**Лексические ошибки машинного перевода с английского языка на русский и пути их исправления (на примере перевода медиатекстов системой SYSTRAN)**

***Булдаков Егор Викторович***

*Студент*

*Московский государственный университет имени М.В.Ломоносова,*

*факультет иностранных языков и регионоведения, Москва, Россия*

*E–mail: yegor.buldakov@mail.ru*

Возросшие масштабы применения машинного перевода и использование такого вида деятельности как постмашинное редактирование обусловливает необходимость совершенствования систем машинного перевода для более эффективного взаимодействия переводчика-человека и переводчика-машины [4]. Научный интерес исследователей машинного перевода направлен, в том числе, на анализ и классификацию наиболее частых ошибок, допускаемых машиной [5]. Наше исследование качества машинного перевода фокусируется на лексическом аспекте перевода с английского языка на русский. Необходимость изучения лексического аспекта машинного перевода заключается в том, что программа машинного перевода неизбежно сталкивается с лексическими трудностями, в частности, с несовпадением значений лексических единиц в исходном языке и языке-цели [7]. Например, серьезную трудность для программы машинного перевода составляют случаи передачи значений безэквивалентной лексики и стилистически окрашенных единиц, а также понятий, для выражения которых в языке-цели могут использоваться несколько единиц в зависимости от контекстуальных условий (напр., русск. «голубой» и «синий» и англ. *blue*) [6; 7].

Предлагаемая нами типология лексических ошибок выявлена при работе с системой нейронного автоматического перевода SYSTRAN при переводе медиатекстов. Анализ текстов производился путем непосредственного сравнения предложений исходного текста и машинного перевода с последующей редакторской правкой. Медиатексты, на материале которых основана представленная в данном докладе типология, – это статьи, взятые из раздела *Lifestyle* в периодическом издании *The Guardian*. По жанру медиатекстов исследуемые статьи следует отнести к жанру текстов очерков по классификации Т. Г. Добросклонской [3]. Большая часть текстов представляют собой путевые очерки, в которых автор описывает путешествия и отдых и делится впечатлениями. Всего было проанализировано 3 текста и выбрано 53 предложения, содержащих одну или несколько лексических ошибок, которые затем были исправлены в отредактированном варианте предложений.

Под лексическими ошибками подразумевается некорректный перевод на уровне лексической единицы. В рамках данной классификации выделены следующие виды ошибок: 1) семантически некорректные соответствия; 2) ошибки лексической сочетаемости; 3) стилистические ошибки; 4) cлова, не распознанные машиной и опущенные или перекодированные путем транслитерации. В качестве единицы анализа было взято предложение, поскольку эта единица представляет оптимальный объем контекста для отражения уместности перевода лексических единиц. Для исправления лексических ошибок в отредактированных предложениях использовались следующие переводческие приемы: 1) добавление отсутствующей единицы; 2) замена на семантически подходящее соответствие; 3) антонимический перевод; 4) компенсация; 5) перестроение фразы или предложения; 6) модуляция.

В результате исследования было выявлено, что семантические некорректные соответствия (единицы, которые затрудняют понимание текста в связи с неверным подбором одного из множественных соответствий) – это самая частотная ошибка, допускаемая системой машинного перевода SYSTRAN. Количество предложений, в которых присутствуют семантически неверные соответствия, составляет 47 случаев, 63% всех ошибок. Например, при переводе предложения, в котором описываемый в статье курорт сравнивается с другими курортами, система SYSTRAN следующим образом переводит фрагмент *…this beautiful eco-friendly resort is hard to beat*: «…этот  красивый  экологически  чистый  курорт  трудно  победить». Глагол *beat*, учитывая контекст, выступает в значении *to be better or more enjoyable than another activity or experience* [9]*.* Следовательно,перевод лексической единицы *beat* некорректен с семантической точки зрения. Используя модуляцию, этот фрагмент можно перевести как: «…найти такой же роскошный и экологически чистый курорт будет довольно трудно». Вторым по частотности видом лексических ошибок, составляющим 15 ошибок (20% всех случаев), являются ошибки лексической сочетаемости, например перевод фрагмента *railways are powered  by hydroelectric energy* как «…железные дороги *питаются* от гидроэлектроэнергии» вместо «…железные дороги *работают* на гидроэлектроэнергии». Менее частотными являются стилистические ошибки, составляющие 10 случаев, 13% всех ошибок (напр., перевод слова *identify* как «идентифицировать» вместо нейтральной единицы «узнать» при переводе предложения, передающего устную речь во время дружеской беседы). К менее частотным типам ошибок также относятся нераспознанные единицы (6 случаев, 8%), которые были опущены или переведены транслитерацией (напр., перевод слова *breezily* (т.е. «беззаботно») как «бризло»).

Таким образом, в проведенном исследовании было выявлено, что выбор семантически неверных множественных соответствий и ошибки лексической сочетаемости оказались наиболее частотными типами лексических ошибок в проанализированных медиатекстах, так на них приходится более 80% всех ошибок. Соответственно, эти типы ошибок более всех остальных увеличивают объемы постмашинного редактирования.

**Литература**

1. Валеева Н. Г. Теория перевода: культурно-когнитивный и коммуникативно-функциональный аспекты. М., 2018.
2. Добросклонская Т. Г. Медиалингвистика: системный подход к изучению языка. М., 2008.
3. Марчук Ю. Н. Проблемы машинного перевода, М., 1983.
4. Нуриев В.А. Методы оценки качества машинного перевода: современное состояние // Информатика и ее применения. М. 2021. Т. 15. вып. 2. С. 104-111.
5. Castilho S. [et al]. Approaches to human and machine translation quality assessment // Translation quality assessment: From principles to practice. 2018. P. 9-38.
6. DiMarco C., Mah K. A model of comparative stylistics for machine translation // Machine translation, Vol. 1. 1994. Issue 1. P. 21-59
7. Hutchins W. J. Introduction to Machine Translation, Great Britain, 1992.
8. ‘A wild swimmer’s paradise’: the beach on top of a Yorkshire moor: https://www.theguardian.com/artanddesign/2023/dec/26/wild-swimmers-paradise-beach-on-top-of-a-yorkshire-moor
9. Cambridge Dictionary Online: https://dictionary.cambridge.org/
10. How to ski greener and switch from plane to trains: https://www.theguardian.com/travel/2023/oct/14/how-to-ski-greener-switch-from-plane-to-train
11. My eco-conscious odyssey in the French Alps: https://www.theguardian.com/travel/2024/jan/24/tignes-france-skiing-odyssey-in-the-alps
12. SYSTRAN Translate: https://www.systransoft.com/translate/