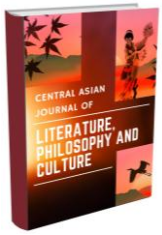

CENTRAL ASIAN JOURNAL OF LITERATURE, PHILOSOPHY, AND CULTURE

https://cajlpc.centralasianstudies.org/index.php/CAJLPC

Volume: 07 Issue: 01 | January 2026 ISSN: 2660-6828

Article

The Importance of Creating a Frequency Dictionary of Morphological Verb Forms in the Uzbek Language

Nasirdinova Oydinniso Dagarovna[*1]

1. Basic Doctoral Student, TSUUL.
* Correspondence: oydinnisonasirdinova0@gmail.com

Citation: Dagarovna N. O. The Importance of Creating a Frequency Dictionary of Morphological Verb Forms in the Uzbek Language. Central Asian Journal of Literature, Philosophy, and Culture 2026, 7(1), 260-264.

Received: 05th Sep 2025
Revised: 13th Oct 2025
Accepted: 21st Nov 2025
Published: 22nd Jan 2026

Abstract: Thus, a frequency compiled dicti onary of verb morphological forms in the Uzbek language is an important task in quantitative detecting the real,known frequency of usage of grammatical forms in real speech and the necessity of help for corpus-based linguistic investigation and language teaching. Verbs are among the most dynamic and complex grammatical classes, and the corpus-based linguostatistical analysis of absolute and relative frequencies, functional gradation, and iterative activity of inflectional forms denoting the functional styles in the "UzVerbCorpus" enable tracking their activity level across functional styles. So far, linguostatistical research on Uzbek has exclusively targeted lexicon and systematic morphostatistical research on verbs is still rare because of the agglutinative nature and structural complexity of Uzbek morphology. In our research we set out to create a methodological basis for the creation of frequency dictionaries of morphological forms of Uzbek verbs and to quantify their distribution in modern speech. This study shows the potential for morphostatistical analysis to be corpus-driven, emphasizes the core and peripheral nature of grammatical forms in terms of usage, and contributes graded data for grammar instruction and linguistic modelling. It is one of the pioneering 21st century efforts to adapt frequency dictionary development for verb grammatical forms in Uzbek using some of the emergent grammar theory and corpus linguistics principles. These results help us to develop more effective grammar instruction, calibrate assessment against international standards such as CEFR, strengthen computational linguistic resources, and further functional grammar research with real world, usage-based data.

Keywords: Verb Morphological Forms, Frequency Dictionary, Linguostatistical Analysis, Speech Gradation, Emergent Grammar, "UzVerbCorpus".

1. Introduction

The compilation of a frequency dictionary of verb morphological forms in the Uzbek language represents one of the first attempts in the 21st century to conduct systematic research on grammatical forms. Quantitative studies of speech units based on linguostatistical analysis substantiate their actual realization in contemporary speech, their degree of activity or inactivity, and their functional gradation [1], [2]. The obtained scientific results serve as a foundational basis for the development of frequency dictionaries.

Among morphological levels, the verb as a part of speech stands out due to its high degree of activity in usage, categorical complexity, and extensive functional range in speech. In determining the degree of usage of verb morphological forms, existing Uzbek language corpora are employed, absolute and relative frequencies of all morphological forms are calculated, and the forms are arranged in special tables from "most active →

least active." The ultimate goal of this methodological approach is to ensure effective teaching of grammar to native speakers and foreign learners according to a graded "simple → complex" instructional scheme, thereby enhancing the practicality, structure, and accessibility of the learning process [3], [4]. Quantitative, scientifically grounded research on verb morphological forms gives rise to the need for a new specialized lexicographic resource, a frequency dictionary.

In Uzbek linguistics, the number of scholars who have conducted linguostatistical research on morphological units and produced frequency dictionaries on this basis is limited [5]. The introduction of quantitative methods into linguistic research, which began to take shape in the 1960s–1970s, primarily focused on lexical units, literary texts classified by functional styles, and relatively small-scale written texts such as individual stories, works, or epics [6].

