

ИСПОЛЬЗОВАНИЕ КОРПУСНЫХ ДАННЫХ В ЛЕКСИКОГРАФИЧЕСКОМ АНАЛИЗЕ

Джуманиязова Интизор Давлатназар кизи

Негосударственное образовательное учреждение «Университет Маъмуна»

Хива, Узбекистан

Юлдашева Севинчой Хамробек кизи

студентка 1 курса «Университет Маъмуна» Хива, Узбекистан

Аннотация: В статье рассматривается роль корпусных данных в современном лексикографическом анализе. С развитием технологий и доступностью электронных ресурсов корпусные данные стали важным инструментом для углубленного и систематического изучения языка. Рассматриваются ключевые методы, такие как частотный, контекстный и коллокационный анализ, которые позволяют лексикографам более точно определять значения слов, их частотность, грамматические особенности и устойчивые сочетания. Также подчеркиваются преимущества использования корпусов, такие как объективность, актуальность и эффективность, которые значительно улучшили процесс создания словарей. Статья демонстрирует, как корпусные данные помогают не только фиксировать языковую практику, но и отслеживать изменения в языке, делая лексикографию более точной и своевременной.

Ключевые слова: лексикографический анализ, частотный анализ, контекстный анализ, коллокационный анализ, статистический анализ.

Annotation. The article examines the role of corpus data in modern lexicographic analysis. With the development of technology and the availability of electronic resources, corpus data has become an important tool for in-depth and systematic language learning. Key methods such as frequency, contextual, and collocation analysis are considered, which

allow lexicographers to more accurately determine the meanings of words, their frequency, grammatical features, and stable combinations. The advantages of using corpora are also highlighted, such as objectivity, relevance, and efficiency, which have significantly improved the dictionary creation process. The article demonstrates how corpus data helps not only to record language practice, but also to track changes in language, making lexicography more accurate and timely.

Keywords: lexicographic analysis, frequency analysis, contextual analysis, collocation analysis, statistical analysis.

С момента внедрения информационных технологий в гуманитарные дисциплины прошло более десяти лет. За этот период появились такие самостоятельные направления в филологии, как компьютерная лингвистика, корпусная лингвистика и компьютерная лексикография. В этих, преимущественно прикладных областях, были разработаны инструменты, существенно упрощающие работу исследователей. Речь идет, прежде всего, об языковых корпусах и электронных лексикографических ресурсах. Процесс разработки и усовершенствования этих инструментов существенно различался. В корпусной лингвистике лидирующие позиции занимали филологи, тогда как в компьютерной лексикографии основную роль играли программисты и предприниматели, что, без сомнения, связано с рыночными реалиями современности: потребность в электронных лексикографических продуктах определяется практическими нуждами. В результате, на сегодняшний день самыми продвинутыми и многофункциональными являются переводные многоязычные словари. В то же время специализированные лексикографические продукты, хотя и достигли определенного прогресса, всё еще не получили должного внимания для дальнейшего развития.

Корпусные данные играют ключевую роль в современном лексикографическом анализе. С развитием вычислительных технологий и

увеличением доступных электронных ресурсов, лексикографы получили мощные инструменты для более точного и систематического анализа языка. Использование корпусов текста позволяет углубить понимание закономерностей языка, а также улучшить точность и актуальность лексикографических работ. В данной статье мы рассмотрим, как корпусные данные помогают в лексикографическом анализе и какие преимущества они предоставляют.

1. Что такое корпус данных?

Корпус данных (или корпус) представляет собой большую коллекцию текстов, собранных в соответствии с определенными критериями. Эти тексты могут быть разных жанров, стилей и времени написания. Корпус может быть как одноязычным, так и многоязычным, и использоваться для анализа языковых особенностей на различных уровнях: лексическом, грамматическом, стилистическом и других.

Корпусы данных могут быть как общими, так и специализированными. Общий корпус включает в себя разнообразные тексты (например, художественную литературу, научные статьи, газетные материалы), в то время как специализированные корпуса могут быть ориентированы на конкретные области знания, такие как медицина, право или техника.

2. Значение корпуса в лексикографическом анализе.

Лексикографический анализ заключается в изучении лексических единиц языка — слов, выражений и их значений. С традиционными методами исследования лексикографы опирались на ограниченные источники, такие как словари, тексты и устные источники. Однако это не всегда позволяло учитывать все возможные нюансы употребления слов.

