

e-ISSN 3060-5059

DOI: 10.67227



IJTIMOYIY-GUMANITAR SOHADA ILMIY-INNOVATSION TADQIQOTLAR

ILMIY METODIK JURNAL

Scientific and Innovative Research in
Social and Humanitarian Studies

Vol. 3 • Issue 7

2026



Founded in 2024



publishscience.uz



publish_science_journal@mail.ru



PEER REVIEWED • OPEN ACCESS

ONOMASTIK TAHLILDA KORPUS LINGVISTIKASINING O'RNI

Xaitqulov Zahriddin Xazratqul o'g'li¹

Shaymardanov Abror Hamid o'g'li²

¹O'zbekiston davlat jahon tillari universiteti, dotsent (PhD).

²"Cyber University" davlat universiteti, katta o'qituvchi.

e-mail: khaitkulovzahriddin@gmail.com ORCID: <https://orcid.org/0009-0002-2661-9584>

e-mail: abror.research@gmail.com ORCID: <https://orcid.org/0009-0003-8313-613X>

Annotatsiya.

Korpus lingvistikasi so'nggi o'n yilliklarda tilshunoslik tadqiqotlarining eng samarali metodlaridan biri sifatida e'tirof etilib, real til materiallari va miqdoriy dalillarni taqdim etish imkonini beradi. Korpus lingvistikasining onomastik tadqiqotlarga tatbiq etilishi turli nutq turlarida uchraydigan atoqli otlarning strukturaviy, semantik, pragmatik va funksional xususiyatlarini o'rganish uchun yangi imkoniyatlarni yaratdi. Mazkur tadqiqot yirik elektron korpuslar yordamida antroponimlar, toponimlar, etnonimlar va boshqa turdagi atoqli otlarning qo'llanish chastotasi, tarqalishi, kollokatsion xususiyatlari hamda kontekstga bog'liq qo'llanish holatlarini tahlil qilish imkoniyatlarini ko'rib chiqadi. Korpusga asoslangan yondashuvlar tadqiqotchilarga nomlash tendensiyalari, til variativligi va ma'no o'zgarishlarini turli tarixiy davrlar hamda ijtimoiy-madaniy muhitlar kesimida o'rganish imkonini beradi. Bundan tashqari, korpus vositalari atoqli otlarning takrorlanuvchi qo'llanish modellarini aniqlashga xizmat qiladi hamda leksikografik va lingvistik tahlillar uchun ishonchli ma'lumotlar bazasini taqdim etadi. Tadqiqotda korpus lingvistikasining zamonaviy onomastik izlanishlardagi o'рни, xususan, badiiy adabiyot, ommaviy axborot vositalari va tarixiy hujjatlar tahlilidagi ahamiyatiga alohida e'tibor qaratilgan. Shuningdek, atoqli otlarning kommunikativ va madaniy funksiyalarini yoritishda korpus metodlarining muhimligini ta'kidlagan taniqli olimlarning ilmiy qarashlari tahlil qilinadi. Tadqiqot natijalari korpus lingvistikasi onomastik izlanishlarning aniqligi va obyektivligini oshirish bilan birga, til, madaniyat va identifikatsiya o'rtasidagi munosabatlarni chuqurroq anglashga ham xizmat qilishini ko'rsatadi. Shu sababli korpusga asoslangan onomastik tahlil tilshunoslik va fanlararo tadqiqotlarning istiqbolli yo'nalishlaridan biri hisoblanadi.

Kalit so'zlar: korpus lingvistikasi, onomastika, atoqli otlar, antroponimlar, toponimlar, korpus tahlili, chastotaviy tahlil, kollokatsiya, semantik tahlil, pragmatika, leksikografiya, til variativligi, nomlash modellari, korpusga asoslangan tadqiqotlar.

РОЛЬ КОРПУСНОЙ ЛИНГВИСТИКИ В ОНОМАСТИЧЕСКОМ АНАЛИЗЕ

Хайткулов Захриддин Хазраткул угли¹

Шаймарданов Абдор Хамид угли²

¹Узбекский государственный университет мировых языков, доцент (PhD).

