

International Conference on Computing, Artificial Intelligence and Information Systems
Hosted Online from Warsaw, Poland
Date: 11th October, 2026
Website: https://econferencia.com

FREQUENCY OF USE OF THE CONCEPT "WEALTH" IN UZBEK AND RUSSIAN LANGUAGE CORPORA AND ITS COMPARATIVE ANALYSIS

Malikov Sardor Ilkhom ugli
Independent Researcher at the Denov Institute of Entrepreneurship and Pedagogy

Abstract
This article presents a comparative analysis of the frequency of lexical units verbalizing the concept of "wealth" in Uzbek and Russian corpora. The study focuses on Uzbek units such as "boylik", "boy", "badavlat", "davlat", "mulk", and "mol-dunyo", as well as Russian units such as "богатство", "богатый", "состояние", "достаток", and "имущество". Particular attention is paid to frequency, contextual variation, genre distribution, and collocational patterns. A corpus-based approach makes it possible to examine lexical units not in isolation but in their actual textual environment. The Uzbek corpus provides search functions based on lemmas, word forms, phrases, and concordances, while the Russian National Corpus provides tools for frequency and collocation analysis. The comparative analysis demonstrates that the concept of "wealth" is verbalized in both languages through a broad range of lexical and contextual units associated with money, property, income, capital, prosperity, social status, and natural resources.

Keywords: wealth, concept, corpus linguistics, frequency, corpus analysis, collocation, concordance, verbalization, comparative analysis, Uzbek language, Russian language.



International Conference on Computing, Artificial Intelligence and Information Systems
Hosted Online from Warsaw, Poland
Date: 11th October, 2026
Website: https://econferencia.com

Introduction

Corpus linguistics is a significant field in modern linguistics that enables the statistical and semantic study of how linguistic units are used in actual speech. Based on corpus data, it is possible to determine the frequency of a specific lexeme's usage, the genres in which it is most active, the words with which it collocates, and the semantic contexts in which it appears. Consequently, corpus analysis serves as an essential methodological tool for studying the verbalization of concepts.

In the context of the Uzbek language, corpus linguistics has developed to include search capabilities for tokens, lemmas, collocations, and concordances within the Uzbek language corpus. The corpus also allows for searches across various stylistic strata—such as literary, scientific, journalistic, official, and conversational registers. This facilitates the comparison of how the concept of "wealth" ("boylik") is utilized across different communicative domains.

Meanwhile, the Russian National Corpus — a large, balanced database of texts — encompasses materials spanning various genres and historical periods. As of 2026, its core corpus contains over 426-million-word occurrences. The corpus provides specialized statistical tools for analyzing the frequency, collocations, and historical evolution of lexical units [1]. The concept of "wealth" holds particular significance for study based on this method. This is because the notion of wealth is verbalized not only through the terms "boylik" and "богатство," but also through a wide range of other lexical units — such as money, capital, income, property, prosperity, assets, opportunities, and luxury.

Analysis and Results

Studying the concept of "wealth" ("boylik") within a corpus is not limited to determining the frequency of a single word. To identify how this concept is verbalized in a corpus, it is necessary to analyze its primary nominative terms, synonyms, derivative forms, collocations, and contextual variants collectively. In the Uzbek language, the lexeme "wealth" ("boylik") itself possesses several semantic layers. According to definitions provided in an educational corpus, "wealth" can denote the state of being wealthy, an accumulation of property or assets, material possessions, or a collection of natural, material, intellectual, psychological, and spiritual resources. This polysemy indicates that the frequency with which the concept of "wealth" appears in a corpus cannot be explained by reference to a single semantic group alone.

