THEME: NATIONAL CORPORA AND THEIR SIGNIFICANCE IN LINGUISTICS

Senior teacher of JSPU

Abdullajonova Hakima

Students: **Farmonova Marjona,**

Ziyodullayeva Mohinur,

Ruzmanova Muxlisabonu,

Qorabekova Shahzoda

Jizzakh State Pedagogical University

Foreign Languages

Annotation: *This article explores the concept of national corpora and their importance in linguistic research. National corpora are large, structured collections of authentic language data that represent a particular language as used by its native speakers. They serve as valuable tools for studying vocabulary, grammar, discourse patterns, and language variation. The paper examines the development, functions, and applications of national corpora, emphasizing their role in promoting linguistic standardization, supporting lexicography, and fostering comparative studies. Examples such as the British National Corpus (BNC) and the Russian National Corpus (RNC) demonstrate how national corpora contribute to the understanding and preservation of linguistic identity.*

Key words: *National corpus, linguistic research, lexicography, standardization, language identity, corpus linguistics, representativeness*

Annotatsiya: *Ushbu maqolada milliy korpuslar va ularning lingvistik tadqiqotlardagi ahamiyati yoritilgan. Milliy korpuslar bu tilning ona tilida soʻzlashuvchilar tomonidan yaratilgan tabiiy matnlardan tashkil topgan yirik, tizimlashtirilgan maʼlumotlar bazasidir. Ular lugʻat boyligi, grammatik tuzilma, nutq uslubi va til oʻzgarishlarini oʻrganishda muhim vosita hisoblanadi. Maqolada milliy korpuslarning shakllanishi, vazifalari va qoʻllanilish sohalari tahlil qilinib,*

ularning til me'yorlashuvi, leksikografiya va taqqoslama tadqiqotlardagi roli yoritiladi. Britaniya milliy korpusi (BNC) va Rossiya milliy korpusi (RNC) kabi misollar milliy korpuslarning tilni saqlash va o'rganishdagi ahamiyatini ko'rsatadi.

Kalit so'zlar: Milliy korpus, lingvistik tadqiqot, leksikografiya, me'yorlashuv, til o'ziga xosligi, korpus lingvistika, vakillik

Аннотация: В статье рассматриваются понятие национальных корпусов и их значение в лингвистике. Национальные корпусы представляют собой крупные структурированные базы данных подлинных текстов, отражающих язык в его естественном употреблении носителями. Они служат важным инструментом для изучения лексики, грамматики, дискурсивных моделей и языковых вариаций. В работе анализируются этапы развития, функции и области применения национальных корпусов, подчеркивается их роль в стандартизации языка, лексикографии и сравнительных исследованиях. Примеры таких корпусов, как Британский национальный корпус (BNC) и Российский национальный корпус (RNC), демонстрируют вклад национальных корпусов в сохранение и изучение языковой идентичности.

Ключевые слова: Национальный корпус, лингвистика, лексикография, стандартизация, языковая идентичность, корпусная лингвистика

Main Part

National corpora are among the most significant developments in modern linguistics. They represent large-scale digital collections of authentic texts in a specific national language, carefully selected to reflect the linguistic, regional, and stylistic diversity of that language. A national corpus functions as an empirical foundation for language description, analysis, and teaching. Its design and structure embody key theoretical principles of representativeness, balance, and authenticity, which ensure that the corpus mirrors real language use across different social and communicative contexts.

The concept of a national corpus originated in the late twentieth century with the creation of the British National Corpus (BNC)—one of the earliest and most influential corpora of English.

It was designed to represent modern British English in spoken and written forms across genres such as newspapers, fiction, academic writing, and everyday conversation. Following this model, many countries have since developed their own corpora: the American National Corpus (ANC), Russian National Corpus (RNC), German Reference Corpus (DeReKo), and Uzbek National Corpus (UzNC), among others. Each of these serves as a linguistic mirror of its language community.

From a theoretical perspective, national corpora are grounded in the principles of descriptive linguistics rather than prescriptive rules. They allow researchers to investigate how language actually functions rather than how it "should" be used. This empirical approach has transformed the study of grammar, semantics, and pragmatics. For example, by analyzing frequency data, linguists can determine how often certain words or grammatical structures appear in natural contexts, leading to more accurate grammatical descriptions and dictionary entries.

National corpora also play a crucial role in lexicography. Modern dictionaries rely heavily on corpus data to ensure that definitions and usage examples are authentic and up to date. The Oxford English Dictionary and other major reference works have integrated corpus evidence to reflect real-world language patterns. Similarly, in the Uzbek language context, national corpora can help identify the most frequent word combinations, idiomatic expressions, and stylistic tendencies, thereby supporting the development of modern and reliable dictionaries.

Another important function of national corpora lies in language standardization. By providing evidence of actual usage, corpora contribute to shaping language norms, educational materials, and official orthography. They also aid in monitoring language change, showing how vocabulary and syntax evolve under the influence of globalization, technology, and cultural exchange. For instance, corpora reveal the growing influence of English borrowings in many languages, including Uzbek and Russian.

National corpora are also invaluable in comparative linguistics and translation studies. Through parallel or comparable corpora, researchers can analyze similarities and differences between national languages, explore translation equivalences, and study cultural nuances in meaning. This enhances intercultural communication and translation accuracy.

Moreover, national corpora are of great significance for language teaching. Teachers can use them to design materials based on authentic examples rather than textbook constructs. Learners benefit from exposure to real language data, which improves their vocabulary acquisition, grammatical awareness, and pragmatic competence. In the digital era, access to national corpora empowers students to engage in data-driven learning, where they independently explore linguistic patterns and usage.

Finally, the creation of national corpora supports cultural preservation. Language is a core element of national identity, and corpora document its evolution, diversity, and richness. By storing historical, literary, and spoken texts, they safeguard linguistic heritage for future generations.

Conclusion

National corpora are not only linguistic tools but also cultural and intellectual assets. They enable linguists, educators, and translators to study language in a scientific, evidence-based manner. By capturing authentic usage across time and region, national corpora promote linguistic research, support lexicography, and preserve cultural identity. As more languages develop their own national corpora, the field of corpus linguistics continues to advance toward a deeper, more comprehensive understanding of language diversity and change.

REFERENCES (BOOKS)

1. Biber, D., Conrad, S., & Reppen, R. (1998). Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press.

2. Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge University Press.

3. Kennedy, G. (1998). An Introduction to Corpus Linguistics. Longman.

4.   Leech, G. (1992). Corpora and Theories of Linguistic Performance. In J. Svartvik (Ed.), Directions in Corpus Linguistics. Mouton de Gruyter.

5.   McEnery, T., & Hardie, A. (2012). Corpus Linguistics: Method, Theory and Practice. Cambridge University Press.

6.   Sinclair, J. (1991). Corpus, Concordance, Collocation. Oxford University Press.

Websites

1.   https://corpus.byu.edu/

2.   https://www.nationalcorpus.ru/

3.   https://www.sketchengine.eu/

4.   https://corpus.britishcouncil.org/