²Государственный университет «Cyber University», старший преподаватель.

e-mail: khaitkulovzahriddin@gmail.com ORCID: <https://orcid.org/0009-0002-2661-9584>

e-mail: abror.research@gmail.com ORCID: <https://orcid.org/0009-0003-8313-613X>

Аннотация.

Корпусная лингвистика (КЛ) в последние десятилетия стала одним из наиболее эффективных методов лингвистических исследований, поскольку предоставляет реальные языковые данные и количественные доказательства. Интеграция корпусной лингвистики в ономастические исследования открыла новые возможности для изучения структурных, семантических, прагматических и функциональных особенностей собственных имён в различных типах дискурса. В данной статье рассматриваются возможности использования крупных электронных корпусов для анализа частотности, распределения, коллокационных моделей и контекстуального употребления антропонимов, топонимов, этнонимов и других видов собственных имён. Корпусно-ориентированные подходы позволяют исследователям изучать тенденции именования, языковую вариативность и семантические изменения в различные

исторические периоды и в разных социокультурных контекстах. Кроме того, корпусные инструменты способствуют выявлению повторяющихся моделей употребления собственных имён и предоставляют надёжную базу данных для лексикографического и лингвистического анализа.

Особое внимание уделяется роли корпусной лингвистики в современных ономастических исследованиях художественной литературы, средств массовой информации и исторических документов. В статье также рассматриваются научные труды известных исследователей, подчёркивающих значимость корпусных методов для выявления коммуникативных и культурных функций собственных имён. Результаты исследования показывают, что корпусная лингвистика не только повышает точность и объективность ономастических исследований, но и способствует более глубокому пониманию взаимосвязи языка, культуры и идентичности. Таким образом, корпусно-ориентированный ономастический анализ представляет собой перспективное направление как для лингвистических, так и для междисциплинарных исследований.

Ключевые слова: корпусная лингвистика, ономастика, собственные имена, антропонимы, топонимы, корпусный анализ, частотный анализ, коллокация, семантический анализ, прагматика, лексикография, языковая вариативность, модели именования, корпусно-ориентированные исследования.

THE ROLE OF CORPUS LINGUISTICS IN ONOMASTIC ANALYSIS

Xaitqulov Zahriddin Xazratqul o'g'li¹

Shaymardanov Abror Hamid o'g'li²

¹Uzbekistan State World Languages University, Associate Professor (PhD).

²Cyber University State University, Senior Lecturer.

e-mail: khaitkulovzahriddin@gmail.com **ORCID:** <https://orcid.org/0009-0002-2661-9584>

e-mail: abror.research@gmail.com **ORCID:** <https://orcid.org/0009-0003-8313-613X>

Abstract.

Corpus linguistics (CL) has been the most effective method of linguistic research in the past few decades; it gives the real language data and quantitative evidence. The integration of corpus linguistics into onomastic studies has opened the door for studying structural, semantic, pragmatic and functional properties of proper names in different kinds of language. The study examines how large electronic corpora can be used to study the frequency, distribution, collocational patterns and context-specific use of anthroponyms, toponyms, ethnonyms and other types of proper names. Corpus-based approaches allow researchers to study naming trends, language variation and semantic changes across different periods and sociocultural contexts. Moreover, corpus tools contribute to identifying the repetitive patterns of proper names and also provide data for lexicographic and linguistic analysis. We focus particularly on corpus linguistics in the context of contemporary onomastic research in literature, media and historical documents. We also review some of the work of famous scholars who point out the importance of corpus linguistics in the context of communicating and cultural aspects of proper names. We find that corpus linguistics not only enhances the accuracy and objectivity of onomastic studies; but it also contributes in a more nuanced way to the discourse about language, culture and identity. Therefore, corpus-based onomastic analysis is an interesting direction for linguistic research and interdisciplinary work.

Keywords: Corpus linguistics, onomastics, proper names, anthroponyms, toponyms, corpus analysis, frequency analysis, collocation, semantic analysis, pragmatics, lexicography, language variation, naming patterns, corpus-based research.

Introduction. Corpus linguistics is one of the most important fields of modern linguistics because it allows **the objective analysis** of linguistic units through large-scale electronic text databases. Onomastics, in turn, is the branch of linguistics that studies proper names, including anthroponyms, toponyms, ethnonyms, hydronyms, and other categories. In recent years, the integration of corpus linguistics and onomastics has enabled researchers to study the frequency of name usage, their semantic properties, pragmatic functions, and historical development on the basis of accurate statistical data [8; 91-p.].

