

## ПРИКЛАДНАЯ ЛИНГВИСТИКА: ПРОБЛЕМЫ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ И СОЗДАНИЯ ПЛАТФОРМ ТЮРКСКИХ ЯЗЫКОВ

Б.Элов\* ,

М.Абжалова\*\* ,

原敏(Yuan Min)\*\*\*

**Аннотация.** В статье рассматриваются проблемы прикладной лингвистики и лингвистических корпусов, наряду с созданием платформ тюркских языков. Затронуты вехи истории компьютерной и корпусной лингвистики, общие принципы построения корпусов лингвистических данных (корпусов текстов) с использованием современных компьютерных технологий, методология сбора реальных языковых явлений письменных и устных текстов, а также методы их хранения и анализа.

**Ключевые слова:** *прикладная лингвистика, лингвистический корпус, платформа тюркских языков.*

**Annotation.** The article examines the problems of applied linguistics and linguistic corpuses, along with the creation of platforms for the Turkic languages. The milestones in the history of computer and corpus linguistics, the general principles of constructing corpuses of linguistic data (text corpuses) using modern computer technologies, the methodology for collecting real linguistic phenomena of written and oral texts, as well as methods of their storage and analysis are also highlighted.

**Key words:** *applied linguistics, linguistic corpus, platform of Turkic languages.*

Термин «прикладная лингвистика» получил широкое распространение в 1950-х годах в связи с появлением первых компьютерных систем для автоматической обработки текстовой информации (машинный перевод, автоматическая абстракция и т. д.); поэтому в русскоязычной литературе вместо термина «прикладная лингвистика» часто используются термины «компьютерная лингвистика», «автоматическая лингвистика», «инженерная лингвистика», что не полностью оправдывает себя, поскольку каждая из

\* доцент Ташкентского государственного университета узбекского языка и литературы, кандидат технических наук, Узбекистан

\*\* Ташкентского государственного университета узбекского языка и литературы, кандидат филологических наук, Узбекистан.



перечисленных дисциплин имеет свои собственные сферы применения и методы в области прикладной лингвистики.<sup>20</sup>

Хотя появление прикладной лингвистики как самостоятельной научной науки восходит к относительно недавнему прошлому (примерно 20-е годы прошлого века), практические проблемы в лингвистике существовали с самого начала. Эти проблемы в конечном итоге привели к оптимизации языковых функций. В частности, она служила для формирования и сохранения традиции чтения и понимания текстов (экзегетика и герменевтика) или для обеспечения общения между разными народами. В философской и лингвистической теории стало обычным различать коммуникативные, эстетические и когнитивные функции языка, которые, в свою очередь, подразделяются на более мелкие функции.

С функциональной точки зрения практическую лингвистику можно определить как учебную дисциплину, в которой целенаправленно изучаются и развиваются методы оптимизации различных областей языковой системы. Языковые функции определяют отправные точки для классификации большой области применения лингвистических знаний. Такие предметы, как теория перевода, машинный перевод, теория и практика преподавания родного и иностранных языков, теория и практика информационных систем, создание информации и, в более широком смысле, искусственные языки, теория кодирования помогают оптимизировать коммуникативную функцию. Исследования в области социолингвистики (особенно изучения и обоснования языковой политики), орфографии и орфоэпии, теории влияния и политической лингвистики важны для оптимизации социальной функции языка. Исследования в области лексикографии (включая компьютерные), терминологии, корпусной и полевой лингвистики помогают оптимизировать эпистемологическую функцию.

Корпусная лингвистика - это независимый раздел компьютерной лингвистики; занимается разработкой общих принципов построения и использования лингвистического корпуса (корпуса текстов) с использованием компьютерных технологий. Под лингвистическим корпусом или корпусом текста понимается крупномасштабный машиночитаемый, комбинированный, помеченный, отформатированный филологически обоснованный набор языковых данных, предназначенный для решения конкретных языковых проблем.



Современные условия языковой деятельности включают не только изучение речевых образов, но и отдельных хронологических случаев, систематизацию речевой практики. Системный подход к обучению и обобщению определяет речевую практику как уникальную единицу. Традиционно только постоянные субъективные факторы позволяют без влияния анализировать эмпирический материал. Конечно, охват речевой практики считается реалистичным из-за объективных критериев, которые применяются сегодня. В то же время через формат «лингво-информация» универсально-статистические и специальные дисциплины открывают новые возможности в выборе языковых форматов. Возникнув в 60-е годы, корпусная лингвистика создала новую методологическую основу речевой практики, исследовательских приоритетов. Корпусная лингвистика как металингвистический инструмент позволяет конструировать, использовать языковую методологию и лингвистическую интерпретацию речевых единиц - при решении компьютерных задач «старого» и «нового» с использованием компьютерных средств.

Корпусная лингвистика - приоритетное направление современной прикладной лингвистики. Корпусная лингвистика определяет общие принципы построения корпусов лингвистических данных (корпусов текстов) с использованием современных компьютерных технологий, разрабатывает методологию сбора реальных языковых явлений письменных и устных текстов, а также методы их хранения и анализа. Работа с текстовыми корпорациями позволяет исследователю в определенной степени абстрагироваться от субъективности и объективно изучать язык.