One of the most significant studies of large-scale linguostatistical analysis in the 20th century is associated with S. Rizaev. As a result, by the 1980s, his *Frequency Dictionary of the Language of Uzbek Children's Literature (compiled using electronic computing machines)* was introduced into scholarly circulation. The dictionary was compiled based on 100,000 words selected from literary works published between 1964 and 1974 using computer-based processing. The number of word forms selected as dictionary units amounts to 26,752 [7], [8]. However, this dictionary focuses primarily on phonetic and lexical statistical analyses rather than morphostatistical data.

Rizaev's research lacks a specialized methodology for applying linguostatistical analysis specifically to parts of speech; instead, phonetic, lexical, and grammatical units are treated collectively [9].

2. Materials and Methods

The method of this research is corpus-driven linguostatistics, which is generally applied in order to determine the frequency of verb morphological forms in modern Uzbek. Empirical facts have been collected from the specially designed research corpus "UzVerbCorpus", including 1276140 running words and balanced in four written functional styles: literary, scientific, journalistic, and official texts. It is important to achieve representativeness of modern written Uzbek in such a corpus and avoid stylistic background bias in lexical frequency calculation. Because Uzbek is agglutinative, all the verb forms appearing in the corpus were obtained automatically with tools for corpus processing, and then were verified manually for their morphology. Calculations were made of absolute and relative frequencies of each verb morphological form, and forms were ordered from most to least active in terms of actual speech use. Linguostatistical methods were used to discover distributional, functional gradation, and center-periphery patterns of grammatical categories consistent with usage and emergent grammar notions. We organized the quantitative results into frequency tables to facilitate systematic comparisons of categorical forms between individual styles. The analysis works on the approach of "from speech to grammar", which gives more importance to data from actual usage of language compared to a rule-based description of grammar. This methodical framework offers an empirically grounded frequency dictionary of verb morphological forms, while also allowing for applications in the areas of language teaching, corpus linguistics, and computational modeling. The study, based on large-scale corpus data and reproducible statistical procedures, guarantees reliability, objectivity, and relevance to the current stage of the development of the Uzbek language.

3. Results and Discussion

The first experimental work in Uzbek linguistics is S. Muhamedov's Alphabetical Frequency Dictionary of the Uzbek Language, compiled on the basis of linguostatistical calculations derived from a 200,000-word sample of newspaper texts. This frequency dictionary covers all parts of speech and is arranged alphabetically. Verbs are presented as word forms with various categorical affixes. The verb *qilmoq* ("to do"), which functions as both a main and auxiliary verb and is among the most frequent verbs in speech, is represented by 40 word forms with different morphological affixes. Among them, the action noun form *qilish* demonstrates the highest frequency [10].

Notably, tense, person-number, and mood forms expressing predicative categories are not presented separately; instead, word forms with these affixes are collectively counted under the base form *qilmoq*. Despite this limitation, the dictionary is significant as a frequency dictionary that partially incorporates morphostatistical analysis. Overall, previous linguostatistical research in Uzbek linguistics has predominantly focused on lexical units, while quantitative analyses of morphological forms have been addressed only marginally. This may be attributed to the agglutinative nature of Uzbek, the abundance of morphological forms, and the complexity of their morphemic structure [11].

Studies indicate that worldwide research on verb frequency began in the 1950s–1960s. After the investigation of linguistic properties of language units, scholarly interest shifted toward determining their degree of usage in speech. In English linguistics, Joan Bybee demonstrated that grammatical forms exhibit "central" or "peripheral" status in usage, revealing that grammatical systems form a quantitative hierarchy. She showed that tense categories in verbs are central, whereas mood forms are secondary in importance.

Paul Hopper, approaching this issue from a similar perspective, introduced the concept of emergent grammar in 1987. Emergent grammar views morphological units as dynamic phenomena shaped by actual speech usage and communicative needs, measuring the centrality of grammatical forms through their frequency. The works of Bybee and Hopper complement each other and are frequently cited together. Their findings highlight the importance of applying emergent grammar principles in Uzbek linguistics by examining the "life of linguistic units in speech" through precise quantitative characteristics.

