



**СОВЕРШЕНСТВОВАНИЕ
МЕТОДИКИ ОБУЧЕНИЯ ЯЗЫКАМ:
ПЛОЩАДКА ОБМЕНА
ПРОГРЕССИВНОЙ ПРАКТИКОЙ**

КОРПУСНАЯ ЛИНГВИСТИКА: ПРОБЛЕМЫ И ЗАДАЧИ

Шарипов С.С.

*Бухарский государственный университет
sharipov.s.s.1777@gmail.com*

Аннотация. Корпусы письменных и устных текстов в настоящее время успешно применяются в лингвистической педагогике и при обучении иностранному языку. Мировая практика развития корпусной лингвистики и ее применения в обучении доказывает эффективность корпусных методов. Современные технологии не только меняют старые лингвистические инструменты (превращая, например, традиционные словари в компьютерные базы данных), но и создают новые. К таким новым лингвистическим ресурсам относятся корпусы текстов.

Ключевые слова: корпусная лингвистика, компьютерные технологии, лингвистика, язык, корпус, анализ, обработка данных.

Корпусная лингвистика в современном ее понимании зародилась в США и Западной Европе в конце 1960-х гг. С ростом возможностей современных компьютерных технологий, с середины 1980-х гг. стали активно появляться корпусные проекты различных масштабов на разных языках и для разнообразных целей. Первый корпус был создан в 1963 г. в США (*The Brown Standard Corpus of American English*). Объем данного корпуса под авторством У. Н. Френсиса (W. Francis) и Г. Кучеры (H. Kucera) составил 1 млн словоупотреблений (500 текстов по 2 тыс. словоупотреблений каждый). Он превратился в популярный объект исследований и пример для создания аналогичных баз данных. В среде ученых появилось понимание того, что ряд корректных лингвистических исследований возможно провести только на большом речевом материале. Все это стимулировало появление подхода, разрабатывающего правила организации текстов в корпус и методики их анализа, а корпусная лингвистика, таким образом, зарождалась как методология такого подхода.

В Узбекистане на сегодняшний день наиболее распространено определение, которое дал этой дисциплине Б. Менглиева: «Корпусная лингвистика – раздел компьютерной лингвистики, занимающийся разработкой общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий». Однако ввиду недостаточной разработанности предмета изучения вопрос дефиниций в данном случае все еще открыт. Ученые ведут споры относительно направления, к которому стоит относить корпусную лингвистику. Так, В. В. Мамонтова считает вышеприведенное определение сужением понимания этой дисциплины и ограничением ее рамками компьютерной лингвистики, в то время как «под компьютерной лингвистикой обычно понимается широкая область использования компьютерных инструментов - программ, компьютерных технологий организации и обработки данных - для моделирования функционирования языка в тех или иных условиях, ситуациях, проблемных

областях, а также сфера применения компьютерных моделей языка не только в лингвистике, но и в смежных с ней дисциплинах». В. В. Мамонтова полагает, что корпусная лингвистика использует компьютеры именно как инструмент, и без них, конечно же, она не имела бы возможности выполнять свои функции. Однако, по ее мнению, это можно отнести практически к любой отрасли современного знания, что не делает их составными частями *computer science* [Авезов: 43].

Изучив современные условия функционирования корпусной лингвистики, сравнив ее различные инструменты, цели, задачи и результаты исследований, произведенных при помощи корпусов, в терминологическом аспекте мы соглашались с мнением Б. Менглиева, поскольку именно глобальная компьютеризация способствовала зарождению этого явления, а также ввиду того, что без машинного сбора и обработки материала исследование в данном случае не представляется возможным, в то время как даже с развитием технологий многие другие прогрессирующие направления лингвистики не испытывают зависимости от компьютеров. Основным предметом изучаемой нами дисциплины является корпус. Под лингвистическим, или языковым, корпусом текстов Б. Менглиев понимает «большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач». Определение корпуса как феномена - также задача дискуссионного характера.

Наиболее точным нам представляется определение корпуса, данное одним из основателей корпусной лингвистики Джоном Синклером. Он понимал под корпусом «собрание отрывков текстов в электронной форме, отобранных в соответствии с внешними критериями, чтобы наиболее полно представлять язык или вариацию языка. Функционирует как источник данных для лингвистических исследований» [Базарова: 89].

