

РОЛЬ ТЕКСТОВЫХ КОРПУСОВ В СОВРЕМЕННЫХ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЯХ

Ю. Ю. Красовская

*Белорусский государственный университет,
Минск, Беларусь, hisradaar@gmail.com*

В данной статье рассматривается роль текстовых корпусов в современных лингвистических исследованиях. Представлены основные методы и подходы к использованию текстовых корпусов в лингвистических исследованиях, включая анализ лексики, грамматики, семантики и дискурса. Рассматриваются преимущества и ограничения данного метода, а также примеры его успешного применения в области обучения иностранным языкам.

Ключевые слова: корпус текстов; метатекст; учебный корпус; корпусная лингвистика; корпусные методы.

THE ROLE OF TEXT CORPORA IN CONTEMPORARY LINGUISTIC RESEARCH

J. Krasowskaja

*Belarusian State University,
Minsk, Belarus, hisradaar@gmail.com*

The article deals with the role of text corpora in contemporary linguistic research. It presents the main methods and approaches to using text corpora in linguistic studies, including analysis of lexicon, grammar, semantics, and discourse. The advantages and limitations of this method are discussed, along with examples of its successful application in the field of foreign language education.

Keywords: text corpus; metatext; educational corpus; corpus linguistics; corpus methods.

В современных гуманитарных исследованиях, таких как литературоведение, социолингвистика, сравнительная лингвистика и переводоведение, активно используются текстовые корпуса. Это объясняется тем, что корпуса текстов позволяют выявлять закономерности в структуре и организации текстов на разных уровнях. Корпус текста можно рассматривать как несколько уменьшенную модель языка, поэтому электронные корпуса позволяют намного быстрее и эффективнее решать ранее фактически невыполнимые для науки задачи: анализировать тысячи словоформ для получения лингвистической информации, получать десятки тысяч примеров словоупотреблений для проведения лингвистических исследований, изучать язык в диахронии, наблюдать изменение значения языковых единиц или употребления тех или иных языковых конструкций и многое другое. Корпусы позволяют проводить как внутриязыковой, так и межьязыковой

сравнительный анализ различных языковых явлений, а также отслеживать тенденции в развитии языка в разные исторические периоды, определять точность перевода, переводческие идиостили и др.

Интерес пользователей корпусов, прежде всего лингвистов, направлен в основном на метатекстовую информацию и примеры использования языковых элементов и конструкций, а не на конкретное содержание текстов. Исходные лингвистические исследования, проводимые с применением корпусов, ранее ограничивались подсчетом частоты встречаемости различных языковых элементов. Статистические методики находят применение в решении сложных лингвистических задач, таких как разработка словарей, составление грамматик, машинный перевод, распознавание и синтез речи, а также средства проверки орфографии и грамматики. Так, статистические методы, используя материал корпуса, позволяют выявить регулярное соседство слов, что может указывать на их вхождение в устойчивые словосочетания. Эти устойчивые словосочетания представляют собой семантически неделимые единицы. Подход корпусной лингвистики обеспечивает репрезентативность и сбалансированность языкового материала, а также предоставляет инструменты для поиска, обычно позволяющие получить хорошую выборку в конкретном корпусе.

Об этом же пишет Н.В. Перцов, опровергающий суждения авторитетнейших лингвистов о русском языке, используя материал Национального корпуса русского языка: «...Следует признать, что возможности корпусов все-таки еще недостаточно усвоены лингвистической общественностью вообще и лингвистами в частности. Обращение к корпусным данным еще не стало столь же привычным и обязательным при формулировке и проверке тех или иных утверждений относительно фактов языка, как обращение к грамматикам и словарям, к работам коллег...» [1].

Целесообразность и необходимость обращения к корпусам и их применение в комплексе с другими исследовательскими инструментами убедительно обосновывает Ч. Филмор: «Во-первых, я не думаю, что могут существовать какие-либо корпуса, какими бы большими они ни были, которые содержали бы все области английской лексики и грамматики, которые я хочу исследовать; все, что я видел, неадекватно. Второе наблюдение заключается в том, что каждый корпус, который я имел возможность изучить, каким бы маленьким он ни был, научил меня фактам, которые я не мог представить никаким другим способом» [2].

Свидетельства из корпусов могут быть использованы для верификации гипотез на каждом уровне языка, начиная от звуков речи и заканчивая целыми разговорами и текстами [3]. В рамках этой структуры анализ может быть повторен и результаты воспроизведены, что невозможно в про-

цессе метода интроспекции. Например, было опровергнуто часто повторяемое утверждение о том, что частицы в глаголах немецкого языка с отделяемыми приставками не могут находиться в начале предложения. Анализ корпусов текстов выявил значительное количество грамматически верных примеров, где частица находится в начальной позиции.

В исследованиях, посвященных лингвистической эволюции, корпусные данные играют критическую роль. Например, древнеанглийский и среднеанглийский являются языками, утраченными для современного общества, что делает невозможным обращение к живым языковым сообществам для получения данных. Тем не менее, существуют фиксированные корпуса, включающие некоторые тексты на этих мертвых языках, сохранившиеся до настоящего времени. Эти корпуса представляют собой ценный инструмент для количественного анализа языковых изменений, поскольку обеспечивают доступ к репрезентативным данным, необходимым для исследования.

