

Google как корпус текстов и его использование в переводе

Мигашикина Ольга Дмитриевна

Студент (магистр)

Челябинский государственный университет, Челябинск, Россия

E-mail: ms.isinka@mail.ru

В настоящее время роль текстов СМИ и, как следствие, роль их перевода огромна. Перевод, в свою очередь, затрудняется передачей различных особенностей языка СМИ, таких как реалии, аллюзии, клише и другие [Микоян, 2003, с. 2]. На начальном этапе исследования мы планировали использовать лингвистический корпус исключительно для проверки полученных в ходе эксперимента результатов. Однако было установлено, что корпусные технологии могут быть использованы на каждом этапе работы с текстом в процессе перевода: при определении жанра и функций текста, при поиске и передаче на язык перевода лингвистических особенностей публицистического текста.

Использование web-механизмов в качестве корпуса - достаточно новая тенденция. Многие ученые определяют корпус как «многочисленную базу аутентичных текстов в электронном формате, созданных для представления общих сведений о языке» [Boulton, 2011a, с. 2]. Задавая себе вопрос, является ли Интернет корпусом, мы смело можем отвечать, что да. В рамках исследования мы рассмотрели плюсы и минусы использования web-механизмов в качестве корпуса текстов. Интернет является самым большим корпусом, по сравнению со специальными корпусами текстов. Один из главных недостатков - это «неизвестный объем интернета как корпуса» [Boulton, 2011a, с. 2]. С другой стороны, в качестве корпуса можно использовать источники, содержащие гораздо больше слов и вхождений, чем самый большой существующий национальный корпус. Еще одним недостатком является «мгновенное пополнение информации в интернете» [Boulton, 2011, с. 2]. Всем известно, что информация в сети обновляется так быстро, что это практически невозможно зафиксировать, что, в свою очередь, не позволяет точно высчитать употребимость слова. Однако главное достоинство использования интернета в качестве корпуса - значительно больший объем информации - выводит сеть на один уровень со специально созданными для лингвистических исследований корпусами. Следовательно, мы можем сделать вывод, что на данном этапе развития корпусной лингвистики существует множество вариантов корпусов, более простых в использовании, как, например, поисковая система Google или другие механизмы сети Интернет. С другой стороны, существуют специальные корпуса с углубленным поиском, позволяющие лингвистам устанавливать сложные параметры поиска вхождений или сортировать вхождения в нужном им порядке. Одна из главных задач корпусного переводоведения на сегодняшний день - сделать поиск в корпусах проще, не уменьшая при этом функциональности самого корпуса и поиска.

В процессе исследования нами был проведен эксперимент по оценке возможности использования поисковой системы Google в качестве корпуса текстов при переводе публицистического текста с английского языка на русский. Эксперимент был проведен удаленно, для того, чтобы участники смогли детально ознакомиться с предоставленными материалами. Однако, мы планируем провести повторный эксперимент в ограниченных временных рамках, проверив гипотезу о том, что использование поисковой системы Google в качестве корпуса способно сократить время выполнения перевода. Большинство участников (89%) использовало Google при переводе предложенного текста, при этом 81% участников ознакомились с инструкцией по поиску в Google [Справочный центр Google, URL] и изучили различные операторы поиска (общие для Google и корпусов текстов) [Boulton, 2011b]. Мы планируем составить собственную пошаговую инструкцию по использованию

web-механизмов в переводе текстов СМИ и апробировать ее в повторном эксперименте. 66% участников анализировали количество результатов, получаемых при поиске в Google вариантов перевода, оценивая употребимость выражения по количеству вхождений, а 88% участников эксперимента анализировали контекст употребления вхождения при просмотре списка ссылок; собственно документ, в котором вхождение употреблено, анализировали 50% участников эксперимента. Данные результаты мы планируем уточнить при повторном эксперименте, попросив участников пошагово описать свою работу с поисковой системой Google при переводе.

В перспективы нашего исследования входит совершенствование методики использования корпусных технологий при переводе текстов СМИ, составление пошаговой инструкции по использованию web-механизмов при переводе текстов СМИ и проведение повторного эксперимента для получения новых и подтверждения уже полученных результатов.

Источники и литература

- 1) Микоян А.С. Проблема перевода текстов СМИ // Язык СМИ как объект междисциплинарного исследования / Под. ред. М.Н. Володиной. М., 2003. URL: <http://evartist.narod.ru/text12/12.htm>
- 2) Справочный центр Google. URL: <https://support.google.com/websearch#topic=3378866>
- 3) Boulton A. Wanted: Large Corpus, Simple Software. No Timewasters. CNRS, University of Lorraine, 2011a, France. 2 p.
- 4) Boulton A. What data for data-driven learning? The University of Nottingham, 2011b. 27 p.

Слова благодарности

Автор выражает огромную благодарность своему научному руководителю к.ф.н. доценту Волковой Татьяне Александровне.