

КОРПУСНОЕ ИССЛЕДОВАНИЕ РУССКОГО ЯЗЫКА

Махамаджонова Гулсанам Валижон кизи*Магистрант 1- курса ТерГУ направления Лингвистика (русский язык)*

Аннотация. В статье рассматриваются теоретические основы корпусного исследования русского языка. Анализируются понятие языкового корпуса, история развития корпусной лингвистики, принципы построения корпусов и методы корпусного анализа. Особое внимание уделяется Национальному корпусу русского языка как основному инструменту эмпирического изучения лексических и грамматических особенностей современного русского языка. Показаны основные направления практического применения корпусных исследований в лингвистике и образовании. Материал статьи может быть использован в учебной и научной работе студентов магистранта.

Ключевые слова: корпусная лингвистика, языковой корпус, Национальный корпус русского языка, корпусный анализ, русский язык, лингвистическое исследование.

Abstract. This article examines the theoretical foundations of corpus-based research on the Russian language. It analyzes the concept of a language corpus, the history of corpus linguistics, the principles of corpus construction, and methods of corpus analysis. Particular attention is paid to the National Corpus of the Russian Language as the primary tool for empirically studying the lexical and grammatical features of the modern Russian language. The article also highlights the key areas of practical application of corpus research in linguistics and education. This article can be used in the academic and research work of graduate students.

Keywords: corpus linguistics, language corpus, National Corpus of the Russian Language, corpus analysis, Russian language, linguistic research.

ВВЕДЕНИЕ.

Современная лингвистика характеризуется активным внедрением цифровых технологий в научные исследования. Одним из наиболее значимых направлений стала корпусная лингвистика, основанная на анализе больших массивов текстов, представленных в электронном виде. Корпусное исследование русского языка позволяет изучать реальные языковые факты, опираясь не на интуицию исследователя, а на объективные данные.

Актуальность данной темы обусловлена ростом интереса к эмпирическим методам анализа языка, а также расширением возможностей Национального корпуса русского языка. Корпусные методы активно

применяются в лексикографии, грамматике, стилистике, социолингвистике и преподавании русского языка. Целью данной работы является рассмотрение теоретических основ корпусного исследования русского языка и анализ основных направлений его применения. В задачи работы входят: определение понятия корпуса, рассмотрение истории развития корпусной лингвистики, описание структуры корпусов русского языка и методов корпусного анализа.

1. Понятие корпусной лингвистики.

Корпусная лингвистика — это раздел языкознания, изучающий язык на основе специально организованных и размеченных электронных корпусов текстов. В отличие от традиционных методов, корпусный подход опирается на большие объемы аутентичных текстов, что обеспечивает объективность результатов.

Корпус представляет собой репрезентативную выборку текстов, отражающую различные жанры, стили, эпохи и формы существования языка. Основной задачей корпусной лингвистики является выявление закономерностей функционирования языковых единиц в реальной коммуникации.

2. История развития корпусных исследований.

Истоки корпусной лингвистики относятся к середине XX века, когда появились первые машинно-читаемые коллекции текстов. Значительным этапом стало создание Брауновского корпуса английского языка. В России активное развитие корпусных исследований началось в конце XX — начале XXI века. Создание Национального корпуса русского языка стало важнейшим событием в отечественной лингвистике, открыв новые возможности для научных исследований и образовательной практики.

3. Принципы построения языкового корпуса.

Основными принципами построения корпуса являются репрезентативность, сбалансированность, электронная форма, разметка и открытость. Репрезентативность предполагает адекватное представление языкового материала. Разметка является ключевой характеристикой корпуса. Она включает морфологическую, синтаксическую, семантическую и прагматическую разметку, что значительно расширяет аналитические возможности исследователя.

4. Национальный корпус русского языка.

Национальный корпус русского языка (НКРЯ) представляет собой крупнейший электронный ресурс для изучения русского языка. Он включает основной корпус, газетный, устный, диалектный, исторический и параллельные корпуса. НКРЯ используется как в научных исследованиях, так

и в образовательных целях. Его интерфейс позволяет осуществлять поиск по словам, грамматическим признакам и контексту.

5. Методы корпусного анализа.

Корпусное исследование использует как количественные, так и качественные методы анализа. Частотный анализ позволяет выявить наиболее употребительные языковые единицы. Контекстуальный анализ применяется для изучения значений слов, коллокаций и фразеологических единиц. Сравнительный анализ используется для сопоставления различных периодов, жанров и стилей.

6. Корпусные исследования лексики.

Корпусные данные широко применяются в лексикологии и лексикографии. Они позволяют выявлять новые значения слов, изучать динамику лексических изменений и фиксировать неологизмы. На основе корпусных данных создаются современные словари, отражающие реальное употребление слов и выражений.

7. Корпусное изучение грамматики.