According to Bybee, grammar is not a static system of rules but a cognitively shaped structure formed through repeated usage in speech. High-frequency constructions indicate ongoing grammaticalization processes and the transition of constructions into central category members. This perspective supports the statistical description of morphological forms and the methodology of compiling frequency dictionaries. Consequently, instead of the traditional approach "from grammatical rules to speech," it becomes methodologically appropriate to adopt the principle "from speech to grammar" when analyzing language and speech phenomena [12].

Based on emergent grammar theories, numerous frequency dictionaries have been created worldwide. In English linguistics, *A Frequency Dictionary of Contemporary American English* by M. Davies and D. Gardner was compiled based on statistical analyses of the 450-million-word COCA corpus [13]. This comprehensive frequency dictionary includes all parts of speech and arranges words from "most frequent to least frequent." Dictionary entries provide numbering, part-of-speech identification, absolute and relative frequencies, and common collocations and phraseological usage, demonstrating practical applicability in language teaching. The dictionary covers the 5,000 most frequently used words across 30 thematic groups and has been digitized and integrated into the COCA corpus following the "from corpus to corpus" principle [14].

In Russian linguistics, *A Frequency Dictionary of Russian* by S. Sharoff, E. Umanskaya, and J. Wilson includes the 5,000 most frequent words and 300 expressions across 26 thematic groups. The dictionary is based on statistical analysis of the Russian National

Corpus, comprising over 150 million words. Each entry presents the Russian word, its English translation, part-of-speech classification, syntactic position, and absolute and relative frequencies. In recent decades, research has increasingly focused on studying the realization of morphological forms in speech, their quantitative and qualitative characteristics, and the evaluation of the Uzbek language's real state through digitization and computer linguistic tools. Establishing a methodology for grammatical statistics and systematizing grammar instruction based on scientifically grounded analyses have become key objectives.

Relying on corpus practice facilitates the creation of new large-scale frequency dictionaries in Uzbek, enhancing accuracy and reliability. Since morphological research in Uzbek language corpora remains an ongoing process, utilizing existing resources is both necessary and effective. Accordingly, linguostatistical practices for frequency dictionaries are conducted in line with learners' needs and current communicative contexts.

The research corpus known as **"UzVerbCorpus,"** consisting of 1,276,140 words and evenly distributed across four written functional styles literary, scientific, journalistic, and official was specifically designed to determine the degree of usage of verb morphological forms in speech. The results of this corpus provide a foundation for compiling another significant frequency dictionary for Uzbek linguistics in the 21st century. Moreover, given the governmental prioritization of enhancing the global status of the Uzbek language through the development of the "Uzbek Language Corpus" and its integration into artificial intelligence systems, the present study gains further relevance as a "corpus-driven corpus-enhancing" endeavor [15].

A frequency dictionary of morphological forms differs fundamentally from traditional lexicographic dictionaries, where the lexeme serves as the primary object. Morphological forms prepare lexemes for speech and can metaphorically be described as the "clothing of lexemes," selected according to communicative context. In 20th-century Uzbek linguistics and lexicographic studies, frequency dictionaries based on morphological form usage are virtually absent. However, it is precisely morphological forms that actualize lexemes in speech and ensure their functional vitality. Given the agglutinative richness of Uzbek and the diversity of grammatical meanings, it is methodologically justified to determine the degree of usage of morphological forms across all parts of speech within specific periods, speech types, and styles using corpus-based linguostatistical analysis. Although this process is complex and labor-intensive, it encourages collaborative, system-oriented research. As a result, new frequency dictionaries reflecting active and inactive word forms, learner-oriented dictionaries, and purpose-specific lexical resources can be developed. Morphological linguostatistics also enables the study of forms in both synchronic and diachronic perspectives, demonstrating the extensive scope of this field.

