

РОЛЬ КОРПУСОВ В ОБУЧЕНИЕ АНАЛИЗЕ ХУДОЖЕСТВЕННОГО ПЕРЕВОДА

Ахмедова Анорхана Насивали кизи

Узбекский государственный университет мировых языков 1 курс докторант

<https://doi.org/10.5281/zenodo.10851527>

Аннотация: Методы корпусной лингвистики, её виды и практика корпусного анализа развиваются в различных сферах лингвистики Узбекистана. Отдельного внимания заслуживают воззрения Н. Абдурахмановой в сфере разработки автоматических методов грамматического аннотирования корпуса и создания для технологии машинного перевода систем, основанных на параллельном корпусе¹.

Ключевые слова: Метод, лингвистика, анализ, различных сферах лингвистики.

THE ROLE OF CORPORA IN TEACHING THE ANALYSIS OF LITERARY TRANSLATION

Abstract: Methods of corpus linguistics, its types and the practice of corpus analysis are developing in various areas of linguistics in Uzbekistan. The views of N. Abdurakhmanova in the development of automatic methods for grammatical annotation of a corpus and the creation of systems based on a parallel corpus for machine translation technology deserve special attention.

Keywords: Method, linguistics, analysis, various areas of linguistics.

ВВЕДЕНИЕ

Размышления касательно изучения функциональных особенностей корпусов на английском языке и создания частотного словаря терминов корпусной лингвистики с учетом результатов корпусного анализа нашли своё отражение в исследованиях Н. Атабаева². Лингвистические основы создания авторского корпуса в своих работах исследовала Ш. Хамраева³, концептуальным исследованием метафор в области гендера на базе корпусного анализа занималась Г. Эргашева⁴, с помощью английского корпуса градуономии в языке анализировала Ж. Джумабаева⁵. В мире на тему корпуса и практики корпусного анализа свои научные изыскания вели многочисленные ученые. Такие ученые, как Г. Лич, К. Макаулиф, М. Малберг, П. Милин, А. Мураками, П. Томпсон, Б. Винтер, В. Захаров, В. Мартин, С. Кублер, А. Лауренс, Э. Этуелл, Боизоу, Маккенери, Ж. Графмиллер, Дж. Грайв, Н. Грум, С. Ханссон, Р. Пейч, А. Счембри⁶ вели исследования в области теоретических концептологических основ корпусной лингвистики.

¹ Abduraxmonova N.Z. Kompyuter lingvistikasi: – Melbrun: Globe edit, 2020,; Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari (monografiya), Toshkent, 2021. – 200 b.

² Атабаев Н.Б. Инглиз тили корпусининг функционал хусусиятлари (СОСА мисолида) дис...ф.ф.док. – Тошкент, 2020. – 189 б.

³ Хамраева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Ф.б.фалс. д-ри (PhD) ... дисс. автореф. – Қарши: 2018. – 52 б.

⁴ Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Фил.фан.д-ри.дисс. автореф. – Тошкент: 2018. – 61 б. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>;

⁵ Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуономия. докторлик диссертацияси автореферати. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/avtoref_jumabayeva.pdf.

⁶ Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p; Седов А.В. Математические модели, методы и алгоритмы построения размеченных корпусов текстов. Автореф. дис. на соиск. учен. степ. канд.

Хотя параллельные тексты в составе корпусной лингвистики рассматриваются как новое направление науки, их история восходит к эпохе до нашей эры. В 196 году до н.э. египетские храмы адресовали царю Птолемею V надпись на камне, в которой прославляли фараона на двух языках (древнегреческом и древнеегипетском), и этот памятник стал прообразом «параллельных текстов». И хотя Розеттский камень является сравнительно новым в плане начертания надписей на камне, полнота его текста и тот факт, что этот текст дан параллельно на двух языках, делает его более законченным на фоне других каменных памятников.

В восьмидесятых годах XIX века параллельные тексты начали применяться в системе автоматической обработки текстов. В конце 50-х годов было сделано несколько попыток их использования в процессе применения в машинном переводе. Но в тот период возможности использования корпуса были ограничены сложностью ввода текстовых данных большого объема из-за ограниченных параметров хранилища данных и вместимости компьютера.

Впоследствии возникли такие идеи, как проведение различных исследований в области хранения, обработки и редактирования образцов перевода в составе корпуса. В конце 70-х годов прошлого столетия Мартин Кей и Мартин Рошайзен⁷ разработали первый автоматический способ параллельного автоматического выравнивания текста. Для выравнивания текстов различного уровня и объема они предложили множество способов, а именно разработали переводящие друг друга единицы. В эти единицы они ввели *параграфы, фразы, слова и обороты речи*.

