

Ловец А.Д.

*Белорусский государственный медицинский университет, г. Минск  
Беларусь*

## **КОРПУСНЫЕ МЕТОДИКИ В КАЧЕСТВЕ ИНСТРУМЕНТА ИССЛЕДОВАНИЯ И РЕПРЕЗЕНТАЦИИ ЯЗЫКОВОЙ ПРАКТИКИ**

**Аннотация:** В данной работе исследуется роль корпусных методик как мощного инструмента для анализа и репрезентации языковой практики. Статья демонстрирует, как использование электронных коллекций текстов (языковых корпусов) позволяет перейти от теоретических предположений к эмпирически обоснованным выводам о функционировании языка. Мы обсуждаем преимущества корпусного подхода, такие как объективность, масштабность и возможность выявления неочевидных языковых тенденций, а также его значение для точного представления реальной языковой картины.

**Ключевые слова:** корпусная лингвистика, методы исследования, репрезентация языкового материала.

Lovets A.D.

*Belarusian State Medical University, Minsk, Belarus*

## **CORPUS TECHNIQUES AS A TOOL FOR RESEARCH AND REPRESENTATION OF LANGUAGE PRACTICE**

**Abstract:** This study examines the role of corpus techniques as a powerful tool for analyzing and representing language practice. The article demonstrates how the use of electronic collections of texts (language corpora) allows us to move from theoretical assumptions to empirically based conclusions about the functioning of language. We discuss the advantages of the corpus approach, such as objectivity, scalability, and the ability to identify non-obvious linguistic trends, as well as its importance for accurately representing the real linguistic picture.

**Keywords:** corpus linguistics, research methods, representation of linguistic material.

За последние годы корпусная лингвистика завоевала статус одного из наиболее эффективных и перспективных направлений в языкознании. Предоставив беспрецедентные возможности для эмпирического исследования языковых феноменов на базе огромных массивов реальных текстов, она кардинально трансформировала методологию множества лингвистических отраслей — от грамматики и лексикологии до стилистики и социолингвистики. Возможность получать количественные показатели относительно частотности слов, коллокаций, паттернов употребления и вариативности языковых единиц позволила исследователям перейти от интроспективных и интуитивных оценок к объективным, поддающимся проверке знаниям.

В последние десятилетия лингвистика всё активнее изучает закономерности совместного использования лексических единиц, что вывело коллокации и их классификацию на передний план комбинаторной лингвистики. Несмотря на то, что термин «коллокация» вошёл в активный оборот сравнительно недавно, учёные давно исследуют устойчивые сочетания и фразеологизмы в живой речи. Само слово впервые было использовано Г. Харрисом в 1750 году, а его научная значимость была признана лишь спустя почти два века благодаря Дж. Р. Фёрсу и Г. Палмеру, включившим его в академический лексикон. Коллокация остаётся сложным феноменом без единого определения: одни авторы (Дж. Р. Фёрс, М. А. К. Халлидей, О. С. Ахманова)

рассматривают её как устойчивую сочетаемость слов, другие (Ф. де Соссюр, В. В. Виноградов) акцентируют внимание на семантической и синтаксической целостности. Однако все существующие трактовки сходятся в одном: в основе коллокации лежит семантико-грамматическая взаимозависимость компонентов словосочетания [3, с. 25].

Понятие «коллокация» было введено в отечественную лингвистику В. В. Виноградовым, который разграничил его с обычным словосочетанием. По его мнению, словосочетание представляет собой устойчивое объединение нескольких знаковых единиц, не являющееся предложением, но способное выражать единое понятие благодаря совокупности своих частей [4, с. 5]. Будучи языковым знаком, такая конструкция служит для обозначения специфических категорий, присущих определенной научной или профессиональной сфере, и органично включается в общую систему. Внутри этой системы она устанавливает связи как с единицами того же уровня, так и с элементами других уровней, тем самым содействуя созданию целостности дискурса.

Ввиду недостаточной изученности функции коллокаций в лингвистике их часто рассматривают как промежуточный тип между свободными и несвободными сочетаниями. В свободных конструкциях слова можно заменять синонимами или опускать, не теряя общего смысла. В несвободных сочетаниях смысл не может быть выведен исключительно из значений отдельных компонентов. В коллокациях же каждый элемент воздействует на понимание целого, порождая новое, оригинальное значение, которое превышает простую сумму составляющих [4, с. 344].

Тем не менее, несмотря на значительные успехи, корпусная лингвистика сталкивается с серьезными методологическими и эпистемологическими трудностями. Одной из ключевых среди них является проблема репрезентативности. Дискуссионным остается вопрос о том, насколько полно, точно и адекватно данные корпусов способны отражать всю сложность, изменчивость и многогранность реальной речевой деятельности, что требует постоянного критического переосмысления.

В нашем исследовании мы постарались дать определения корпуса текстов. Если говорить простыми словами, корпус – это огромная, тщательно собранная коллекция текстов и/или аудиозаписей. Это может быть всё что угодно: от классической литературы и научных статей до постов в социальных сетях, диалогов из фильмов или записей уличных разговоров. Главное, чтобы эти материалы были репрезентативными, то есть отражали разнообразие языкового употребления. Корпусы – это не просто «склады» текстов. Они становятся мощным инструментом для исследователей языка, потому что позволяют увидеть язык в действии, выявить закономерности и тенденции, понять, как люди действительно говорят и пишут, создавать более точные словари и учебники. Разрабатывать технологии, связанные с языком.

Речевая практика — это не просто набор текстов или высказываний; это живой, интерактивный процесс, обусловленный контекстом и глубоко укорененный в социокультурной среде. Она включает невербальные элементы, такие как интонация, паузы, жесты и мимика, а также сложную систему

социальных ролей, намерений и фоновых знаний, которые часто остаются вне зоны внимания традиционного корпусного анализа [1, с. 46].

