(cc) BY

Березина Юлия Валерьевна
канд. филол. наук, доцент
Байкова Ольга Владимировна
д-р филол. наук, заведующий кафедрой
Скурихина Ольга Витальевна
канд. пед. наук, доцент
ФГБОУ ВО «Вятский государственный университет»
г. Киров, Кировская область

ЛИНГВИСТИЧЕСКИЙ АНАЛИЗ ЛЕКСИКО-ГРАММАТИЧЕСКИХ ОШИБОК МАШИННОГО ПЕРЕВОДА

Аннотация: *в данной статье определены наиболее частые лексико-грамматических ошибки, возникающие в процессе автоматического перевода научно-публицистических текстов. Представленная тема является актуальной, так как в области переводоведения все больше внимания уделяется анализу и созданию компьютерных систем, которые могли бы оптимизировать и ускорить процесс перевода. Результаты исследования могут быть использованы для дальнейшего решения проблем теории перевода, а именно, описания лингвистических особенностей машинного перевода и определения типичных ошибок, возникающих в процессе данного вида перевода.*

Ключевые слова*: автоматический перевод, лексико-грамматические ошибки.*

Научно-технический прогресс привел к росту объема информации и расширению межъязыковых границ. Так называемый «информационный взрыв» – увеличение скорости и объема информации в мире, поставил перед переводчиками крайне трудную задачу – перевод огромного числа текстов с разных языков. Ввиду постоянного роста информационных потоков вопрос о скорости и качестве перевода встал особенно остро. Именно поэтому, переводчики начали прибегать к помощи систем машинного перевода, так как обработка текстов, без

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

использования компьютера, связана с высокими затратами как временного, так и материального характера.

Данная статья направлена на изучение наиболее частых лексико-грамматических ошибок, возникающих в процессе автоматического перевода научно-публицистических текстов.

В представленной работе лингвистический анализ лексико-грамматических ошибок проводится по классификации ошибок машинного перевода, предложенной Дж. Карбонеллом и А. Лави [1]. Внутри данной классификации выделяют следующие классы ошибок: 1) пропущенное слово; 2) неверный порядок слов; 3) неверное слово; 4) неизвестное слово.

Ошибка класса «пропущенное слово» возникает, когда в тексте машинного перевода отсутствует та или иная лексическая единица. Неверный порядок слов в переводном тексте может наблюдаться как на уровне предложения, так и на уровне словосочетаний. Ошибка класса «неверное слово», включает в себя: 1) смысловые ошибки, возникающие при неверном выборе переводческого эквивалента; 2) ошибки при выборе формы слова (при верно переведенной основе); 3) ошибки при переводе идиом и др. При возникновении ошибок класса «неизвестное слово», системы машинного перевода не осуществляют перевод слов или словосочетаний, а оставляют их в исходном виде, т. е. в тексте перевода присутствуют фрагменты на языке оригинала. Чаще всего ошибки данного класса возникают при переводе имен собственных, слов, принадлежащих узкой тематике, аббревиатур и сокращений.

Материалом исследования послужила статья доктора филологических наук Питера Кастберга – «Machine Translation Tools». Выбор материала научно-публицистического стиля речи обусловлен тем, что перевод текстов данного стиля является актуальным для специалистов различных сфер деятельности.

Итак, по результатам анализа ошибок машинного перевода научно-публицистического текста, выполненного переводчиком «Google-Translate», сделаны выводы о частоте допущения ошибок того или иного класса:

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

1) Класс «неверное слово» является самым распространенным типом ошибок машинного перевода. Значительную часть ошибок данного класса составляет неверный перевод специальной лексики, свойственной научно-публицистическим текстам.

2) Класс «неизвестное слово» является вторым по численности. Рассматриваемый переводчик не осуществил перевод сокращений и аббревиатур, а перенес их исходный вариант в текст перевода.

3) «Неверный порядок слов» является третьим по частоте классом ошибок машинного перевода. Большинство ошибок данного класса допускались на уровне предложения и для получения верного перевода необходимо было перенести слово или словосочетание в другую часть предложения.

4) Самым малочисленным классом ошибок машинного перевода оказался класс «пропущенное слово». Практически все пропущенные слова являлись словами-филлерами, т. е. словами, не несущими в себе главного смысла предложения. Пропуск ключевого слова был осуществлен в одном предложении. Безошибочно было переведено лишь одно предложение из ста.

Из всего вышесказанного можно сделать вывод о том, что даже нейронный машинный перевод, на принципе которого работают современные онлайн переводчики, не может осуществить перевод научно-публицистических текстов, который бы не требовал последующего редактирования.

Список литературы

1. Carbonell J. A Framework for Interactive and Automatic Refinement of Transfer-Based Machine Translation / J. Carbonell, A. Llitj´os, A. Lavie. // Proceedings of the 10th Annual Conference of the European Association for Machine Translation (EAMT) / – Budapest, 2005 – P. 87–96.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)