Использование выровненных параллельных текстов было предложено Харрисом⁸. Предложенная им теория включает в себя *составление памяти переводов, словарей и списка двуязычных терминов*. Харрис выдвинул теорию, которая должна была способствовать получению межъязыковых данных параллельных корпусов, обучению при помощи компьютера или служить источником исследований для сравнительного языкознания и переводоведения.

Первый этап создания проекта любого параллельного корпуса представляет собой отбор соответствующей параллели или сравнительного текстового материала. Первый этап этого проекта чаще всего требует большого терпения и много времени, потому что включает в себя регистрацию авторского права, набор или сканирование текста и правку ошибок. Иногда для создания параллельных и сравнительных корпусов осуществляется автоматический сбор данных. При этом материалы *WWW (World Wide Web) (Всемирной паутины)* в основном рассматриваются в качестве настоящей альтернативы. Следует обозначить *URL*⁹ и

тех. наук (05.13.18) Петрозаводск, 2013. – 22 с.; Atwell E. Development of tagsets for part-of-speech tagging. An international handbook. Corpus Linguistics: Mouton de Gruyter. 2008.; Laurence Anthony, AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom, 2005 IEEE International Professional Communication Conference Proceedings, – P 729-737; Баранов А.Н., Михайлов М.Н., Сидоров Г.О. «Динамический корпус текстов» как новая технология прикладной лингвистики // Труды международного семинара Диалог по компьютерной лингвистике и ее приложениям. – Т.2. 1998.

⁷ Kay M., Roscheisen M. Text-translation alignment. Computational Linguistics, 1993. № 19 (1), –P. 121-142.

⁸ Harris B. Bitexts: A new concept in translation theory. Language Monthly, 1988. – P. 54,

⁹ URL считается интернет-адресом ресурсного источника.

параллельный *HTML*¹⁰. В дальнейшем Ресник¹¹ добавил в процесс создания параллельного корпуса фильтр идентификаторов.

Параллельные тексты называют *bitexts* – *двуязычными* или, если языков больше двух, *multitext* – *многоязычными*. На сегодняшний день благодаря улучшению в них средств хранения и архивирования электронных документов параллельных текстов становится всё больше, их охват расширяется, поэтому различные фирмы и веб-сайты используют параллельные корпуса, которые охватывают многоязычные документы.

Параллельные корпуса, используемые в корпусном переводе, можно разделить на следующие виды:

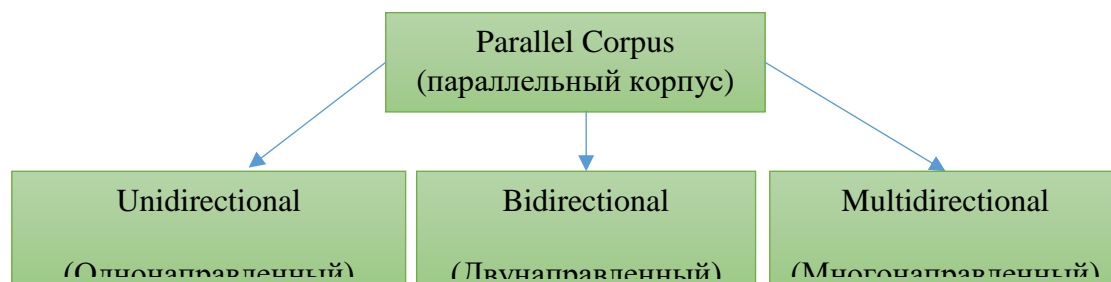


Рисунок 1. Классификация параллельного тела по свойству текста

Под однонаправленными (англ. *unidirectional*) параллельными корпусами понимаются тексты, переведенные с одного языка на другой. Например, произведения, переведенные с узбекского на английский язык. В двунаправленном (англ. *bidirectional*) параллельном корпусе понимается совокупность текстов, переведенных с узбекского на английский и наоборот с английского на узбекский. В разнонаправленном (полиязычном) параллельном корпусе текст состоит из текстов, переведенных с узбекского языка на английский, русский, испанский, немецкий и другие языки. Параллельный корпус сравнения относится к типу двунаправленных (косвенных) параллельных корпусов, т.е. образцы текстов, входящие в этот корпус, состоят из одинаковых частей текста, некоторые из которых переведены с узбекского на английский язык и наоборот.

Еще одна проблема в области создания параллельного корпуса заключается в том, что если одно понятие выражается в одном языке одним словом, то во втором языке оно может выражаться двумя и более словами. Если в оригинальном тексте мысль выражена посредством короткой фразы, то в переводе она может выражаться при помощи длинного предложения.

Параллельный корпус состоит из одного и того же текста, переведенного на два или несколько языков. Тексты выровнены, то есть связаны с соответствующими сегментами, фразами. Корпус даёт возможность поиска или сопоставления переводов на одном или на обоих языках.

