Секция «Общая теория, история и дидактика перевода»

Участие переводчиков в расширении корпусов межъязыковых больших данных

Научный руководитель – Гарбовский Николай Константинович

Цзинь Ифан

Аспирант

Московский государственный университет имени М.В.Ломоносова, Высшая школа перевода (факультет), Кафедра перевода и переводоведения, Москва, Россия

E-mail: 1554583267@qq.com

В «цифровую эпоху» переводчики не только используют «большие данные» для получения наиболее достоверных результатов в кратчайшие сроки благодаря современным информационно-коммуникационным технологиям, но и сами вносят существенный вклад в пополнение корпуса «межъязыковых больших данных». Поэтому работа с корпусами больших данных на сегодняшний день стала одной из самых важных тем методологии перевода

Корпус представляет собой совокупность текстов, объединенных на основе тех или иных критериев.

Корпуса можно разделить на национальные и межъязыковые корпуса. В национальные корпуса, как правило, собираются тексты только на одном языке. Межъязыковые корпуса включают тексты на двух или более языках. С конца 20-го века благодаря интенсивному развитию компьютерных технологий различные страны приступили к созданию национальных корпусов, среди которых хорошо известны Британский национальный корпус (BNC), американский национальный корпус (OANC), а также Национальный корпус русского языка (НКРЯ), существующий в Интернете с 2003 года и включающий разные тексты на современном русском языке общим объемом более 600 млн слов. Корпус русского языка — это информационно-справочная система, основанная на собрании русских текстов в электронной форме и предназначенная «для всех, кто интересуется самыми разными вопросами, связанными с русским языком: профессиональных лингвистов, преподавателей языка, школьников и студентов, иностранцев, изучающих русский язык» [http ://www.ruscorpora.ru].

Корпуса межъязыковых больших данных содержат тексты на одном языке и их переводы на другой язык. Такие тексты, называемые параллельными или битекстами, не только дают важную информацию для принятия решений переводчиками, но и позволяют «обучаться» программам автоматического перевода.

Корпус межъязыковых больших данных создается с помощью специальных компьютерных программ, которые приводит в соответствие два текста (оригинал и перевод) по каждому предложению. Собрание битекстов может использоваться в качестве справочных материалов для поиска нужных сочетаний. Самым ранним из известных межъязыковых корпусов в истории человечества является розеттский камень, обнаруженный в Египте в 1799 г., с выбитыми на нем тремя идентичными по смыслу текстами, написанными древнеегипетскими иероглифами, египетским демотическим письмом и на древнегреческом языке. Текст камня относится ко времени правления Птолемея V Эпифана.

В настоящее время примером создания подобного рода корпуса можно назвать платформу YeeSight, применяемую с целью стимулирования роста знаний в широком спектре различных областей знаний, объединяющую преимущества мощного машинного перевода и семантический поиск. С помощью платформы YeeSight люди могут искать информацию

на своем родном языке и получать все соответствующие результаты на любом другом языке.

Достоинство корпуса межъязыковых больших данных состоит в том, что он может помочь людям сравнивать различные языковые рамки, способствуя развитию сравнительного-исторического языкознания и переводческих исследований. Межъязыковая платформа больших данных, основанная на межъязыковом корпусе, может помочь пользователям получать информацию на разных языках по одной и той же теме и расширять сферу применения исследований перевода, потому что, изучая перевод как процесс межъязыковой и межкультурной коммуникации, как коммуникацию с использованием двух языков, как контакт языков, мы со всей очевидностью обнаруживаем межъязыковую и межкультурную асимметрию [Гарбовский Н.К, 2007: 10].

Несмотря на то, что корпус межъязыковых больших данных быстро развивается, он также столкнулся с многочисленными проблемами:

- во-первых, составление корпуса межъязыковых больших данных - трудоемкая работа. Количество и типы корпусов, созданных в каждой стране, пока не могут эффективно удовлетворить потребности исследований;

- во-вторых, корпус дает недостаточный контекст, а некоторые корпуса содержат несоответствующие параллельные тексты.

Источники и литература

1) Гарбовский Н.К. Теория перевода: учебник / Н.К. Гарбовский. М., 2007. 544 с.

2) Национальный корпус русского языка // http://www.ruscorpora.ru