



ANDUJON DAVLAT
CHET TILLARI
INSTITUTI



O'ZBEKISTON RESPUBLIKASI OLIY TALIM,
FAN VA INNOVATSIYALAR
VAZIRLIGI



ANDUJON LUG'ATSHUNOSLIK VA
TARJIMASHUNOSLIK
AMALIY-ILMIY INNOVATSIYALAR MARKAZI



Xalqaro ilmiy-amaliy anjuman materiallari

1

**“O‘ZBEKISTONDA ZAMONAVIY LUG‘ATSHUNOSLIK VA
TARJIMASHUNOSLIKNING DOLZARB MASALALARI
(bir tilli, ko‘p tilli, terminologik, frazeologik hamda
elektron lug‘atlar tuzish masalalari)”**

15-SENTYABR 2023

family problems for years?" she had demanded after parroting the same words his grandmother had hurled at him last night. "You are impossible! It was bad enough that you stayed away for fifteen years without a word, but overthrowing all our plans is too much. We will never live down the ignominy". [3, 66].

Tahdid holatlar kommunikantlar o'rtasidagi emotsional vaziyat va, shuningdek, psixologik xususiyatlarning in'ikosi sifatida yuzaga chiqadi. Albatta, voqea-hodisalar jarayonida yuzaga keladigan konflikt emotsiyalarni jilovlay olmaslik sabab qilib ko'rsatiladi. Mazkur kommunikatsiya jarayonida dominant subyektning raqibiga nisbatan verbal, keyinchalik esa kontaktli vaziyatlarni keltirib chiqarishi aniqlandi.

FOYDALANILGAN ADABIYOTLAR

1. Pavlenko A. Temple University *Pragmatics & Cognition* 10:1/2 John Benjamins Publishing Company. 2002. –P.208-237.

2. Апресян. Ю.Д. Избранные труды. Том II. Интегральное описание языка и системная лексикография. – М.: Языки русской культуры, 1995. – 766.

3. Allison L. Christmas collection. Belgrave House/Regency Reads, 2010. - 268 p.

О ПЕРСПЕКТИВАХ КОРПУСНЫХ ИССЛЕДОВАНИЙ И ПРАКТИКЕ ИХ ПРИМЕНЕНИЯ В ЛИНГВИСТИКЕ

Мусаева Гулистан Косбергеновна
Чирчикский государственный педагогический университет

Аннотация. Анализируются магистральные направления современных корпусных исследований и возможности их применения в современной лингвистике и филологии. Описываются основополагающие положения корпусной лингвистики в рамках отечественного и зарубежного языкознания. Проводится классификация лингвистических корпусов различных типов и обзор их инструментария. Рассматриваются проблемы анализа массива текстовых данных и индивидуальных текстовых корпусов. Проводится статистический анализ объема публикаций в наукометрических базах данных. Делаются выводы о перспективах корпусных методик в современных языковых исследованиях.

Ключевые слова: корпус, корпусная лингвистика, лингвистический корпус, корпусное исследование, корпус текстов, массив текстовых данных, автоматизированная обработка данных.

Аннотация. Замонавий корпус тадқиқотларининг асосий йўналишлари ва уларни замонавий тилшунослик ва филологияда қўллаш имкониятлари таҳлил қилинган. Корпус тилшунослигининг маҳаллий ва хорижий тилшунослик доирасидаги асосий қоидалари тавсифланган. Ҳар хил турдаги лингвистик корпусларни таснифлаш ва уларнинг воситаларини кўриб чиқиш амалга оширилади. Бир қатор матнли маълумотлар ва алоҳида матн корпусларини таҳлил қилиш муаммолари кўриб чиқилади. Ссиентометрик маълумотлар

базаларидаги нашрлар ҳажмининг статистик таҳлили ўтказилади. Замонавий тил тадқиқотларида корпус техникасининг истиқболлари тўғрисида хулосалар чиқарилади.

Калит сўзлар: корпус, корпус лингвистикаси, лингвистик корпус, корпус тадқиқотлари, матнлар корпуси, матнли маълумотлар массиви, маълумотларни автоматлаштирилган қайта ишлаш.

On the prospects of corpus research and the practice of its application in linguistics

Abstract. The study aims to analyze main fields of modern corpus research and their application in modern linguistics and philology. The fundamentals of corpus linguistics in the framework of domestic and foreign linguistics are described. The study provides classification of linguistic corpora of various types and a review of their tools. The problems of analyzing an array of text data and individual text corpora are considered. The study carries out statistical analysis of the volume of publications in scientometric databases. The prospects of corpus methods in modern language studies have been summarized.

Keywords: corpus, corpus linguistics, linguistic corpus, corpus research, corpus of texts, array of text data, automated data processing.

Введение. Зарождение интереса к исследованию больших объемов текстовой информации относят к середине XX века, когда впервые в истории начали появляться технические средства и возможности для создания автоматизированных систем обработки текстовых данных. «Рост объема информации в современном информационном обществе затрудняет обработку, хранение, распространение, поиск информации, то есть результативную работу с ней. Данную проблему можно решить с помощью перевода информации в электронную форму и использования информационно-коммуникационных технологий» [1: 133]. Развитие компьютерных технологий и возрастание вычислительных мощностей дало толчок в развитии направления языковых исследований, получившему название корпусная лингвистика [2]. Термин введен в широкое употребление в 1960-е годы, что явилось результатом создания массивов текстовых данных, содержащих многочисленные примеры актуальных языковых контекстов. В основе «фундаментального множества информационных технологий лежит коммуникация, неразрывно связанная с языком и, следовательно, лингвистикой» [3: 102]. Считается, что первым языковым корпусом в прямом смысле этого слова стал созданный в 1961 г. в университете Брауна (США) корпус английского языка, содержащий 500 фрагментов текстов по 2 тысячи слов в каждом. Позднее общепризнанным международным стандартом представительности стал объем языкового корпуса равный 1 млн. словоупотреблений. В.В. Рыков дает следующее определение корпуса: «некоторое собрание текстов, в основе которого лежит логический замысел, логическая идея, объединяющая эти тексты и воплощенная в правилах организации текстов в корпус, алгоритме и программе анализа корпуса текстов, сопряженной с этим идеологии и методологии» [4]. Языковой