Similarly, the Russian lexeme "богатство" can refer to material property, large financial resources, or valuable items; in a figurative sense, it may also denote the spiritual and cultural potential of an individual or a society. Consequently, when conducting a corpus analysis, it is essential to examine not only the frequency of the terms "wealth" ("boylik") and "богатство" but also the specific semantic contexts in which they are actively used. Within the Uzbek language, the lexeme "wealth" ("boylik") serves as one of the primary nominative centers for the concept of "wealth". It is semantically linked to terms such as "boy" (rich), "badavlat" (wealthy), "davlat" (wealth/state), "mulk" (property), "mol-mulk" (assets/possessions), "mol-dunyo" (worldly goods/wealth), and "sarvat" (riches/fortune). An analysis of corpus materials reveals the use of "boylik" (wealth) across various contexts. For instance, combinations such as "material wealth," "natural wealth," "national wealth," "spiritual wealth," and "cultural wealth" demonstrate the broad semantic scope of the concept of wealth. The

inclusion of terms like “natural wealth,” “national wealth,” and “spiritual wealth” in the entry for “boylik” within an educational corpus further confirms this characteristic.

From the perspective of corpus linguistics, these combinations are of particular significance. They demonstrate that the lexeme “boylik” (wealth) is actively used not only in economic contexts but also in ecological, social, and cultural ones. The search system of the Uzbek language corpus allows for searches based on lemmas, enabling various grammatical forms of “boylik” to be viewed within the scope of a single lexeme [2]. The concordance function, meanwhile, serves to examine the immediate contexts of the lexeme and to identify its semantic characteristics.

On this basis, a corpus model of the concept of “boylik” in the Uzbek language can be summarized as follows: “boylik” → property/assets → money → income → state/wealth → prosperity → natural resource → spiritual wealth. This model indicates that the verbalization of the concept of “boylik” in the Uzbek language encompasses a broad semantic field.

Corpus-based verbalization of the concept of “богатство” (wealth) in the Russian language: In Russian, the primary nominative center for the concept of “богатство” is the lexeme “богатство”, around which units such as “богатый” (rich), “разбогатеть” (to become rich), “состояние” (fortune/assets), “достаток” (affluence/means), “имущество” (property), “капитал” (capital), and “доход” (income) are clustered. Data from the national corpus make it possible to study the frequency of use and the contextual characteristics of these units within actual texts. The NKRЯ search system provides the capability to search by lemma and word form, identify collocations, and obtain statistical data

[3]. The corpus's statistical features are particularly significant for determining the frequency of collocations [4].

In contexts involving the term "богатство" (wealth), combinations such as "материальное богатство" (material wealth), "духовное богатство" (spiritual wealth), "природные богатства" (natural wealth/resources), "национальное богатство" (national wealth), and "огромное богатство" (immense wealth) actualize various semantic layers of the concept. This demonstrates that, in the Russian language, the concept of wealth encompasses both material and non-material dimensions. For instance, in the phrase "духовное богатство," wealth refers to a person's spiritual potential, whereas "природные богатства" refers to natural resources.

Consequently, associating the Russian concept of "богатство" exclusively with money and capital narrows its actual semantic potential as evidenced in a corpus.

Comparative analysis of frequency and semantic features: When comparing corpora for the Uzbek and Russian languages, it is essential to consider the size of each corpus alongside frequency data. Directly comparing raw frequency counts across corpora of different sizes is methodologically unsound. Therefore, in cross-corpus research, it is standard practice to use frequency normalized per million words — that is, relative frequency [5].

For instance, the frequency of the lexeme "boylik" in an Uzbek corpus and the frequency of "богатство" in a Russian corpus can only be directly compared if the total sizes of the corpora are identical. Otherwise, it is appropriate to use the following formula: Relative frequency = number of occurrences of a unit / total number of tokens in the corpus. This approach allows for a more accurate comparison of the actual functional activity of the units "boylik" and "богатство" across the two corpora. Furthermore, it is necessary to compare not only the

individual units “boylik” and “богатство” but also their respective semantic fields. Analyzing terms such as “boylik,” “boy,” “badavlat,” “mulk,” “mol-dunyo,” and “davlat” in Uzbek, alongside “богатство,” “богатый,” “состояние,” “достаток,” “имущество,” and “капитал” in Russian, provides a more comprehensive view of how the concept is verbalized within the corpora.

An important aspect of corpus analysis is the identification of collocations. Examining which units frequently co-occur with a given word helps to define its semantic characteristics. In Uzbek, for example, the collocation of “boylik” with terms such as “moddiy,” “ma’naviy,” “tabiiy,” “milliy,” “madaniy,” “katta,” and “ulkan” reveals various facets of the concept. “Material wealth” denotes economic content, “spiritual wealth” refers to a person’s mental and intellectual capacities, and “natural wealth” signifies a collection of resources.