The term *corpus* is used in two main ways in onomastic studies. The more traditional usage refers to the linguistic data used as the basis for an onomastic study. Such data are usually historical and non-computerized and may consist of lists of names or sources such as address books, telephone directories, or name indexes. This is not the type of corpus relevant to the present study. The more recent concept of a corpus, which forms the basis of this article, refers to a

collection of electronic textual data that can be searched using computer software, where proper names (and other linguistic units) occur as parts of larger syntactic constructions [6; 344-p].

Literature review. A number of foreign and Uzbek scholars have contributed to research in the corpus linguistics of onomastic analysis. Among the leading foreign researchers are H. Motschenbacher, *Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names*; R. Coates, *The Pragmatic Theory of Properhood*; W. van Langendonck, *Theory and Typology of Proper Names*; J. M. Anderson, *The Grammar of Names*; C. Hough, *The Oxford Handbook of Names and Naming*; R. Nazar, *Corpus-Based Methods for Recognizing the Gender of Anthroponyms*; M. Sfeir, *A Corpus Linguistic Approach for Onomastics in Drama*; and F. Mitzlaff and G. Stumme, *Onomastics: The Power of Social Co-Occurrences*.

Among Uzbek scholars who have made significant contributions to corpus linguistics and onomastic analysis are N. Ataboyev, *Problematic Issues of Corpus Analysis and Its Shortcomings and Analysis of the Media Text Corpus Through the Prism of Existing Diachronic Corpora of the English Language*; K. Rizayeva, *Uzbek Corpus Linguistics*; S. Abdurahmanova, *The Formation and Development of Uzbek Corpus Linguistics*; M. Samatboyeva, *Automated Recognition of Proper Nouns in NER Objects*; Y. Avlaqulov, *Linguistic Study of Uzbek Proper Names*; and N. Mahmudov, *Uzbek Lexicology and Semantics*, among others.

Methods. Onomastic research has historically relied on a wide variety of methods to shed light on topics such as the history and etymology of names, the meaning and grammar of names, and the social relevance of naming practices. Corpus linguistic methods, however, have only played a minor role in the field [2; 78-p]. This is a disappointing fact, as corpus-based investigations are a very powerful tool for studying and confirming the linguistic features of names in actual language use at the semantic and grammatical levels. The performance focus of corpus linguistics, however, is different from that of other methods of onomastics, which are competence-centred and rely on introspection or reference works.

J. O. Butler and colleagues emphasize that corpus linguistics is an important tool for historical onomastic research. They show that historical corpora allow us to identify orthographic, abbreviated, and other forms of toponyms. This is, in turn, useful for tracing the history of names, improving onomastic databases, and developing automatic named-entity recognition systems. The authors also show that research on regional and historical variants of names using corpora can open up new possibilities for onomastic studies [4; 80-p].

In Germany, H. Motschenbacher is one of the leading scholars who has established a relationship between corpus linguistics and onomastics. He clearly states that corpus linguistics has not been widely applied to onomastic studies and calls for the study of names in a genuine linguistic context. According to him, "Corpus linguistics is, to date, still an underexplored methodology in onomastics" [8; 91-p]. He further argues that the actual functions of proper names in the language system can be identified by examining frequency, concordance, collocation, and keyword analysis [9; 93-p].

Discussions. Proper names have often been compared to common nouns and personal pronouns, with which they share central characteristics [5; 65-p]. For example, proper names, just like common nouns and unlike pronouns, are an open word class that can easily be expanded by adding new members. They can be combined with adjectives and other modifiers. In English, neither common nouns nor proper nouns are case-inflected, unlike pronouns, which show case distinctions (e.g., *I – me, he – him*). In grammatical terms, both common nouns and proper nouns have a grammatical gender value, and pronouns often agree with controller nouns; thus, they also show different grammatically gendered forms.

As with pronouns, proper names are inherently definite. Common nouns, on the other hand, have to take a definite article to be definite (*she – Mary – the woman*). Proper names and pronouns are referential, but not descriptive. Both are not associated with determiners, restrictive adjectives, or restrictive modifiers such as relative clauses and prepositional phrases.