корпус, созданный в университете Брауна, стал моделью и прообразом других национальных корпусов, создание которых продолжается по сей день. Актуальность исследования заключается в том, что современные технические средства дают возможность создания индивидуально-авторских текстовых корпусов, а также автоматизированной системы для их обработки, позволяющей мгновенно получать разнообразные данные о тексте и его содержании, а также представлять их в удобном для дальнейшего использования виде. Материалы и методы. В исследовании использовались такие методы, как лингвистическое описание, дефиниционный метод, метод систематизации и классификации материала, наблюдение, а также, метод статистической обработки данных и количественный сравнительный анализ. В качестве материала исследования были использованы данные из наукометрических баз данных, таких как Elibrary (РИНЦ и Science index), WoS, Scopus, Google Scholar, а также данные национальных языковых корпусов, таких как BNC, COCA, Национальный корпус русского языка. Результаты. Одним из самых популярных национальных корпусов является BNC (British National Corpus) [5]. Данный корпус содержит наиболее широко используемые онлайн-базы данных, которые применяются для самых разных целей преподавателями и исследователями в университетах по всему миру [6]. Кроме того, данные корпуса (например, полнотекстовые, частотность слов) использовались широким кругом компаний в самых разных областях, особенно в области технологий и изучения языков. Британский национальный корпус (BNC) был первоначально создан издательством Oxford University press в 1980-х – начале 1990-х годов и содержит 100 миллионов слов текста из широкого спектра жанров (например, разговорный, художественная литература, журналы, газеты и академические издания).

Еще одним примером весьма представительного корпуса можно считать корпус современного американского английского языка (Corpus of Contemporary American English) [7]. Он является крупным и репрезентативным корпусом американского английского языка, а также наиболее широко используемым корпусом английского языка, и связан со многими другими корпусами английского языка. Эти корпуса ранее были известны как «Корпуса BYU», и они дают беспрецедентное представление о вариациях английского языка. Корпус содержит более миллиарда слов текста (более 25 миллионов слов в год с 1990 по 2019 год) из восьми жанров:

- устная речь;
- художественная литература;
- популярные журналы;
- газеты;
- академические тексты;
- субтитры к телевидению и фильмам;
- блоги; – другие веб-страницы.

В качестве примера русскоязычного корпуса можно привести Национальный корпус русского языка [8], который позиционируется как представительная

коллекция текстов на русском языке общим объемом более 2 млрд слов, оснащенная лингвистической разметкой и инструментами поиска. Он создан по образцу англоязычных корпусов и содержит следующие разделы .

ЛИТЕРАТУРА

1. Хурум Р.Ю. Использование информационно-коммуникационных технологий в лингвистике как фактор формирования информационных компетенции // Вестник Адыгейского государственного университета. Сер.: Филология и искусствоведение. Майкоп, 2017. Вып. 4. С. 133-138.
2. Баранов А.Н. Проблема репрезентативности корпуса данных (на примере политической метафорики) // Труды Международного семинара «Диалог-2001» по компьютерной лингвистике и ее приложениям. Аксаково, 2001.
3. Птущенко Е.Б. Перспективные информационные технологии как инструмент познания в лингвистике // Вестник Адыгейского государственного университета. Сер.: Филология и искусствоведение. Майкоп, 2017. Вып. 2. С. 102-108.

GLOBALLASHUV JARAYONIDA AXBOROT TEXNOLOGIYALARI VA KOMPYUTER LINGVISTIKASI FANLARINING INTEGRALLASHUVI

Toxta Ametovna Yuldasheva,
Chirchiq davlat pedagogika universiteti

Annotatsiya: maqolada globallashuv jarayonida axborot texnologiyalari va kompyuter lingvistikasi fanlarining integrallashuvi haqida gap borgan.

Kalit soʻzlar: kompyuter lingvistikasi, filologiya, tillarni oʻrganish metodikasi.

Аннотация: в статье говорится об интеграции информационных технологий и компьютерной лингвистики в процессе глобализации.

Ключевые слова: компьютерная лингвистика, филология, методика изучения языка.

Abstract: the article talks about the integration of information technologies and computer linguistics in the process of globalization.

Keywords: computer linguistics, philology, language learning methodology.

Uchinchi mingyillikning konseptual voqeligi-globallashuv, yuqori surʼatlarda axborot almashinuvi, fan texnika taraqqiyoti, maʼnaviy-axloqiy qadriyatlarni qayta koʻrib chiqishga intilish, siyosiy va ekologik xatarlarbularning barchasi inson, jamiyat hamda taʼlim oldiga yangi vazifalarni qoʻydi.

Jamiyat rivojining hozirgi bosqichida ijodiy, nostandart fikrlaydigan, fan va ishlab chiqarishda yangi usullarni topa oladigan professional mutaxassislariga ehtiyoj tobora ortmoqda.

Taʼlimning sifati unda faoliyat olib borayotgan professoroʻqituvchining kasbiy