In the Russian language, the combination of the word “богатство” (wealth) with terms such as “материальное” (material), “духовное” (spiritual), “природное” (natural), “национальное” (national), “огромное” (enormous), and “несметное” (immense/countless) forms distinct semantic groups. Such collocations are of particular importance for comparative analysis, as they reveal which semantic features of a given concept are most prominent or activated in each of the two languages.

The availability of tools such as “Word Portrait” and collocation statistics within the Russian National Corpus allows for a deeper investigation into the combinability and frequency profile of a specific lexeme [4]. In the Uzbek language corpus, functions for concordance, collocation, and lemma-based searching provide the foundation for such an analysis.

Based on corpus data, it is possible to distinguish between the universal and the national-cultural components of the concept of “wealth” (“boylik”). Universal

components include concepts such as money, property, income, capital, housing, and prosperity. Because these elements are rooted in the economic experience of humanity, they form similar semantic associations across different languages.

National-cultural components, by contrast, manifest in the value judgments assigned to wealth and in its relationship with other cultural values. In the Uzbek language, concepts such as “baraka” (divine blessing/abundance), “rizq” (sustenance/provision), “nasiba” (allotted share/destiny), and “qanoat” (contentment) reinforce a spiritual and value-based interpretation of wealth. In Russian, the association of the term “богатство” (wealth) with concepts like “успех” (success), “возможности” (opportunities), “роскошь” (luxury), “власть” (power), and “положение” (status/position) highlights the social and status-related components of wealth.

Conclusion

Thus, corpus frequency reveals the quantitative aspect of a concept, while collocational and contextual analysis uncovers its qualitative dimensions. The scientific novelty of this research lies in demonstrating the possibility of a comprehensive comparison of the verbalization of the concept of “wealth” in Uzbek and Russian corpora, based on frequency, collocational, and semantic criteria. Through a corpus-based approach, not only the primary nominative expressions of a concept are considered, but also other units that constitute its semantic field.

In this study, the manifestation of the concept of “wealth” in Uzbek and Russian corpora is characterized through a core and peripheral layers. The core includes direct nominative expressions for wealth and units denoting material prosperity, while the periphery comprises units representing spiritual, social, natural, and

axiological components. The results of this research can be applied in fields such as the comparative lexicology of Uzbek and Russian, cognitive linguistics, corpus linguistics, linguacultural, and concept ology. Furthermore, these findings have practical significance for defining the semantic field of the concept of "wealth" in both languages, creating corpus-based dictionaries, selecting equivalents during the translation process, and developing language-teaching materials. In particular, the use of frequency and collocation data makes it possible to determine which units of the concept of "wealth" are active in actual speech, and which are relatively peripheral.

REFERENCES:

1. Рахилина Е.В. О Национальном корпусе русского языка // Вестник Российской академии наук. — 2024. — Korpus statistikasi va kollokatsion tahlil imkoniyatlariga oid materiallar.
2. O‘zbek tili korpusi. Lemma, token, birikma va konkordans bo‘yicha qidiruv tizimi. — O‘zbek tili korpusi.
3. Национальный корпус русского языка. Поиск по лемме и словоформе, лексико-грамматический и коллокационный поиск. — НКРЯ.
4. Рахилина Е.В. О национальном корпусе русского языка: статистика и возможности корпусного анализа // Вестник Российской академии наук.
5. Бирюлин Л.А. Частотность слов и проблема сопоставимости корпусов // Русский язык в научном освещении. — 2017. — № 2. — С. 70–91.
6. Ибрагимова Б. O‘zbek tilining korpus lingvistikasi asosida so‘z chastotasi va ma’no o‘zgarishi tahlili // Science and Culture: Innovations of the 21st Century. — 2025. — Vol. 1. — № 2. — B. 127–135.
7. Бегматов Э., Ҳожиев А., Мадвалиев А. ва бошқ. O‘zbek tilining izohli lug‘ati. 6 jildlik. — Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2022.