Использование корпуса в лексикографии даёт возможность:

1) Изучать частотность употребления слов. Корпус позволяет анализировать, как часто и в каких контекстах встречаются те или иные лексемы. Это помогает

составителям словарей учитывать в своей работе не только те слова, которые известны и часто употребляются, но и те, которые могут быть в моде в данный момент или актуальны в определенных профессиональных областях.

2) Определять полные и актуальные значения слов. Благодаря корпусу можно выявить все возможные значения слова, а также различия в его употреблении в разных контекстах. Например, слово может иметь несколько значений в зависимости от жанра текста, аудитории или исторического периода.

3) Описывать коллокации и устойчивые выражения. Корпусные исследования позволяют выявлять коллокации — устойчивые сочетания слов, которые часто встречаются вместе. Это важно для точности определения значений слов в словарях, поскольку такие сочетания могут сильно менять или уточнять смысл слов.

4) Оценивать грамматические и синтаксические особенности. Корпусные данные помогают изучать, как слова и фразы функционируют в предложениях, выявлять их синтаксическую связь и способы образования различных грамматических форм.

5) Анализировать языковую эволюцию. С помощью корпусных данных можно отслеживать изменения в употреблении слов, например, изменение их частотности, стилистического окраса или расширение значений.

3. Методы лексикографического анализа с использованием корпуса

Для проведения лексикографического анализа используются различные методы, включая автоматические и ручные. В первую очередь, лексикографы используют программное обеспечение для анализа текста, которое позволяет быстро и точно извлекать нужную информацию из больших массивов данных.

1. Частотный анализ. Это один из наиболее распространенных методов. Он позволяет выявить самые часто употребляемые слова и их сочетания в тексте.

Частотный анализ помогает лексикографам выбрать слова для словаря, а также установить их актуальность в разных областях.

2. Контекстный анализ. Программное обеспечение может искать контексты, в которых встречается определённое слово. Это важно для понимания, какие значения оно может приобретать в зависимости от контекста, а также для выявления синонимов и антонимов.

3. Коллокационный анализ. Этот метод помогает выделить устойчивые словосочетания и выражения, которые часто встречаются вместе с определённым словом. Это знание крайне полезно для точного описания значений слов в словаре.

4. Статистический анализ. С помощью статистических методов можно определить, как распределяются различные лексемы по корпусу, а также оценить их значимость в текстах разных жанров и стилей.

4. Преимущества использования корпуса в лексикографическом анализе.

Корпусные данные предоставляют несколько ключевых преимуществ для лексикографов:

А) Объективность. Использование реальных текстов позволяет избежать субъективности, которая может быть присуща традиционному подходу, основанному только на авторских наблюдениях.

Б) Актуальность. Корпус может включать тексты, отражающие современную языковую практику, что помогает создавать актуальные словари, соответствующие современным нормам и тенденциям языка.

В) Полнота. С помощью корпусов можно учитывать разнообразие употреблений слов в различных контекстах, что позволяет создавать более полные и точные словарные статьи.

Г) Эффективность. Корпус позволяет автоматически анализировать огромные объемы текстов за короткое время, что значительно ускоряет процесс лексикографической работы.

5. Заключение

Использование корпусных данных в лексикографическом анализе стало неотъемлемой частью современной лексикографии. С помощью корпусов лексикографы могут глубже понять реальные закономерности в употреблении слов, выявлять их различные значения и коллокации, а также отслеживать изменения языка. Это делает процесс составления словарей более точным, актуальным и эффективным. В будущем с развитием технологий и расширением доступа к большим данным, корпусная лексикография будет продолжать развиваться, открывая новые возможности для анализа и описания языка.

Список литературы

1. Баранов, С. Ю. (2013). *Корпусная лексикография: Теория и практика*. Москва: Наука.
2. Кузьменко, Н. В. (2007). *Корпусная лексикография и её методы*. Санкт-Петербург: Издательство РГПУ.
3. Лопатин, В. М. (2006). *Лексикография и корпуса: Проблемы и перспективы*. Москва: Издательство МГУ.
4. Виноградова, Е. М. (2014). *Основы корпусной лексикографии*. Санкт-Петербург: Алетейя.
5. Гаврилова, Л. И., & Никитина, О. В. (2011). *Методы анализа лексики с использованием корпусных данных*. Москва: Высшая школа.
6. Воскресенский, А. И. (2015). *Словарь русского языка в эпоху цифровых технологий*. Москва: Академический проект.