В парадигме современного языкознания лингвистика отличается методологической универсальностью и эффективностью. Сегодня корпусная лингвистика часто понимается в лингвистике как новый подход к обучению использованию языка в «реальной жизни» с использованием компьютера и электронного корпуса. Корпусная лингвистика имеет как минимум две особенности, чтобы претендовать на независимую позицию:<sup>21</sup>

- 1) характер использованного устного материала;
- 2) особенности средств.

В конце девятнадцатого и начале двадцатого веков корпус начал создаваться для лингвистических исследований или, в целом, для решения практических задач (например, расчета частотности языковых единиц). С

изобретением и широким использованием компьютеров для корпусной лингвистики начался новый этап развития - созданные корпорации отличаются от существующих не только форматом хранения, но и размером. Национальный корпус - это сокровище национального языка. Он широко используется лингвистами, лексикографами, компьютерными лингвистами, программистами, редакторами, переводчиками, журналистами, издателями, учеными, учителями, учащимися и любыми другими специалистами в данной области.

Лингвистический корпус - это набор текстов, которые объединены по определенным характеристикам (язык, жанр, текст и т. Д.) И снабжены поисковой системой. Лингвистический корпус может также включать письменные тексты (тексты газет, журналов, литературные произведения), транскрипции для радио и телевидения. Организация корпуса может быть самой разной. В зависимости от цели создания корпус может содержать тексты (конкретный язык, литературные жанры, написанные в определенный исторический период).

Национальный лингвистический корпус - это язык социальных возможностей, который представлен в различных жанрах, стилях, региональных и письменных текстах, и его крайне интересно изучать. Национальный лингвистический корпус предназначен для чтения. Поскольку основная цель в национальном корпусе - показать живой язык, тексты разных стилей и жанров используются как единица корпуса, преобладает репрезентативность (жанровое разнообразие текста).

Поскольку корпус - это зеркало, возможность, отражающая сущность каждого языка, создание и практическое использование национально-образовательного корпуса узбекского языка является актуальной задачей на сегодняшний день.

Основными характеристиками текста являются: 1) размер основного текста; 2) представительность дела; 3) разметка .

Размер корпуса может меняться со временем, но эти изменения не должны влиять на репрезентативность корпуса, либо ее рациональную составляющую.

Репрезентативность - это его отдельные части (разные характеристики - время, жанры, стили, авторы и т. д.).

Сами тексты отсутствуют для решения различных лингвистических задач. Для этого тексты должны включать в себя ряд дополнительной лингвистической и экстралингвистической информации - разметку. Макет



определяет множество возможностей, которые корпус предлагает исследователям.

Корпусная лингвистика - это отрасль лингвистики, которая занимается разработкой, созданием и использованием корпусного текста. Этот термин был придуман в 1960-х годах для обозначения развития практики создания корпусов и внес свой вклад в развитие компьютерных технологий с 1980-х годов.

Корпус состоит из релевантных текстов и предназначен для адекватного отражения лексико-грамматических явлений, присущих всему объему текстов на соответствующем языке. И размер, и структура корпуса важны для репрезентативности. Размер представления зависит от задачи, так как он определяется тем, сколько примеров можно найти для изучаемых событий.

Многие тексты, доступные в Интернете, можно использовать в качестве корпуса (т. е. использовать миллиард слов для основных мировых языков). Самый распространенный способ работы лингвистов с интернетом - это создавать запросы к поисковой системе и интерпретировать результаты на основе количества найденных страниц или первых возвращенных ссылок. На английском эта методика получила название Googleology, но Яндексология могла бы стать более подходящим названием для русского языка. Следует отметить, что этот подход годен для решения ограниченных задач, потому что инструменты разметки текста, используемые на веб-сайте, не описывают ряд лингвистических особенностей текста (выделение, классы грамматики, границы фраз и т. Д.). Более того, эта ситуация осложняется низкой распространенностью смыслового порядка.<sup>522</sup>

Электронный корпус можно использовать как богатую сокровищницу для успешного решения ряда задач, обучения языку, а также сбора различных документов и информации. Используя функциональные возможности созданного корпуса, программные ресурсы корпуса текста и особенности текстов, входящих в корпус, он может быть использован при создании тюркоязычных платформ.

В рамках проекта будет рассмотрено изучение любого пласта языка, в том числе создание многотомных словарей, исследование грамматики. Лингвист работает с широким спектром текстов. Для этого нужны не только интеллектуальные силы, но и много времени, очень трудоемкий процесс, который требуется, особенно на подготовительных этапах. Поэтому необходим переход к новым методам сбора материала, его анализу и новым



формам лингвистических источников, что не только значительно повысит производительность труда, но и откроет путь к новым методам и направлениям изучения турецкого языка. Анализ этой проблемы показывает, что создание электронного корпуса и языковой платформы важно для ее решения.

#### **Список использованной литературы:**

1. Барсегян А.А. и др. Анализ данных и процессов: учеб. пособие /. – 3-е изд., перераб.и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.: ил. + CD-ROM – (Учебная литература для вузов)
2. Баркович А.А. Корпусная лингвистика: специфика современных метаописаний языка // Вестник Томского государственного университета. 2016. № 406. С. 5–13
3. [https://www.krugosvet.ru/enc/gumanitarnye\\_nauki/lingvistika/PRIKLADNAYA\\_LINGVISTIKA.html](https://www.krugosvet.ru/enc/gumanitarnye_nauki/lingvistika/PRIKLADNAYA_LINGVISTIKA.html)
4. <http://lab314.brsu.by/kmp-lite/CL/CL-Lect/LecCL51.htm>