При создании параллельного корпуса *слова* и *фразы* выравниваются. В английском языке этот процесс именуется термином *alignment*. Для создания параллельного корпуса требуется

¹⁰ *HTML* или язык гипертекстовой разметки – язык программирования, используемый для создания передней части веб-сайтов. Подразумевается размещение написанных для ввода структуры веб-страницы и на веб-сайты, просматриваемые нами через веб-браузеры.

¹¹ Resnik Philip “Mining the Web for bilingual text”, in: Proceedings of the 37th annual meeting of the Association for computational linguistics. 1999.

разделить тексты корпуса на сегментные единицы. Сегментные единицы, являясь выровненными словами или фразами, считаются отделяемыми частями параллельного (многоязычного) корпуса. Выровненное слово или фраза дают сведения о том, переводом какого сегмента с одного языка на другой является данный сегмент. Насколько короткими являются сегменты, настолько просто бывает найти в сегменте переведённое слово или выражение. Обычно сегменты даются в виде фраз, но некоторые корпуса можно выравнивать на уровне параграфа или документа. Самый простой способ ввести данные по выравниванию для параллельного корпуса это загрузить сведения в виде таблицы. Чаще всего таблицы подготавливаются при помощи программы *Excel*.

Для правильного использования параллельного корпуса исходный текст и его перевод взаимно корректируются. Это означает, что необходимо будет определить пару или набор предложений, фраз и слов в исходном тексте и их переводы на другие языки. Параллельное чередование текста очень важно, ведь переводчик в процессе перевода может выделять, объединять, удалять, вводить дополнительные лексические единицы или грамматически изменять порядок предложений в тексте второго языка в соответствии со степенью возможности языка. Параллельные корпусные тексты различаются по жанровому признаку.

В художественном тексте переводчику предоставляется большая свобода, при которой переводчик может опускать некоторые слова в оригинале, заменять слова, добавлять слова для уточнения смысла, поэтому создание параллельного корпуса художественного произведения по сравнению с другими стилями требует времени, отраслевой квалификации, знания стилистических единиц, культуры языков.

При сборке параллельных корпусов текста важным является определение структуры и темы текста. Одним из таких требований является наличие статического или динамического характера текстов или предоставление текста полностью или частично. Автор должен определить объем, тему, жанр и стиль текста.

ЛИТЕРАТУРЫ

1. Abduraxmonova N.Z. Kompyuter lingvistikasi: – Melbrun: Globe edit, 2020
2. Abduraxmonova N. O‘zbek tili elektron korpusining kompyuter modellari (monografiya), Toshkent, 2021. – 200 b.
3. Атабоев Н.Б. Инглиз тили корпусининг функционал хусусиятлари (СОСА мисолида) дис...ф.ф.док. –Тошкент, 2020. – 189 б.
4. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Ф.б.фалс. д-ри (PhD) ... дисс. автореф. – Қарши: 2018. – 52 б.
5. Эргашева Г. Турли тизимдаги тилларда гендерга оид терминологиянинг шаклланишида лингвистик ва экстралингвистик омиллар: Фил.фан.д-ри.дисс. автореф. – Тошкент: 2018. – 61 б. URL: <http://demo.navoiy-uni.uz/sites/default/files/ziyonet.pdf>
6. Джумабаева Ж. Ўзбек ва инглиз тилларида лексик ҳамда стилистик градуонимия. докторлик диссертацияси автореферати. – Тошкент: 2016. – 82 б. URL: http://tashgiv.uz/wp-content/uploads/2017/01/ avtoref_jumabayeva.pdf.
7. Hunston S. Corpora in Applied Linguistics: Cambridge University Press, 2002. – 234 p
8. Седов А.В. Математические модели, методы и алгоритмы построения размеченных корпусов текстов. Автореф. дис. на соиск. учен. степ. канд. тех. наук (05.13.18) Петрозаводск, 2013. – 22 с.

9. Atwell E. Development of tagsets for part-of-speech tagging. An international handbook. Corpus Linguistics: Mouton de Gruyter. 2008.
10. Laurence Anthony, AntConc: Design and Development of a Freeware Corpus Analysis Toolkit for the Technical Writing Classroom, 2005 IEEE International Professional Communication Conference Proceedings, – P 729-737
11. Баранов А.Н., Михайлов М.Н., Сидоров Г.О. “Динамический корпус текстов” как новая технология прикладной лингвистики // Труды международного семинара Диалог по компьютерной лингвистике и ее приложениям. – Т.2. 1998.
12. Kay M., RoscheisenM. Text-translation alignment. Computational Linguistics, 1993. № 19 (1), –P. 121-142.
13. Harris B. Bitexts: A new concept in translation theory. LanguageMonthly, 1988. – P. 54
14. Resnik Philip “Mining the Web for bilingual text”, in: Proceedings of the 37th annual meeting of the Association for computational linguistics. 1999.