В современных условиях корпус - это мощный и эффективный инструмент научных исследований, в том числе по теории и практике перевода.

Задачей корпуса является предоставление всевозможных справок из разных областей (лексики, грамматики, акцентологии, истории языка и др.) для обучения родному или иностранному языку. Одно из преимуществ корпусного исследования в лексикографии состоит в том, что корпус можно использовать для демонстрации множества контекстов, в которых употребляется то или иное слово. Затем из этих контекстов можно выделить разные смыслы, ассоциируемые со словом. По мнению специалистов, методологический аппарат корпусной лингвистики является перспективным инструментом в теоретическом и практическом обучении иностранному языку. Результаты корпусных поисков (*конкордансов*) в распечатанном виде могут быть легко инкорпорированы в раздаточный материал, методические пособия и т.п. и использованы в процессе традиционного преподавания.

Кроме прямого применения на занятии по иностранному языку, корпус как метод может быть использован для критической оценки существующих

методических материалов. Так, в ходе большинства таких исследований ученые обнаружили, что существуют значительные расхождения между тем, что предписывается учебными пособиями по грамматике английского языка, и тем, как язык действительно используется носителями, о чем свидетельствует корпус разговорного языка.

Однако анализ тематически обусловленной литературы показывает, что пока корпуса письменных и устных текстов применяются в лингвистическом анализе значительно успешнее, чем при обучении иностранному языку и в лингвистической педагогике.

В целом же специалисты сходятся во мнении, что методы корпусной лингвистики должны быть обязательны при разработке и оценке эффективности учебных материалов и методических пособий с тем, чтобы наиболее распространенные употребления получали приоритетное внимание, а периферийные употребления занимали соответствующее им место.

Важно отметить, что корпусная лингвистика как научная методология и отрасль языкознания очень молода и развивается в разных странах неравномерно. Для некоторых языков, таких как английский, немецкий, финский или японский, созданы обширные и репрезентативные аннотированные корпуса, в то время как для других языков, включая и русский, процесс создания полноценных корпусов переживает период становления. Ситуация с мультязычными (параллельными или сравнительными) корпусами еще более сложная, что значительно ограничивает невероятный потенциал корпусных методов в области лингвистических исследований, языковых технологий и преподавании иностранных языков.

Корпусная лингвистика в англоязычных странах начала развиваться гораздо раньше, чем в Узбекистане. Как следствие, относящаяся к ней терминосистема до сих пор испытывает недостаток эквивалентов на русском языке. Современным лингвистам приходится осваивать понятия «конкорданс», «токенизация», «лемматизация», «стемминг», «парсинг», «тэггер» и многие другие [Нигматова: 123].

Мнения специалистов разнятся даже относительно русскоязычной нормы множественного числа отправной лексемы «корпус»: часть лингвистов, наряду с общепринятой нормой «корпусы»^б, вводит в узус единицу «корпуса», что у нас вызывает лишь стойкую ассоциацию с распространенной современной просторечной ошибкой «полис» - «полюса».

Тем не менее в последние десятилетия обширные корпуса текстов, использованные исследователями методик преподавания иностранных языков для оценки реалий языка в его естественном состоянии, в значительной степени повлияли на повышение качественного уровня выпускаемых языковых пособий. Особого упоминания заслуживают новые словари, создаваемые с использованием методик корпусной лингвистики, такие как *Longman*, *Oxford*, *Collins*, начиная с опыта критического переосмысления постулатов описательной грамматики английского языка. Одним из первых в этой серии

стало издание *Longman Grammar of Spoken and Written English*, опубликованное в 2000 г.

По мнению Т. Н. Сергеевой, основные задачи корпусной лингвистики могут быть сведены к следующим: 1) разработка теоретических оснований данного направления; 2) анализ опыта создания и применения корпусов различных видов; 3) формулирование общих требований к корпусу; 4) создание корпусов для различных исследовательских и учебных задач; 5) формирование эффективных способов применения корпусов текстов в различных областях языкознания [Sharipov: 61].

Как нам видится, именно последняя задача представляет практическую ценность для наиболее широких слоев исследователей и преподавателей иностранного языка.