4. Conclusion

The primary motivation for compiling a frequency dictionary of verb morphological forms in the Uzbek language lies in identifying the quantitative representation of verb categorical forms in real speech. Such a dictionary enables the analysis of usage ratios between lexical and predicative verb forms, their degrees of activity or inactivity, and provides a scientific explanation of the discrepancies between language and speech. Secondly, it contributes to clarifying the normative state of modern Uzbek, allowing grammatical rules and textbooks to be aligned with actual language usage. For textbooks and teaching materials designed for native and foreign learners, the degree of activity of linguistic units in speech is crucial; however, due to the lack of targeted linguostatistical research, this aspect has largely been overlooked. Determining the activity level of linguistic units facilitates the adaptation of Uzbek language assessment to international

standards, including CEFR requirements. Thirdly, a frequency dictionary of verb morphological forms serves as an essential empirical foundation for corpus and computational linguistics. In developing automatic text analysis systems, morphological analyzers, and syntactic parsers for Uzbek, frequency-based models of verb forms provide reliable statistical support. Fourthly, in enhancing the scientific basis and effectiveness of Uzbek language teaching methodology, it becomes evident that teaching all verb forms at the same level is an outdated and inefficient approach. Frequency dictionaries clearly demonstrate the necessity of prioritizing high-frequency forms. Finally, updating functional grammar knowledge from a contemporary perspective and accurately identifying the stylistic distribution of verb morphological forms contribute to the advancement of functional grammar research in Uzbek linguistics.

REFERENCES

[1] M. Askarova, "The emergence and development of corpus linguistics," *Til va adabiyot.uz*, no. 5, pp. 49–51, 2024.

[2] J. Bybee, "From usage to grammar: The mind's response to repetition," *Language*, vol. 82, no. 4, pp. 711–733, 2006, doi: 10.1353/lan.2006.0186.

[3] M. Davies and D. Gardner, *A Frequency Dictionary of Contemporary American English*. London, UK: Routledge, 2008.

[4] P. J. Hopper, "Emergent grammar," in *Proceedings of the Berkeley Linguistics Society*, vol. 13, Berkeley, CA, USA, 1987, pp. 139–157.

[5] S. Muhamedov, *An Alphabetical Frequency Dictionary of the Uzbek Language*. Tashkent, Uzbekistan: Fan Publishing House, 1982.

[6] S. Sharoff, E. Umanskaya, and J. Wilson, *A Frequency Dictionary of Russian: Core Vocabulary for Learners*. London, UK: Routledge, 2013.

[7] B. Yo'ldoshev, *Linguostatistical Methods of Text Analysis: A Methodological Guide (Part 1)*. Samarkand, Uzbekistan: Samarkand State University Press, 2008.

[8] National Database of Legislation of the Republic of Uzbekistan, "On priority projects for the implementation of artificial intelligence technologies in 2025–2026," Lex.uz, No. 425, Jul. 10, 2025. [Online]. Available: https://lex.uz

[9] S. A. Rizayev, *A Linguostatistical Study of the Uzbek Language*, Doctoral dissertation abstract. Tashkent, Uzbekistan, 2008, 50 p.

[10] F. Usmanov, "Problems and prospects of compiling a frequency dictionary of the modern Uzbek language," *SSRN Electronic Journal*, Art. no. 5563839. [Online]. Available: https://ssrn.com/abstract=5563839

[11] I. Z. Soliyevich, "The problem of morphological construction of the verb in improving the system of digital automatic translation of words in the Uzbek language," *Journal of Positive School Psychology*, vol. 7, no. 2, 2023.

[12] N. Murzintceva and S. S. Yuldasheva, "Uzbek language morphology analyser," *Natural Language Processing Journal*, Art. no. 100195, 2025.

[13] N. F. Kamolovna, "Modeling analytical verb forms for the morphological analyzer of the Uzbek language," *American Journal of Philological Sciences*, vol. 4, no. 4, pp. 111–116, 2024.

[14] D. Khudoynazarova, "Comparative analysis of word formation systems in English and Uzbek languages," *Lingvospektr*, vol. 1, no. 1, pp. 139–145, 2025.

[15] S. Omarova, D. Ospanova, N. Aitova, G. Tokenkyzy, A. Ormanova, and M. Alshynbekova, "A corpus approach in language discovery: A word frequency analysis based on the corpus outcomes in Kazakh," *IEEE Access*, 2025.