In onomastic and linguistic analysis, **proper nouns** and **proper names** are often used interchangeably, although they are different from one another. A simple understanding of this distinction is extremely important for the proper classification of naming expressions and their role in language. A proper noun is a grammatical category of nouns used to refer to specific people, places, organizations, countries, rivers, mountains, and other unique referents. Proper nouns are orthographic in nature and, in English, are always written with an initial capital letter and are usually written as one word. For example, *Mary, John, London, Paris, Uzbekistan, and Amazon*. Like common nouns that refer to a class of entities, proper nouns refer to singular and unique entities.

The concept of a proper name, however, is much more complicated. A proper name is not limited to one word but can be any linguistic expression that is meant to distinguish a particular entity. Thus, proper names may consist of one word, multiple words, noun phrases, or even descriptive expressions. From this point of view, proper nouns are only a subtype of proper names.

J. Anderson, R. Coates, and W. van Langendonck stress that proper names should be viewed as a semantic and functional category rather than simply a grammatical one. A proper name can take a variety of structural forms: simple proper nouns (*London, Mary, Shakespeare, Oxford*); combinations of proper nouns (*London Heathrow, New York City, Oxford University*); modified proper nouns (*Greater London, South Korea, Northern Ireland*); combinations of common and proper nouns (*London Bridge, Oxford Street, Lake Michigan*); noun phrases containing proper nouns (*Tower of London, University of Oxford, Bank of England*); and descriptive noun phrases functioning as proper names (*Tower Bridge, The White House, The British Museum*) [2; 5; 7; 67–127; 65-p].

A more recent corpus linguistic study investigating the grammar of English country names is that of Motschenbacher (2020). The study uses the Corpus of Contemporary American English (COCA) to shed light on the grammatical behavior of English country names, more specifically on their co-occurrence with the definite article. It is shown that virtually all country names occur with a definite article to some extent and that certain grammatical

constructions require the use of a definite article with country names [10; 99-p].

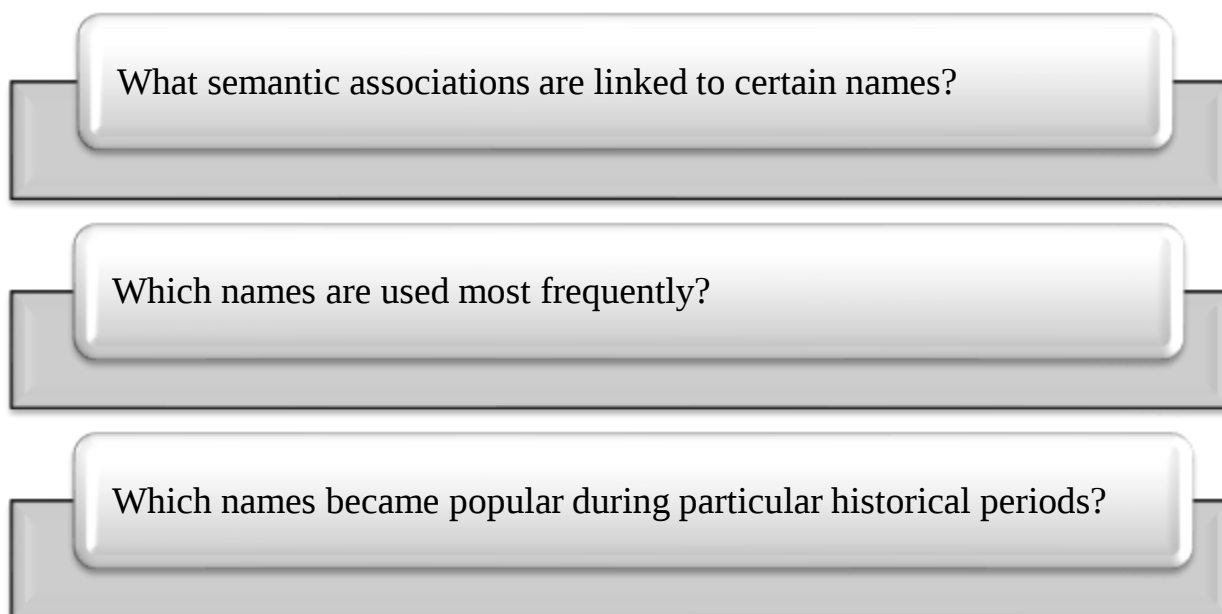
From an onomastic point of view, all of the examples above are proper names since they identify unique entities. However, not all of them are proper nouns in the grammatical sense. For instance, *Tower Bridge* is a proper name even though it consists of the common nouns *tower* and *bridge*. *The White House* is also a proper name even though it is a descriptive noun phrase.