Корпусные методы зарекомендовали себя также и в мировой практике преподавания языка профессионального общения, являясь высокоэффективным инновационным дополнением к традиционным образовательным технологиям. Эти методы сочетают в себе такие аспекты, как междисциплинарность, эмпирическая адекватность, аутентичность, гибкость и адаптация к конкретным задачам и целевым группам, возможность самостоятельной работы студента.

Поиск по корпусу может выдать:

1) все употребления выбранного слова в непосредственном контексте, на основании чего студент сможет определиться с выбором точного эквивалента при переводе;

2) управление глагола, существительного и т.д. (слова, которые чаще всего стоят рядом с выбранным словом);

3) интертекстуальность: значение слова как сумма его употреблений;

4) развитие концептов во времени.

Многие лингвисты используют корпус как «банк примеров», т.е. пытаются найти. Подход корпусной лингвистики обеспечивает репрезентативность и сбалансированность языкового материала, а также поисковый инструмент, который обычно дает возможность хорошей выборки в данном корпусе.

С течением времени доступ к языковым корпусам упростился. Еще 10 лет назад, как правило, требовались специальные программы, серверы, CD-диски. На сегодняшний день данные ресурсы доступны для бесплатного скачивания или в режиме онлайн.

По мнению З. К. Гаджиевой, в пользу этого метода говорят следующие определяющие его надежность и достоверность факторы:

- корпусный метод принято считать эмпирическим, его главной целью является анализ реальных словоупотреблений в естественной языковой среде;
- использует достаточно большую, репрезентативную подборку текстов;
- является целевым, т.е. должен быть ориентирован на реальное приложение и результаты.

В больших корпусах возникает проблема, которая становится все более актуальной: поиск по запросу может выдавать сотни и даже тысячи результатов

(контекстов употребления), которые физически невозможно просмотреть в ограниченный отрезок времени. Для решения этой проблемы разрабатываются системы, позволяющие группировать результаты поиска и автоматически разбивать их на подмножества (кластеризация результатов поиска) либо выдающие наиболее устойчивые словосочетания со статистической оценкой их значимости. Проблемой в данной области является то, что осмысление теоретических оснований нового направления в определенной мере отстает от конкретных исследований с применением корпусов. Несмотря на то, что проблемы корпусной лингвистики рассматриваются в учебных пособиях, им посвящены специальные выпуски научных журналов, в которых публикуются статьи по общим и специальным проблемам создания и функционирования корпусов текстов, ряд задач остается нерешенным. Это определение корпусной лингвистики и основных понятий, ее места в системе языкознания, поиск и ассимиляция русскоязычных эквивалентов для обозначения ее методов и др. Кроме того, применение корпусов в виде исследовательского инструментария в Узбекистане только начинает свое становление в качестве предмета языковедческой рефлексии.

Однако все же можно прийти к выводу, что, хотя в настоящее время возможности методов корпусной лингвистики в Узбекистане пока не находят должной реализации в обучении родному и иностранному языку, анализ корпусов текстов, методик и наработок корпусной лингвистики является достаточно перспективным направлением в области преподавания иностранных языков не только студентам лингвистических специальностей, но и студентам неязыковых вузов. С развитием вычислительной техники, способной хранить и обрабатывать огромное количество лингвистических данных, и с развитием корпусной лингвистики стало возможным основывать лингвистическое суждение на чем-то гораздо более надежном и разнообразном, чем личный опыт или интуиция отдельно взятого автора учебного пособия по иностранному языку.

Литература

1. Аvezов С. О корпусной лингвистике, трудностях перевода и принципах организации параллельных корпусов текстов //«Узбекские национальные образовательные здания теоретическое и практическое создание вопросы" Международная научно-практическая конференция. – 2022. – Т. 1. – №. 1.

2. Базарова Б. Б. Корпусная лингвистика и преподавание иностранных языков // Вестник Бурятского государственного университета. — 2015 — Вып. 15. — С. 88—92.

3. Нигматова Л. Параллельные корпусы, как эффективный метод обучения иностранного языка //«Узбекские национальные образовательные здания теоретическое и практическое создание вопросы" Международная научно-практическая конференция. – 2022. – Т. 1. – №. 1.

4. Sharipov S. S. " LEXICOGRAPHY" AND DICTIONARY COMPILATION //Scientific reports of Bukhara State University. – 2021. – Т. 5. – №. 2. – С. 52-66.