Грамматические исследования на корпусном материале позволяют анализировать функционирование морфологических и синтаксических конструкций в реальной речи. Корпусный подход способствует пересмотру традиционных грамматических описаний и уточнению языковых норм.

8. Практическое применение корпусных исследований.

Корпусная лингвистика активно используется в преподавании русского языка, переводоведении, компьютерной лингвистике и разработке систем искусственного интеллекта. Корпусные данные позволяют создавать учебные материалы, основанные на реальных языковых примерах.

9. Преимущества и ограничения корпусного подхода.

К основным преимуществам корпусного подхода относятся объективность, масштабность и воспроизводимость результатов. Однако корпусные исследования зависят от качества разметки и состава корпуса. Ограничением также является невозможность полного отражения всех аспектов языковой системы.

ЗАКЛЮЧЕНИЕ

Корпусное исследование русского языка представляет собой одно из наиболее значимых и методологически продуктивных направлений современной лингвистики, объединяющее достижения теоретического языкознания, информационных технологий и статистического анализа. Использование языковых корпусов позволяет исследователю выйти за пределы интуитивных наблюдений и оперировать объективными, эмпирически подтверждёнными данными, отражающими реальное функционирование языковых единиц в различных сферах коммуникации.

В ходе исследования было показано, что корпусная лингвистика обеспечивает принципиально новый уровень анализа языкового материала, поскольку позволяет учитывать частотность, контекстуальную обусловленность и вариативность языковых явлений. Особенно важным является тот факт, что корпусные методы дают возможность выявлять скрытые закономерности языка, которые не всегда фиксируются в традиционных грамматиках и словарях. Это способствует уточнению лингвистических теорий и пересмотру устоявшихся представлений о норме и узусе.

Национальный корпус русского языка занимает центральное место в корпусных исследованиях, выступая универсальным инструментом для анализа лексических, грамматических и стилистических особенностей русского языка. Его многоуровневая структура и развитая система разметки позволяют проводить как масштабные статистические исследования, так и детальный анализ отдельных языковых единиц в контексте. Благодаря этому НКРЯ является надёжной эмпирической базой для магистерских и диссертационных работ.

Особую значимость корпусная лингвистика приобретает в междисциплинарном аспекте. Корпусные данные активно используются в социолингвистике, когнитивной лингвистике, переводоведении, прикладной лингвистике и методике преподавания русского языка как родного и как иностранного. Это расширяет границы традиционного языкознания и способствует интеграции лингвистических исследований в более широкий научный контекст.

Для магистрантов освоение корпусных методов анализа является не только важным элементом научной подготовки, но и необходимым условием формирования профессиональных исследовательских компетенций. Умение работать с корпусами развивает навыки критического мышления, аналитической обработки данных и аргументированного научного изложения. Кроме того, корпусный подход повышает практическую ценность магистерских исследований, делая их результаты более доказательными и востребованными в научном сообществе.

В перспективе дальнейшее развитие корпусной лингвистики связано с расширением объёма корпусов, углублением семантической и прагматической разметки, а также с внедрением методов автоматической обработки больших данных и искусственного интеллекта. Это открывает новые возможности для более точного и комплексного изучения русского языка, а также для создания инновационных лингвистических ресурсов и образовательных технологий.

Таким образом, корпусное исследование русского языка выступает важнейшим инструментом современной лингвистики, обеспечивающим научную объективность, методологическую строгость и практическую значимость исследований. Его активное использование в магистерской подготовке способствует повышению качества научных работ и формированию нового поколения исследователей, ориентированных на современные методы анализа языка.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ:

1. Норма языковая // Большая российская энциклопедия. Том 23. — М., 2013. — С. 326—327.
2. Язык в законе. В каком? // Российская газета. — 2002. (О попытке законодательного регулирования употребления языка в России.)
3. Митрофанова А. Глобализация и языковая политика // Топ-менеджер. — 2004. — № 625. (О языках, политике и их влиянии друг на друга.)
4. Норма и вариативность в языке и речи: Сб. науч. трудов / РАН. ИНИОН. Центр гуманитар. науч.-информ. исслед. Отд. языкознания; Отв. ред. Казак Е. А. — М., 2017. — 193 с. (Сер. Теория и история языкознания).
5. Баранов, А. Н., Плунгян, В. А., Сичинава, Д. В. Национальный корпус русского языка: 2006–2021. — М.: РАН, 2021.
6. Копотев, М. В. Корпусная лингвистика: методы, инструменты и практика. — М.: ЯСК, 2018.
7. Плунгян, В. А. Корпусная грамматика русского языка. — М.: РГГУ, 2019.
8. Ляшевская, О. Н. Корпусное исследование русской лексики и грамматики. — М.: ЯСК, 2015.
9. Апресян, Ю. Д. Новые подходы в компьютерной лексикографии. — М.: РАН, 2012.
10. Назарова, Т. Б. Цифровая лингвистика: методы анализа языковых данных. — СПб.: СПбГУ, 2020.