

**КОРПУСНАЯ ЛИНГВИСТИКА КАК НОВЕЙШЕЕ
НАПРАВЛЕНИЕ В ЯЗЫКОЗНАНИИ**

Джуманиязова Интизор Давлатназар кизи
преподаватель, НОУ «Университет-Маъмуна», Хива

**CORPUS LINGUISTICS AS THE NEWEST
DIRECTION IN LINGUISTICS**

Djumaniyazova Intizor Davlatnazar qizi
teacher, NOU "University-Mamun", Khiva

**KORPUS TILSHUNOSLIGI TILSHUNOSLIKNING
ENG YANGI YO'NALISHI SIFATIDA**

Djumaniyazova Intizor Davlatnazar qizi
"Ma'mun-universiteti" NTM o'qituvchisi, Xiva



intizor1828@gmail.com

<https://orcid.org/0009-0005-6284-2320>

Tel: +998 99 544 10 20

Аннотация: В настоящее время существует довольно большое число языковых корпусов, в том числе и корпусов русского языка, которые отличаются друг от друга по самым разным параметрам. В статье рассматривается краткий анализ истории возникновения корпусной языковой лингвистики и определяются ее задачи, методы, выявляется её взаимодействие с иными языковыми дисциплинами, даётся общая характеристика корпуса как основного понятия корпусной лингвистики, представляется классификация корпусов текстов.

Ключевые слова: лингвистический корпус, культурометрия, методология, корпусно-ориентированный подход, корпусно-управляемый подход.

Annotation. Currently, there is a fairly large number of language corpora, including Russian language corpora, which differ from each other in a variety of ways. The article considers a brief analysis of the history of corpus linguistic linguistics and defines its tasks and methods, identifies its interaction with other linguistic disciplines, gives a general description of the corpus as the basic concept of corpus linguistics, and presents a classification of text corpora.

Key words: linguistic corpus, culturometry, methodology, corpus-oriented approach, corpus-guided approach.

Annotatsiya: Hozirgi vaqtda turli xil parametrlarda bir-biridan farq qiladigan juda ko'p sonli til korpuslari, shu jumladan rus tilidagi korpuslar mavjud. Maqolada korpus tilshunosligining paydo bo'lishi tarixining qisqacha tahlili ko'rib chiqiladi va uning vazifalari, usullari aniqlanadi, uning boshqa til fanlari bilan o'zaro ta'siri aniqlanadi, korpus lingvistikasining asosiy tushunchasi sifatida korpusning umumiy tavsifi beriladi, matn korpuslarining tasnifi taqdim etiladi.

Kalit so'zlar: lingvistik korpus, kulturometriya, metodologiya, korpusga yo'naltirilgan yondashuv, korpusga asoslangan yondashuv.

Введение

С развитием информационных технологий в настоящее время всё больше информации, которая используется нами ежедневно, существует в электронном формате. Данные из традиционных печатных источников

переносятся в цифровое пространство, поскольку такая форма организации информации позволяет не только обеспечить ее долговременное хранение, но и дает возможность работать с ней на совершенно другом уровне. Одной из подобных форм организации данных является

языковой корпус, который по определению В. П. Захарова представляет собой «многогранное собрание естественных случаев употребления языка в виде текстов разной жанровой и стилистической направленности и хранящееся в электронном формате».

Основная часть.

Прежде, чем говорить о корпусной лингвистике, необходимо определить само понятие лингвистического корпуса. По-английски это будет *linguistic corpus* или *text corpus*, множественное число *linguistic corpora* (*corpora* употребляется реже). Существует довольно много определений, которые сходятся в одном: корпус есть «некоторый филологический объект» [1].

Корпусная лингвистика – это лингвистика корпусов, то есть собраний текстов. Для начала такого «определения» вполне достаточно. Такое направление лингвистики существует чуть более полувека, а в России это, по сути, наука XXI века: ее активное развитие пришлось на самое начало третьего тысячелетия.

В настоящее время существует множество определений понятия «корпус». Например, определение, приведенное в учебнике Э.Финегана, гласит: *корпус* — репрезентативное собрание текстов, обычно в машиночитаемом формате, включающее информацию о ситуации, в которой текст был произведен, такую как информация о говорящем, авторе, адресате или аудитории [2].

Т.Мак-Энери и Э.Вилсон дают следующее определение: *корпус* — это собрание языковых фрагментов, отобранных в соответствии с четкими языковыми критериями для использования в качестве модели языка [3].

Сегодня корпусная лингвистика часто понимается как новая лингвистическая дисциплина, которая связана с изучением использования языка в реальной жизни с помощью компьютеров и электронных корпусов. Корпусная лингвистика имеет по крайней мере две черты, дающие ей основание претендовать на положение самостоятельной дисциплины: 1) характер используемого словесного материала, а именно размеченные тексты; 2) специфика инструментария.

Если такие разделы лингвистики, как синтаксис, семантика и социолингвистика, имеют целью описание или оценку языковой структуры или языкового использования, то корпусная лингвистика является более широким понятием, методологией, которую можно применить ко многим аспектам как языковых исследований, так и не только языковых. Корпусные методы лежат в основе новой дисциплины, которая получила название «культурометрия» (*culturomics*) и распространяется на все области гуманитарных исследований. [4].

