

*Хеббиккулиева Дуньягозель,
преподаватель кафедры теории и практики русского языка
Туркменского национального института
мировых языков имени Довлетмаммеда Азади,
Ашхабад, Туркменистан.*

**ЛЕКСИКОГРАФИЯ И КОРПУСНАЯ ЛИНГВИСТИКА:
ПЕРЕСЕЧЕНИЕ ТЕОРИИ И ПРАКТИКИ**

Аннотация: Лексикография и корпусная лингвистика являются двумя важными направлениями современной лингвистики, каждое из которых имеет свои теоретические и практические задачи. В данной работе рассматривается взаимодействие этих дисциплин, которое значительно изменило методы создания словарей, анализа языковых единиц и изучения языковых закономерностей. В статье приводятся примеры интеграции корпусных данных в лексикографическую практику, обсуждаются перспективы использования искусственного интеллекта и машинного обучения в лексикографии, а также анализируются ключевые проблемы и пути их решения.

Ключевые слова: лексикография, корпусная лингвистика, анализ, словари.

*Hebbikulyeva Dunyagozel,
lecturer
at the Department of Theory and Practice of the Russian Language
Turkmen National Institute of
world languages named after Dovletmammet Azady
Ashgabat, Turkmenistan.*

**LEXICOGRAPHY AND CORPUS LINGUISTICS: THE
INTERSECTION OF THEORY AND PRACTICE**

Abstract: Lexicography and corpus linguistics are two important areas of modern linguistics, each of which has its own theoretical and practical tasks. This paper examines the interaction of these disciplines, which has significantly changed the methods of creating dictionaries, analyzing linguistic units and studying linguistic patterns. The article provides examples of the integration of corpus data into lexicographic practice, discusses the prospects for using artificial intelligence and machine learning in lexicography, and analyzes key problems and ways to solve them.

Keywords: lexicography, corpus linguistics, analysis, dictionaries.

Лексикография традиционно рассматривалась как искусство составления словарей, объединяющее в себе интуитивные и эмпирические подходы к описанию языка. Однако с развитием корпусной лингвистики, основанной на анализе больших массивов текстов, эта дисциплина приобрела новый импульс к развитию. Корпусная лингвистика предлагает методы анализа данных, которые обеспечивают более точное, детализированное и актуальное представление языковой системы.

Современные корпусные технологии позволяют не только анализировать частотность слов и устойчивость фраз, но и выявлять скрытые семантические и прагматические закономерности. Это открывает новые горизонты для лексикографии, позволяя составлять словари, которые не только фиксируют языковую норму, но и описывают её динамику. Основной задачей данной работы является анализ взаимодействия корпусной лингвистики и лексикографии с точки зрения теоретического и прикладного аспекта, а также выявление перспективных направлений их синтеза.

Теоретические основания взаимодействия

Лексикография всегда нуждалась в эмпирической основе. В традиционной практике эта основа представляла собой ограниченный набор текстов, которые лексикографы анализировали вручную. Однако

такой подход имел очевидные ограничения: низкая репрезентативность данных, субъективность интерпретации и невозможность охватить весь спектр вариативности языка. Корпусная лингвистика устраняет эти недостатки, предоставляя систематизированные и репрезентативные данные о языке.

Основным инструментом корпусной лингвистики являются текстовые корпуса — большие коллекции текстов, которые охватывают различные жанры, стили, регионы и временные периоды. Анализ корпусных данных осуществляется с помощью автоматизированных программ, что позволяет проводить сложные вычисления, такие как анализ частотности, выявление коллокаций, построение конкордансов и исследование грамматических конструкций. Применение таких методов в лексикографии способствует созданию словарей, которые более точно отражают реальные языковые процессы.

Анализ взаимодействия методов

Примером успешной интеграции корпусной лингвистики и лексикографии является создание новых типов словарей, таких как корпусные и фразеологические словари. Корпусные словари используют данные текстовых корпусов для определения значений слов, их сочетаемости и частотности. Например, в процессе создания корпусного словаря русского языка использовались данные Российского национального корпуса, что позволило выявить не только основные значения слов, но и частотные контексты их употребления.

Фразеологические словари, созданные на основе корпуса, содержат информацию об устойчивых выражениях, их семантике и употреблении. Например, фразеологизм «ломать голову» может быть описан не только с точки зрения его значения, но и с учётом частотности употребления, типичных контекстов и вариативности. Корпусный анализ показал, что это выражение чаще встречается в публицистике, чем в научной литературе,

что является важным для описания прагматических характеристик фразеологизма.

Ещё одним направлением применения корпусной лингвистики является создание словарей неологизмов. Корпусные методы позволяют оперативно выявлять новые слова и выражения, анализировать их частотность и контексты употребления. Например, слова, связанные с современными технологиями («гаджет», «подписчик», «инфлюенсер»), становятся объектами изучения с помощью корпусных данных, которые помогают определить их статус в языке: временные это явления или устойчивые единицы.

Новые подходы и перспективы

Современная лексикография активно использует методы машинного обучения и искусственного интеллекта. Эти технологии позволяют автоматизировать процесс создания словарей. Например, алгоритмы машинного обучения могут анализировать большие массивы текстов, выявлять значимые статистические закономерности и автоматически формировать словарные статьи. Такой подход особенно полезен при создании многоязычных словарей, где требуется сопоставить языковые единицы разных языков на основе их использования в корпусах.

Одним из перспективных направлений является создание интерактивных словарей, которые не только фиксируют существующую языковую норму, но и позволяют пользователям добавлять новые значения и контексты. Такие словари могут обновляться в реальном времени, интегрируя данные из новых текстов, что обеспечивает их актуальность и универсальность.

Лексикография и корпусная лингвистика, взаимодействуя, создают новые возможности для описания языка. Использование корпусных данных позволяет не только повышать точность словарей, но и открывает новые горизонты для изучения языковой системы. В то же время

интеграция методов искусственного интеллекта и машинного обучения делает процесс создания словарей более эффективным и технологически продвинутым. Однако остаются и проблемы, такие как выбор репрезентативных данных для корпусов, сложности в интерпретации полученных результатов и необходимость адаптации традиционных методов лексикографии к новым реалиям. Решение этих вопросов станет важной задачей будущих исследований в данной области.

Литература:

1. Быков В.В., Плунгян В.А. Российский национальный корпус: структура и использование в лингвистических исследованиях // Вопросы языкознания. 2018. № 3. С. 64–78.

6. Баранов А.Н. Лексическая семантика и корпусные исследования. Москва: Языки славянской культуры, 2021.

5. Российский национальный корпус. URL: <https://ruscorpora.ru>.

4. Рецкер Я.И. Теория перевода и проблемы лексикографии. Москва: Прогресс, 2019.

3. Лопатин В.В., Михайлов А.А. Корпусная лингвистика: методы и технологии. Санкт-Петербург: Наука, 2021.

2. Грамматическая информация в корпусах русского языка / Под ред. Н.Ю. Шведовой. Москва: Издательство РАН, 2020.