



EOC
EUROASIAN
ONLINE
CONFERENCES

ENGLAND CONFERENCE

**INTERNATIONAL CONFERENCE ON
MULTIDISCIPLINARY STUDIES AND
EDUCATION**



Google Scholar

zenodo

OpenAIRE

doi digital object
identifier

eoconf.com - from 2024



INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY STUDIES AND EDUCATION: a collection scientific works of the International scientific conference – London, England, 2026. Issue 7

Languages of publication: Uzbek, English, Russian, German, Italian, Spanish

The collection consists of scientific research of scientists, graduate students and students who took part in the International Scientific online conference «**INTERNATIONAL CONFERENCE ON MULTIDISCIPLINARY STUDIES AND EDUCATION**». Which took place in London 2026.

Conference proceedings are recommended for scientists and teachers in higher education establishments. They can be used in education, including the process of post - graduate teaching, preparation for obtain bachelors' and masters' degrees. The review of all articles was accomplished by experts, materials are according to authors copyright. The authors are responsible for content, researches results and errors.





LEXICAL-SEMANTIC DIFFERENTIATION OF SYNONYMS IN TURKIC LANGUAGES: AN AI AND PARALLEL CORPUS-BASED APPROACH TO TRANSLATOR TRAINING

Hasanova Gulhida Abdulaziz qizi

PhD student

Uzbekistan State World Languages university

+998909284198

gulhida0898@gmail.com

orcid: 0009-0000-6972-6772

Abstract. Despite extensive research on synonymy in linguistics, the problem of accurate synonym selection in translation remains unresolved, particularly in Turkic languages where subtle lexical-semantic distinctions significantly affect meaning and stylistic adequacy. This paper proposes an AI-supported, parallel corpus-based approach to teaching synonym usage differentiation for translator training. Drawing on Uzbek–English parallel texts, the study demonstrates how corpus evidence, combined with AI-assisted analysis, can reveal fine-grained distinctions between synonymous units, especially within action verbs. The findings suggest that integrating parallel corpora and artificial intelligence tools into translator education enhances learners’ ability to identify contextual, stylistic, and semantic nuances of synonyms, thereby improving translation quality. The paper contributes both a methodological framework and pedagogical implications for adapting Turkic languages to modern language technologies.

Keywords: synonymy, lexical-semantic differentiation of action verbs, parallel corpora, artificial intelligence, translator training, translation studies; collocational behavior; semantic nuance

Introduction. In contemporary linguistics and translation studies, synonymy is recognized as a complex and multidimensional phenomenon. Although synonymous units share a core semantic component, they differ in stylistic, pragmatic, and contextual features. These distinctions become particularly critical in translation, where inappropriate synonym selection may lead to semantic distortion or stylistic inadequacy. In Turkic languages, including Uzbek, synonymy is highly developed and context-sensitive. However, traditional translator training often treats synonyms as interchangeable units, which results in errors in both meaning transfer and discourse appropriateness. At the same time, recent advancements in artificial intelligence (AI) and corpus linguistics provide new opportunities to model and teach these distinctions more effectively. This paper aims to develop a methodology for teaching lexical-semantic differentiation of synonyms using parallel corpora and AI-assisted tools. The focus is placed on action verb synonyms, as they demonstrate high variability in semantic and functional properties.

Literature Review. The phenomenon of synonymy has long occupied a central position in linguistic research, yet it remains one of the most theoretically complex and methodologically challenging areas within lexical semantics.





Traditional approaches often treated synonyms as words with identical or nearly identical meanings; however, contemporary linguistic theory emphasizes that complete synonymy is rare and that most synonymous units differ along multiple semantic and pragmatic dimensions. These differences may include variations in intensity, evaluative meaning, stylistic marking, and contextual distribution. Within this framework, lexical-semantic differentiation becomes a crucial analytical task, requiring the identification of fine-grained distinctions between units that share a common semantic core. The cluster-based approach to synonymy, as developed in computational and theoretical linguistics, further reinforces this perspective by modeling synonyms as networks of related meanings rather than discrete equivalents [Edmonds, 1999].

In Uzbek linguistics, synonymy has also been extensively studied, particularly in relation to lexical systematization and stylistic variation. Research has shown that Uzbek synonyms often form rich and hierarchically organized leksik-semantik guruhlar, within which individual units differ in usage frequency, expressive value, and combinability. Scholars have emphasized the importance of analyzing synonyms not only at the level of abstract meaning but also in terms of their functional behavior in actual discourse [Abdurahmonov, 1980; Rahmatullayev, 2006]. Furthermore, the distinction between ma'no sinonimlari, uslubiy sinonimlar, and nutqiy sinonimlar highlights the multidimensional nature of synonymy in Uzbek and underscores the need for context-sensitive analysis [Jamolxonov, 2005]. Despite these advances, the application of such theoretical insights to translation practice remains insufficiently developed, particularly in the context of systematic translator training.