Задаваясь вопросом о месте корпусной лингвистики в лингвистике вообще, видимо, правильнее всего будет сказать, что это методология лингвистического исследования, применимая практически к любой области лингвистики. Однако существует и другой взгляд: корпусная лингвистика — это, собственно, и есть настоящая научная лингвистика. В англоязычной литературе эти подходы — корпусная лингвистика как методология лингвистики и как отдельная наука — получили название *corpus-based* (корпусно-ориентированный подход) и *corpus-driven* (корпусно-управляемый подход).

Первый подход предполагает, что корпусы используются для проверки лингвистических теорий или гипотез, чтобы их подкрепить, подтвердить, опровергнуть или уточнить. Второй подход провозглашает, что корпус сам является главным и единственным источником наших теорий о языке, корпусная лингвистика получает здесь статус теории [5] и рассматривается как «важнейший концепт в лингвистической теории» [Stubbs, 1993, p. 24]. Это значит, что корпус неявно содержит в себе теорию языка и нужно ее оттуда только «добыть» [6]. «Теория не существует независимо от данных» [Tognini-Bonelli, 2001, p. 84–85]. Это понимание возвращает нас к работам американских структуралистов первой трети XX в.

Можно также привести высказывание В.А.Плунгяна (лекция в Европейском университете в Санкт-Петербурге) о том, что если раньше лингвистика стояла на двух «китах»

— лексике и грамматике, то теперь к ним добавилась третья ипостась — корпус.

В русском языке терминология корпусной лингвистики пока окончательно не установилась в силу ряда причин: зарождение корпусной лингвистики в США и Великобритании и ее более позднее развитие в России обусловили тот факт, что терминология складывалась и продолжает развиваться в недрах английского языка. Русские термины в основном представляют собой английские заимствования, некоторые из них, но в других значениях, давно существуют в русском языке. Так, русское слово «корпус» стало многозначным задолго до своего появления в качестве термина корпусной лингвистики. Употребление форм этого существительного в лингвистике является проблематичным, поскольку возможны варианты множественного числа «корпусы» и «корпуса». Для значения «массив», которое имеет место в случае языковых корпусов, именительный падеж множественного числа должен быть «кóрпусы», и, соответственно, прилагательное «кóрпусный» должно произноситься с ударением на первом слоге [Большой толковый словарь русского языка, 1998]. В то же время наблюдение над узусом специалистов пока свидетельствует в пользу форм «корпусá», «корпуснóй», «корпуснáя», которые используются часто, так что можно, видимо, с осторожностью сказать, что в настоящее время этот вопрос остается открытым. Правила, регламентирующие употребление той или иной формы применительно к корпусной лингвистике, пока нет, хотя, как представляется, «победить» должен вариант «кóрпусы».

В первой половине 1990-х годов корпусная лингвистика окончательно сформировалась как новое лингвистическое направление. Но правильно ли будет сказать *новое*?

Знаменитая ключевая фраза мольеровского героя из пьесы «Мещанин во дворянстве» звучит так: «...я и не подозревал, что вот уже более сорока лет я говорю прозой». Открытие, сделанное господином Журденом, должно изобличать, конечно, его безграмотность, однако можно сказать, что мы

действительно говорим прозой. Так же можно сказать, что лингвисты всю жизнь занимались корпусной лингвистикой, не подозревая об этом. Не случайно одна из статей В. Н. Фрэнсиса (W. N. Francis) называется «Corpora V. C.» (*before Christ* — до рождества Христова). Здесь, конечно, игра слов: автор имел в виду *корпусы до компьютеров* (*Corpora Before Computers*). В статье речь идет о том, что идеи корпусной лингвистики действительно зародились задолго до компьютерной эры [Francis, 1992]. Лингвисты и лексикографы уже давно в своей работе используют эмпирический материал, цитаты из текстов, которые выписывались на карточки и образовывали «корпусы» под названием «картотеки».

Выводы. Корпусная лингвистика – область лингвистики, в которой создаются, совершенствуются корпусы текстов, а также применяются в качестве инструмента лингвистического исследования.

Под корпусом в современной лингвистике нужно понимать совокупность текстов в электронной форме. Тексты собраны по определенным критериям, в соответствии с конкретной исследовательской задачей, и отражают ту или иную сферу языка [7].

Список использованной литературы:

1. Баранов, А.Н. (2001). Корпусная лингвистика. Введение в прикладную лингвистику. -М., – С.112-137.
2. Information Structure Categories in Discourse Finegan, 2004
3. McEnery, T. & Wilson, A. (2001). Corpus Linguistics (second edition)
4. Захаров, В.П., Богданова, С.Ю. (2020). Корпусная лингвистика: учебник. 3-е изд., перераб. — СПб.: Изд-во С.-Петербур. ун-та, 234 с.
5. Tognini-Bonelli, 2001, p. 1 Corpus Linguistics at Work
6. Sinclair, 2004, p. 191 Trust the text language, corpus and discourse
7. Аблаева Н., Джуманиязова И. Возможности корпусной лингвистики в изучении русского словообразования. Vol. 2 no. 22.04 (2024): «contemporary technologies of computational linguistics - ctcl.2024»
8. <https://znanium.ru/catalog/documentid=373779>