Использование текстовых корпусов является разумным и в контексте дискурсивных исследований. С.Ю. Богданова предлагает обратиться к таким корпусам с целью выявления характеристик определенных типов дискурса и оценки степени соответствия отдельного текста моделям дискурса в данном регистре [4].

Исследование, проведенное Д. Байбером, представляет собой пример анализа лексической семантики, в рамках которого проанализировано употребление английских прилагательных *big*, *large* и *great*, которые, на первый взгляд, считаются синонимами. Исследование основано на анализе частотности употребления этих прилагательных на основе выборки из 5.7 миллионов словоупотреблений из корпуса Longman-Lancaster. Кроме того, были рассмотрены ближайшие коллокации для каждого из этих слов. Результаты анализа указали на то, что упомянутые прилагательные не являются эквивалентными семантически и имеют различные предпочтительные сочетания и распределение в текстах. [5]

Корпусные технологии играют ключевую роль в обучении языкам, предоставляя студентам доступ к реальным языковым материалам. Репрезентативность корпуса обеспечивает широкий спектр примеров, что позволяет студентам изучать языковые явления в различных контекстах. Это способствует более эффективному закреплению теоретических знаний на практике. Кроме того, конкордансы могут использоваться для самоконтроля и при выполнении переводов. Например, студенты могут использовать конкордансы для проверки правильности употребления выражений или словосочетаний, что помогает им лучше усвоить языковые правила.

Необходимо отметить, что, помимо анализа употребления лексических единиц, использование поиска единиц посредством конкордансеров

также способствуют выявлению их сочетаемости с другими словами, определению наиболее часто употребляемых предлогов и прочих параметров. Можно ввести параметр для отображения однокоренных слов, что упростит поиск студентам и позволит им изучать не только исходное слово, но и его производные. В случае с грамматическими конструкциями конкордансы также облегчают анализ их употребления, что способствует практическому применению теоретических знаний.

Вместо предварительно подготовленных материалов с примерами использования конкретных слов или грамматических времен, учащиеся самостоятельно обращаются к онлайн-версии лингвистического корпуса, проводят поиск и анализ данных. Это способствует развитию у обучающихся необходимых стратегий в области языкового образования и самообразования. Затем они могут обратиться к описанию соответствующего правила в учебнике или грамматическом справочнике, сравнить его с собственными выводами и начать использовать новую структуру в коммуникативно-ориентированных заданиях. Здесь в первую очередь следует сказать о коллигации, то есть грамматической сочетаемости слов. Примером коллигации служит управление глаголов. С использованием конкорданса можно проверить, требует ли определенный глагол сопровождения предлогом, а также определить возможность использования после него другого глагола в форме инфинитива или герундия. Корпусные технологии также предоставляют возможность исследования фразовых глаголов, представляющих сложный аспект в процессе обучения.

Так, исследование, проведенное в университете Чан-Ан в Южной Корее, сфокусировалось на оценке способности студентов к самостоятельному использованию корпусов данных. В результате было обнаружено, что общие языковые корпуса служат студентам как источник справочной информации, в то время как специализированные корпуса оказывают непосредственную помощь при работе с научными текстами. Студенты инженерных специальностей, чьим родным языком не является английский, стремятся поднять уровень своей профессиональной речи до уровня носителей языка [6]. В результате этого делается вывод о необходимости развития специализированных корпусов и их регулярной модернизации совместно со студентами. На данный момент существует несколько проектов, в рамках которых студенты активно участвуют в создании специализированных языковых корпусов.

Таким образом, можно утверждать, что корпусная лингвистика охватывает широкий спектр исследований в области языкознания. Эти исследования направлены не только на изучение структуры языка, но также фокусируются на практических аспектах, таких как методы обучения иностранному языку с использованием текстовых корпусов. Этот подход

предоставляет не только понимание языковых явлений, но и помогает в разработке эффективных стратегий обучения, основанных на реальных языковых контекстах.

Библиографические ссылки

1. *Перцов Н.В.* О роли корпусов в лингвистических исследованиях // Корпусная лингвистика 2006 / СПб : Изд-во С.-Петерб. ун-та; Изд-во РХГА, 2006. С. 318-331.
2. *Fillmore C.* Corpus linguistics or Computer-aided armchair linguistics // Directions in Corpus Linguistics: Proceedings of the Nobel Symposium 82, Stockholm, 4–8 August 1991 / Mouton de Gruyter. Berlin, 2011. pp. 35-60.
3. *Захаров В.П.* Корпусная лингвистика: Учебник для студентов направления «Лингвистика». 2-е изд., перераб. и дополн. / В.П. Захаров, С.Ю. Богданова. СПб : СПбГУ. РИО. Филологический факультет, 2013.
4. *Богданова С. Ю.* Возможности корпусной методологии в решении лингвистических задач [Электронный ресурс] URL: <https://cyberleninka.ru/article/n/vozmozhnosti-korpusnoy-metodologii-v-reshenii-lingvisticheskikh-zadach> (дата доступа: 10.01.2024).
5. *Biber D.* Corpus Linguistics. Investigating Language Structure and Use. Cambridge Univ. Press, 1998.
6. *Chang Ji-Yeon* The use of general and specialized corpora as reference sources for academic English writing: A case study // ReCALL. 2014. Volume 26. P. 243-259.