This distinction is particularly important in corpus-based onomastic research. Traditional linguistic approaches typically focus on individual proper nouns, whereas corpus linguistics studies the entire structure of proper names, including their collocations, semantic associations, and contexts. In large corpora such as COCA, BNC, and COHA, proper names are found in various structural patterns, demonstrating that naming practices involve more than simply single-word proper nouns.

Thus, proper nouns are a grammatical category, whereas proper names are a broader onomastic category encompassing all expressions used to identify distinct entities. This distinction allows researchers to conduct more accurate studies of naming systems and gain deeper insights into language, culture, and identity [1; 69-p].

R. Coates' studies help answer the following questions through corpus-based analysis [5; 127-p]:

1. **diagram.** R. Coates' through corpus-based analysis.



The following examples illustrate this approach. In corpora such as **COCA** (Corpus of Contemporary American English) and the **BNC** (British National Corpus), the following collocations can be observed:

- **London** – *capital, city, government*
- **Shakespeare** – *literature, drama, author*
- **Harry Potter** – *magic, wizard, Hogwarts*

These collocations help identify the semantic fields and associative meanings of proper names.

Among the scholars who have examined the grammatical status of proper nouns, the Scottish linguist John M. Anderson occupies a prominent place. He argues that proper names possess syntactic characteristics that distinguish them from common nouns. Corpus-based research makes it possible to identify and verify such grammatical patterns. For example, corpora can be used to examine proper names that typically occur with the definite article:



2. **diagram.** Names occur with the definite article

However, the following proper names are usually used without an article:



3. diagram. Names occur without an article

With thousands of real examples, corpus studies provide statistical evidence to confirm these grammatical regularities.

W. van Langendonck is probably the most influential Belgian thinker on the theory of proper names in the world. He studies the referential function of names and language communication. Through corpus-based analysis, one can then investigate the following types of proper names: Anthroponyms, *John. William. Elizabeth. Michael.* Toponyms, *London. New York. Berlin.* Ergonyms, *Google. Microsoft. Apple.* Corpus tools can be used to find out the frequency of these names in texts and to analyze the semantic environments in which they occur.

Results. In 2021, R. Nazar and his colleagues developed methods for the automatic identification of the gender of personal names using large linguistic corpora. Their study examined the grammatical features surrounding personal names and proposed algorithms for identifying gender markers based on contextual information.

For example, in the following corpus sentences:

- *Mary said she was happy.*
- *Elizabeth gave her speech.*

the pronouns *she* and *her* function as female gender markers, indicating that the personal names refer to female individuals.

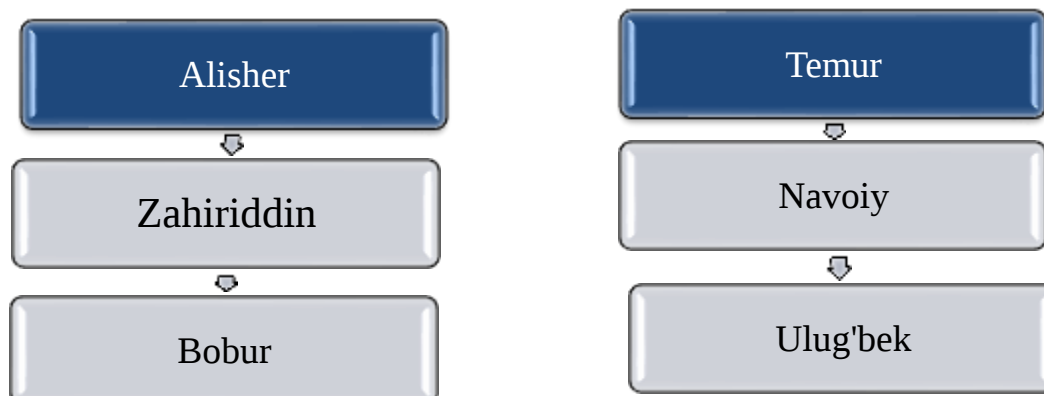
Similarly, in the following sentences:

- *John said he was ready.*
- *Michael completed his work.*

the pronouns *he* and *his* function as male gender markers, indicating that the personal names refer to male individuals.