Parallel corpora have emerged as a powerful empirical resource in translation studies, enabling researchers to observe how lexical units are used and translated across languages in authentic contexts. Unlike traditional bilingual dictionaries, which often provide static and decontextualized equivalents, parallel corpora offer dynamic evidence of actual translation choices, including variation and inconsistency. Corpus-based approaches have demonstrated that translators frequently rely on a limited set of equivalents, leading to the neutralization of semantic distinctions present in the source language [Baker, 1992; Sinclair, 1991]. Further developments in corpus linguistics have introduced quantitative methods for analyzing frequency, collocation, and distribution, thereby enabling more systematic investigation of lexical patterns [Stubbs, 2001; Gries, 2009; McEnery & Hardie, 2012].

The integration of artificial intelligence into linguistic and translation studies has significantly expanded the analytical possibilities of corpus-based research. AI-driven tools, particularly those grounded in natural language processing, allow for the automated extraction, classification, and comparison of large-scale linguistic data. These technologies facilitate the identification of subtle semantic distinctions by detecting patterns that may not be immediately visible through manual analysis. Early work on computational linguistics highlighted the importance of large corpora





for modeling language behavior and improving lexical choice in automated systems [Church & Mercer, 1993]. More recent developments in AI and NLP have further enhanced these capabilities, enabling more sophisticated representations of meaning and context [Jurafsky & Martin, 2023]. In the field of translation studies, such tools also contribute to improving translation quality assessment by providing objective, data-driven criteria for evaluating lexical choices [House, 2015].

Despite these advancements, the application of AI and corpus-based methods to the teaching of synonym differentiation in Turkic languages remains limited. Most existing studies focus either on theoretical aspects of synonymy or on technical developments in NLP, without adequately bridging the gap between linguistic analysis and pedagogical practice. Consequently, there is a need for an integrated approach that combines lexical-semantic theory, corpus methodology, and AI technologies in order to address the specific challenges of translator training. The present study responds to this need by proposing a framework in which parallel corpora and AI-assisted analysis are used not only as research tools but also as instructional resources for developing translators' competence in synonym usage differentiation.

Methodology. The present study is based on the analysis of Uzbek–English parallel texts, primarily drawn from literary works and their professionally produced translations. The selection of the corpus was guided by the need to ensure contextual diversity, the presence of clearly identifiable synonymic clusters, and a high level of translation reliability. Literary texts were prioritized because they provide a rich environment for observing nuanced lexical choices, particularly in the domain of action verbs, where synonymy is both frequent and functionally significant.

The analytical framework of the research is grounded in a lexical-semantic approach to synonym differentiation, with a particular focus on action verbs belonging to the same leksik-semantik guruh. Within this framework, synonymous units are not treated as interchangeable elements but are examined in terms of their distinctive semantic and functional properties. The differentiation process is carried out through several interrelated parameters, including denotational nuance, which reflects subtle differences in core meaning; degree and intensity of the action; stylistic marking and register; frequency of usage in authentic discourse; and collocational behavior, or *to'ldiruvchi xususiyatlari*, which determine the compatibility of verbs with specific types of objects and syntactic environments.

In order to enhance the empirical rigor of the analysis, AI-assisted tools are employed throughout the research process. These tools facilitate the extraction of concordance lines from the parallel corpus, enabling systematic observation of usage patterns across a wide range of contexts. Furthermore, frequency analysis allows for the identification of dominant and peripheral synonyms within a given cluster, while clustering algorithms assist in grouping lexical units according to contextual similarity. AI-based alignment techniques are also used to detect translation correspondences between Uzbek and English units, making it possible to trace how synonymic distinctions are preserved, neutralized, or altered in translation. The





integration of these technologies ensures both scalability and precision in the analysis, allowing for a more comprehensive understanding of synonym behavior than would be achievable through manual methods alone.