Прежде чем обсуждать вопросы репрезентации, важно четко определить понятие речевой практики. В широком понимании, это совокупность всех видов речевой деятельности, реализуемых индивидами и сообществами в реальных жизненных ситуациях. Она охватывает не только создание и восприятие языковых знаков, но и весь комплекс социокультурных, психологических и когнитивных процессов, которые сопровождают и формируют это взаимодействие.

Анализ ключевых характеристик речевой практики показывает, почему ее репрезентация в корпусах представляет особую сложность:

1) Динамичность и процессуальность: речь — это не статичный продукт, а непрерывный процесс, разворачивающийся во времени. Она отличается спонтанностью, незавершенностью, самокоррекцией, переключением кодов, гезитациями и повторами. Корпус же по своей природе фиксирует уже свершившееся, представляя собой «замороженный» срез речи.

2) Зависимость от контекста: любое высказывание тесно связано с конкретной ситуацией — физической, социальной, культурной или интертекстуальной. Чтобы верно истолковать смысл реплики, необходимо учитывать обстоятельства общения, роли участников, их фоновые знания и коммуникативные намерения. Однако крупные текстовые базы данных часто «сглаживают» или полностью игнорируют этот микроконтекст, рассматривая тексты как изолированные единицы.

3) Интерактивность и диалогичность: значительная часть речевого общения носит диалогический характер. Диалог и дискуссия представляют собой совместный процесс создания смысла, где реплики участников взаимозависимы. Корпусы, ориентированные на письменную речь или монологи, плохо фиксируют эту динамику. Даже в базах устной речи сложность заключается в корректном аннотировании таких элементов, как перебивания, эллипсы, реакции слушателей и невербальные сигналы.

4) Многоканальность и мультимодальность: живая коммуникация не ограничивается только словами. Интонация, тембр, громкость, паузы, жесты, мимика и зрительный контакт несут важную смысловую нагрузку. Традиционные текстовые корпуса игнорируют эти аспекты. Хотя мультимодальные базы данных представляют собой прогресс, их создание и анализ сопряжены с серьезными техническими и методологическими трудностями.

5) Социокультурная и прагматическая сложность: речевая практика неразрывно связана с социальными нормами, культурными конвенциями, речевыми актами, имплицатурами, вежливостью и стратегиями убеждения. Эти аспекты глубоко укоренены в сознании носителей и проявляются в нюансах лексики, синтаксиса и интонации. Корпусы способны зафиксировать лишь поверхностные проявления этих явлений, но не их глубинные механизмы и интенции.

б) Вариативность и идиолект: речь каждого индивида уникальна (идиолект), но в то же время вписывается в рамки диалектов, социолектов, регистров и стилей. Язык постоянно эволюционирует, адаптируясь к новым условиям и социальным группам. Поскольку корпуса фиксируют язык на определенный момент времени, они могут не отражать всей динамики изменений и богатства вариативности [2].

Таким образом, речевая практика — это сложный многомерный феномен, который невозможно полностью «упаковать» в статичную линейную структуру корпуса без потери существенных аспектов его сущности. Проблемы репрезентации речевой практики в корпусах проявляются на каждом этапе исследования: от формирования базы данных до интерпретации результатов.

Ключевым постулатом корпусной лингвистики выступает репрезентативность, то есть способность языкового массива достоверно отражать изучаемый язык или его специфические варианты. Тем не менее, реализация этой цели представляет собой одну из наиболее трудных проблем в данной области.

Ввиду отсутствия единого общепринятого толкования термина, современные исследователи сталкиваются с тем, что каждый созданный корпус служит лишь фрагментарным зеркалом действительности, адаптированным под узкие научно-исследовательские задачи.

Традиционно крупнейшие и наиболее ранние лингвистические базы данных фокусировались преимущественно на письменных текстах, таких как художественная литература, публицистика и научные работы. Подобный подход спровоцировал существенный дисбаланс в изучении языка, поскольку были упущены из виду характеристики спонтанной устной коммуникации, доминирующей в повседневном общении. Устная речь отличается фрагментарностью, эллиптичностью, частым использованием повторов и междометий; эти особенности зачастую остаются без должного отражения в корпусах, если они не были разработаны специально для фиксации устных высказываний и не снабжены соответствующей аннотацией [4, с. 27].

Формирование языкового корпуса во многом обусловлено тем, насколько легко и доступно можно извлечь материал. Так, интернет-контент (блоги, форумы, соцсети) добывается без труда, однако он зачастую не отражает всю полноту языка, фиксируя лишь узкие стилистические ниши или особенности отдельных групп. Напротив, сбор спонтанной устной речи сопряжен с серьезными трудностями: процесс требует больших затрат времени, средств и усилий, а также решения этических вопросов, что существенно сужает возможности масштабирования таких исследований.

### **Литература**

1. Гришина, Е. А. Корпусные исследования в лингвистике: учебное пособие / Е.А. Гришина. – М.: Флинта, 2015. – 168 с.
2. Захаров, В. П. Корпусная лингвистика: методология, инструменты, приложения / В.П. Захаров, О.В. Богданова. – Изд-во СПбГУ, 2018. – 184 с.
3. Кибрик, А. А. Корпусные методы в лингвистике / А.А. Кибрик // Вестник Московского университета. Серия 9. Филология. – 2001. – № 2. – С. 104-118.

4. Рыков, В. В. Корпусная лингвистика и ее применение в исследовании русского языка / В.В. Рыков. – М.: Флинта, 2011. – 208 с.