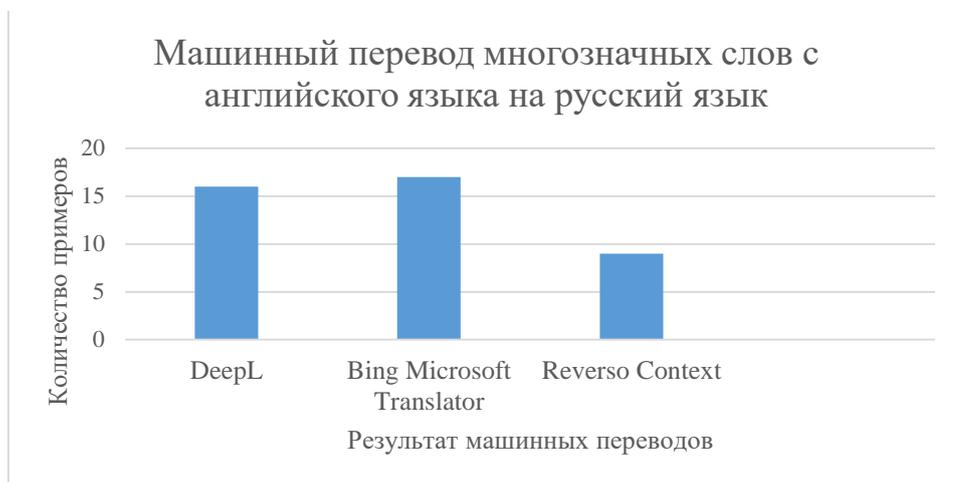


Рис. 1. Статистика успешных машинных переводов многозначных слов с английского языка на русский язык

Наилучший результат в переводе полисемичных слов с английского языка на русский показывает Bing Microsoft Translator с точностью 85 %. DeepL достигает необходимого результата с точностью 80 %. В то же время система Reverso Context в данном исследовании демонстрирует наименьшую эффективность в переводах, достигнув лишь 45 % точности. Данные значения мы получили на основе проведенной выборки, где за 100 процентов мы взяли 20 переведенных примеров. Количество успешно переведенных предложений мы разделили на 20 и умножили на 100. По такому же принципу мы получили статистику переводов многозначных слов с русского языка на английский.

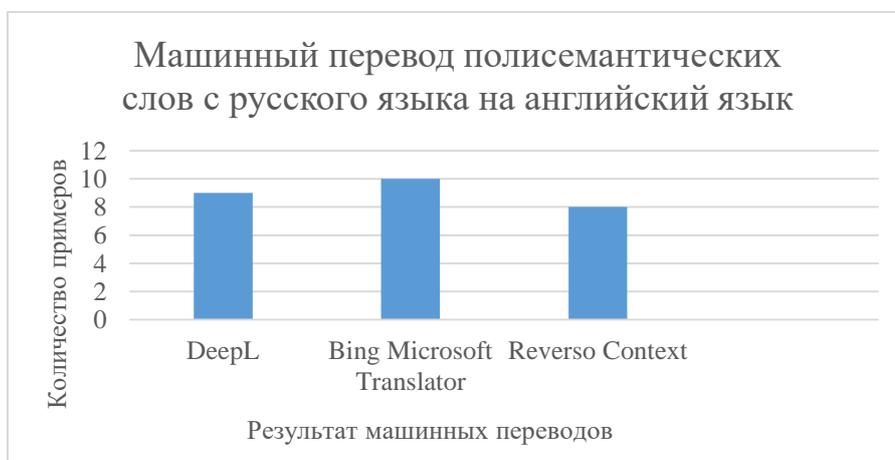


Рис. 2. Статистика успешных машинных переводов многозначных слов с русского языка на английский язык

Bing Microsoft Translator также продемонстрировал наилучший результат в переводе многозначных слов с русского языка на английский, выдав верные переводы в 83 % запросов. Система DeepL достигла точности в 75 % запросов. В данной статистике система Reverso Context занимает последнюю позицию с 66 % точности. Тем не менее мы отметили более высокие результаты Reverso Context при переводах с русского языка на английский по сравнению с ее же переводами с английского языка на русский. Это может быть обусловлено различиями в лингвистической структуре и семантике русского и английского языков. Погрешности в переводах присутствовали во всех системах, что лишь подтверждает всю важность и необходимость проведения дальнейших исследований в области изучения систем автоматического перевода и специфики их работы с целью повышения ее качества, для которого необходимо расширить объем обучающих данных, улучшить алгоритмы машинного обучения и интегрировать контекстуальные данные.

Таким образом, в результате проведенного исследования мы пришли к выводу о том, что использование автоматических систем перевода способно упростить и ускорить процесс перевода текстов, что повлечет за собой, с одной стороны, увеличение объема работ, который переводчик может успешно освоить, с другой – уменьшение количества времени, затраченного на перевод. Приобретение же студентами навыка перевода текстов с использованием автоматических систем перевода и их постредактирования приведет к тому, что рынок переводческих услуг обогатится высокопрофессиональными конкурентоспособными специалистами.

ЛИТЕРАТУРА

1. Набиджонова З. А. Лексико-грамматическая полисемия в английском языке // Достижения науки и образования. Иваново, 2020. С. 18–20.
2. Песина С. А. Полисемия в когнитивном аспекте : монография. Санкт-Петербург : Изд-во РГПУ им. А. И. Герцена, 2005. 325 с.
3. Серебренников Б. А. Роль человеческого фактора в языке. Язык и мышление. М. : Наука, 1988. 242 с.
4. Утробина А. А. Компьютерная лингвистика и машинный перевод : об истории становления [Электронный ресурс] // Научная электронная б-ка «КиберЛенинка». 2022. № 2. URL: <https://cyberleninka.ru/article/n/kompyuternaya-lingvistika-i-mashinnyu-perevod-ob-istorii-stanovleniya> (дата обращения: 28.06.2024).
5. Hutchins W. J. The development and use of machine translation systems and computer-based translation tools // International Journal of Translation. 2003. P. 113–114.
6. Deep Learning: что это и где применяется [Электронный ресурс] // Yandex Practicum. URL: <https://practicum.yandex.ru/blog/deep-learning-cto-eto-i-gde-primenyetsya> (дата обращения: 25.12.2023).

7. Reverso Corporate Translation Technologies [Electronic resource] // Reverso Corporate. URL: <https://www.corporate-translation.reverso.com/translation-technologies> (accessed: 10.04.2024).