https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Abduraximova Feruza Boxadirovna
Senior Teacher (PhD)
Faculty of English Language and Literature
Fergana State University
feruz.abdurakhimova818@gmail.com
Diyorakhon Ibrohimova Xusanovna
Student
Faculty of English Language and Literature
Fergana State University
ibrohimovadiyora9904@gmail.com

CORPUS-BASED THEORIES OF THE SEMANTIC SYSTEM OF SYNONYMS IN UZBEK AND ENGLISH

Abstract

This article investigates corpus-based theoretical approaches to the semantic system of synonyms in Uzbek and English. Unlike traditional synonym studies grounded in intuition and lexicographic description, corpus linguistics provides empirical, usage-based evidence drawn from authentic language data. The research integrates data from major English corpora, including the British National Corpus and the Corpus of Contemporary American English, alongside the Uzbek National Corpus. Using concordance analysis, collocational statistics, and frequency distribution, the study reveals systematic semantic differentiation among near-synonyms across both languages. The paper incorporates established theories of usage-based semantics, distributional meaning, and semantic prosody to strengthen analytical validity. A contrastive perspective highlights typological differences between English as an analytic language and Uzbek as an agglutinative language. The study proposes an expanded corpus-based semantic classification model for Uzbek synonyms, representing a methodological innovation in Uzbek lexical

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

semantics and contributing to future corpus-driven linguistic and computational research.

Keywords: corpus linguistics, synonymy, semantic system, Uzbek language, English language, usage-based semantics, collocation, frequency analysis

Introduction

Synonymy occupies a central yet controversial position in lexical semantics. Classical semantic theory traditionally defined synonyms as lexical units sharing identical or near-identical meanings. However, modern linguistics increasingly rejects the notion of absolute synonymy, emphasizing that even closely related words differ in distribution, stylistic value, pragmatic function, and collocational behavior (1). These differences become especially visible when language is analyzed through empirical data rather than intuition.

The rise of corpus linguistics in the late twentieth century marked a methodological shift in semantic research. Instead of relying on introspection, corpus linguistics examines meaning as it emerges from repeated patterns of actual language use (2). This approach aligns with usage-based theories of language, which argue that semantic knowledge is shaped by frequency, context, and cognitive entrenchment (8).

In English linguistics, corpus-based synonym studies have benefited from large, balanced corpora such as the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). These resources enable fine-grained analysis of register variation, collocational strength, and semantic prosody (3,5). By contrast, Uzbek synonymy has been predominantly studied within descriptive, normative, and lexicographic frameworks, with limited empirical corpus support (7).

The recent development of the Uzbek National Corpus represents a turning point in Uzbek linguistics, enabling data-driven semantic research for the first time (4). This study aims to integrate corpus-based semantic theories into the analysis of

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Uzbek and English synonym systems, addressing the gap between traditional semantic description and empirical linguistic evidence.

Methods

This study employs a mixed qualitative–quantitative corpus-based methodology grounded in distributional and usage-based semantic theory. English data were extracted from the British National Corpus and the Corpus of Contemporary American English, while Uzbek data were obtained from the Uzbek National Corpus.

Synonymous lexical pairs were selected based on three criteria: dictionary equivalence, minimum frequency thresholds, and cross-register occurrence. This ensured representativeness and avoided idiosyncratic or marginal lexical items.

Three core corpus-linguistic techniques were applied. First, concordance analysis was used to examine contextual realizations of synonyms, allowing observation of pragmatic roles and semantic restrictions (2). Second, collocation analysis measured statistically significant co-occurrence patterns, revealing semantic prosody and functional preference (5,9). Third, frequency analysis identified distributional dominance across genres, supporting usage-based semantic interpretation (6).

For Uzbek data, additional normalization procedures were applied due to agglutinative morphology. Lemmatization and morphological filtering were necessary to unify inflectional variants under a single lexical head (7). The analytical framework was informed by distributional semantics, according to which meaning emerges from patterns of co-occurrence (10).