These examples demonstrate how corpus-based methods can be used for the automatic identification and analysis of personal names in onomastic research.

Uzbek scholar M. Samatboyeva has conducted research on the automatic identification of anthroponyms using Named Entity Recognition (NER) technology. According to M. Samatboyeva, "Anthroponyms constitute the largest category of proper nouns." The following examples illustrate anthroponyms that can be automatically identified in a corpus using NER techniques:



4 diagram. Anthroponyms are used via corpus in using NER techniques

Such corpus-based approaches facilitate the rapid identification, classification, and analysis of proper names in large text collections.

In his research on the linguistic study of the Uzbek onomastic system, Y. Avlaqulov provides a comprehensive classification of onomastic units. According to the scholar, the main categories of onomastic units are:

- anthroponyms (personal names);
- toponyms (place names);
- ethnonyms (names of ethnic groups);
- zoonyms (names of animals);
- theonyms (names of deities or gods);
- phytonyms (names of plants);
- ideonyms (names of intellectual, cultural, or artistic works);
- hydronyms (names of water bodies).

This classification provides a systematic framework for understanding the structure and diversity of proper names in the Uzbek language and its onomastic system.

Conclusion. The above-mentioned studies by foreign and Uzbek researchers demonstrate that corpus linguistics significantly enhances the reliability and accuracy of onomastic analysis. The examples discussed above further support this conclusion. In addition, corpus linguistics provides extensive opportunities for the analysis of proper names, including anthroponyms and toponyms. Specifically, corpus-based methods make it possible to determine the frequency of proper names and their collocations, trace historical changes in their usage, identify their semantic and pragmatic functions, and automatically detect onomastic units using Named Entity Recognition (NER) systems. Therefore, corpus linguistics represents one of the most promising and modern methodological approaches to the study of onomastics in English and Uzbek. Corpus-based methods enable researchers to achieve a detailed, objective, and data-driven understanding of the structure, usage, and development of proper names across different linguistic and cultural contexts.

REFERENCES

1. Allerton D. J. The Linguistic and Sociolinguistic Status of Proper Names: What Are They, and Who Do They Belong To? // *Journal of Pragmatics*. – 1987. – Vol. 11. – P. 61–92.
2. Anderson J. M. *The Grammar of Names*. – Oxford: Oxford University Press, 2007. – P. 67.
3. Biber D., Reppen R. (Eds.). *The Cambridge Handbook of English Corpus Linguistics*. – Cambridge: Cambridge University Press, 2015.
4. Butler J. O., Donaldson C. E., Taylor J. E., Gregory I. N. Alts, Abbreviations, and AKAs: Historical Onomastic Variation and Automated Named Entity Recognition // *Journal of Map & Geography Libraries*. – 2017. – Vol. 13. – No. 1. – P. 80.
5. Coates R. A New Theory of Properhood // *Proceedings of the 21st International Congress of Onomastic Sciences* (Uppsala, 19–24 August 2002) / ed. by E. Brylla, M. Wahlberg. – Vol. 1. – Uppsala, 2005. – P. 125–138.
6. Greule A. Methoden und Probleme der corpusgebundenen Namenforschung [Methods and Problems of Corpus-Based Name Research] // *Name Studies: An International Handbook of Onomastics* / ed. by E. Eichler, G. Hilty, H. Löffler, H. Steger, L. Zgusta. – Vol. 1. – Berlin: De Gruyter, 1995. – P. 344.
7. Van Langendonck W. *Theory and Typology of Proper Names*. – Berlin: Mouton de Gruyter, 2007. – P. 65.
8. Motschenbacher H. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names // *Names*. – 2020. – P. 89.
9. Motschenbacher H. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names // *Names*. – 2020. – P. 91.
10. Motschenbacher H. Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names // *Names*. – 2020. – P. 99.