Analysis and Results. The analysis reveals that action verbs in Uzbek frequently form dense synonymic clusters characterized by fine-grained semantic differentiation. A representative example is the cluster *kesmoq* – *qirqmoq* – *chopmoq*, which shares a common denotational core related to the act of cutting but diverges significantly in terms of semantic nuance and functional application. The verb *kesmoq* functions as a neutral and broadly applicable unit, typically used in contexts where the action of cutting is not further specified in terms of manner or intensity. In contrast, *qirqmoq* often implies a more controlled, precise, or repeated action, frequently associated with activities requiring careful manipulation. The verb *chopmoq*, on the other hand, conveys a sense of forcefulness and intensity, often involving rough or vigorous action directed at more resistant objects.

Parallel corpus evidence demonstrates that these distinctions are not consistently maintained in translation. In many cases, all three Uzbek verbs are rendered by the single English equivalent “to cut,” which results in a loss of semantic specificity. This phenomenon can be described as neutralization, where the target language fails to reflect the internal differentiation present in the source language. However, the corpus also provides instances where translators employ more specific equivalents such as “to chop” or “to slice,” indicating an awareness of these nuances, albeit not in a systematic manner.

Further analysis identifies several recurrent translation patterns and problems. Among these, the most prominent is the tendency toward overgeneralization, where a semantically broad equivalent replaces multiple distinct source-language units. This is often accompanied by the loss of stylistic nuance, particularly in cases where the original verb carries expressive or evaluative connotations. Additionally, instances of contextually inappropriate synonym choice are observed, suggesting that translators may rely on dictionary equivalence rather than contextual analysis. The use of AI-assisted tools proves particularly effective in uncovering these patterns, as frequency distributions and concordance data make discrepancies and inconsistencies readily visible.

Discussion. The findings of the study highlight the limitations of traditional approaches to teaching synonymy in translator training, which often rely on static lists of equivalents and insufficiently account for contextual variability. In contrast, the integration of parallel corpora and artificial intelligence offers a dynamic and data-driven alternative that reflects actual language use. By exposing learners to authentic examples and enabling them to observe patterns across multiple contexts, this approach fosters a more nuanced understanding of lexical-semantic relationships.

From a methodological perspective, the use of AI tools introduces a level of analytical depth and efficiency that is particularly valuable in the study of synonymy. Automated extraction and classification of data allow researchers and learners to





process significantly larger datasets, thereby increasing the reliability of their conclusions. Moreover, the ability to visualize frequency patterns and contextual distributions supports the development of analytical skills that are essential for professional translators.

In the context of Turkic languages, the proposed approach is especially relevant due to the richness and complexity of their synonymic systems. As these languages increasingly interact with digital technologies and AI-driven applications, the need for precise lexical modeling becomes more urgent. The incorporation of corpus-based and AI-assisted methods into translator education thus contributes not only to improved translation quality but also to the broader process of adapting Turkic languages to contemporary technological environments.

Conclusion. The present study demonstrates that lexical-semantic differentiation of synonyms, particularly within the domain of action verbs, can be effectively addressed through an integrated approach combining parallel corpus analysis and artificial intelligence. By systematically examining semantic nuance, stylistic features, and collocational behavior, the research provides a detailed account of how synonymous units function in authentic discourse and how their distinctions are treated in translation.

The results indicate that AI-supported corpus analysis enhances the accuracy and depth of synonym differentiation, thereby offering significant benefits for translator training. Learners who engage with this approach are better equipped to recognize subtle differences between lexical units and to make contextually appropriate choices in translation. At the same time, the methodology contributes to the development of more sophisticated language technologies capable of handling the complexity of Turkic languages.

Future research may extend this approach by incorporating larger and more diverse corpora, exploring additional semantic domains, and applying advanced AI techniques such as semantic vector modeling. Such developments would further strengthen the interface between linguistics, translation studies, and artificial intelligence.





References

1. Baker, M. (1992). *In Other Words: A Coursebook on Translation*. London: Routledge.
2. Edmonds, P. (1999). *Semantic Representations of Near-Synonyms for Automatic Lexical Choice*. PhD Dissertation, University of Toronto.
3. Gries, S. T. (2009). *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge.
4. Jurafsky, D., & Martin, J. H. (2023). *Speech and Language Processing* (3rd ed.). Draft version. Stanford University.
5. Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
6. Stubbs, M. (2001). *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell.
7. Church, K. W., & Mercer, R. L. (1993). Introduction to the special issue on computational linguistics using large corpora. *Computational Linguistics*, 19(1), 1–24.
8. McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
9. Nida, E. A. (1964). *Toward a Science of Translating*. Leiden: Brill.
10. House, J. (2015). *Translation Quality Assessment: Past and Present*. London: Routledge.