Results

The corpus findings demonstrate that English synonyms exhibit systematic semantic specialization. Near-synonyms such as begin and start share a core semantic concept but differ in register distribution and collocational strength.

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

Begin frequently occurs in academic and formal written contexts, while start dominates in spoken discourse and informal genres (3).

Similarly, Uzbek synonym pairs such as *boshlamoq* and *yo‘lga qo‘ymoq* show clear functional differentiation. Corpus evidence indicates that yo‘lga qo‘ymoq strongly collocates with abstract and institutional nouns, reflecting a procedural and administrative semantic profile. In contrast, boshlamoq displays broader semantic scope and higher overall frequency.

Frequency analysis further confirms that high-frequency synonyms tend to be semantically generalized, whereas lower-frequency items encode more specialized meanings. In Uzbek, morphological affixation contributes to subtle semantic variation, reinforcing the role of form–meaning interaction in agglutinative languages (7).

Discussion

The results strongly support usage-based and corpus-driven theories of synonymy, which view meaning as emergent from contextualized language use rather than static dictionary definitions (1,8). Corpus evidence reveals that synonyms form dynamic semantic networks shaped by frequency, collocation, and pragmatic constraints.

A major theoretical contribution of this study is the development of an expanded corpus-based semantic classification model for Uzbek synonyms. While inspired by English corpus methodologies such as Word Sketch theory (5), the model incorporates Uzbek-specific morphological and stylistic features. This adaptation represents an important methodological innovation for Uzbek lexical semantics.

The findings also reinforce the relevance of distributional semantics, demonstrating that semantic similarity and difference can be quantitatively modeled through corpus patterns (10). Integrating Uzbek data into this framework

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

aligns national linguistic research with international standards and supports the development of computational linguistic resources.

Conclusion

This study introduces a theoretically enriched, corpus-based approach to the analysis of synonymy in Uzbek and English. By combining usage-based semantics, distributional theory, and empirical corpus analysis, the research demonstrates that synonymy is best understood as a gradient, context-sensitive phenomenon.

The novelty of the article lies in:

1. integrating Uzbek corpus data into internationally recognized corpus-semantic theories;

2. proposing a corpus-based semantic classification model adapted to agglutinative morphology;

3. empirically demonstrating cross-linguistic differences in synonym systems.

Future research directions include expanding the corpus dataset, developing Uzbek WordNet-style semantic networks, investigating diachronic synonym change, and applying corpus-based synonym analysis to machine translation and language teaching contexts.

References

1. Cruse A. Meaning in Language: An Introduction to Semantics and Pragmatics. 3rd ed. Oxford: Oxford University Press; 2011.

2. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press; 1991.

3. Davies M. The Corpus of Contemporary American English (COCA): 560 million words, 1990–present. Int J Corpus Linguist. 2008;13(2):159–190.

4. Uzbek National Corpus [Internet]. Tashkent; 2023 [cited 2025]. Available from: https://uzbekcorpus.uz

https://wordlyknowledge.uz/index.php/gpble
Qabul qilingan sana: 07.12.2025
Nashr etilgan sana: 30.01.2026

5. Kilgarriff A. Word sketches, collocations and semantic similarity. Int J Lexicogr. 2010;23(1):1–20.

6. Stubbs M. Words and Phrases: Corpus Studies of Lexical Semantics. Oxford: Blackwell; 2002.

7. Jalolov J. O‘zbek tilshunosligida leksik-semantik tadqiqotlar. Tashkent: Fan; 2019.

8. Bybee J. Language, Usage and Cognition. Cambridge: Cambridge University Press; 2010.

9. Hunston S. Corpora in Applied Linguistics. Cambridge: Cambridge University Press; 2002.

10. Firth JR. A synopsis of linguistic theory. Stud Linguist. 1957;11:1–32.