

**CRITERIA FOR INCLUDING NEOLOGISMS FROM ENGLISH AND UZBEK INTO
BILINGUAL CORPORA AS LEXICAL UNITS**

Ergashev Tulkinjon Tukhtamurodovich

Senior lecturer (PhD), department of philology University
of information technologies and management

Abstract

The dynamic nature of language is most clearly reflected in the continuous emergence of neologisms. In an era of rapid technological and social change, both English and Uzbek are generating new lexical units at an unprecedented rate. This article provides a detailed examination of the criteria and requirements for including neologisms from English and Uzbek into bilingual corpora as fully recognized lexical units. It analyzes semantic, pragmatic, functional, and cross-linguistic factors that determine whether a new word has achieved sufficient stability and relevance for corpus integration. The study compares neologism formation processes in both languages and proposes a comprehensive set of criteria, including frequency and dispersion, semantic stability, collocational productivity, cross-linguistic equivalence, and temporal persistence. Within the anthropocentric paradigm, the inclusion of neologisms is interpreted as evidence of human communicative creativity and adaptation to new social realities. The findings offer practical guidelines for corpus linguists, lexicographers, and language technology specialists working on English-Uzbek bilingual resources.

Keywords: neologisms, bilingual corpora, lexical units, English-Uzbek, corpus linguistics, semantic stability, digital communication, anthropocentric paradigm, cross-linguistic equivalence.

Introduction. Language is a living system that constantly evolves in response to human needs, technological progress, and social transformations. Neologisms – newly coined or newly adopted words – represent the most visible manifestation of this dynamism. In the digital age, the speed and volume of neologism creation have increased dramatically, particularly in domains related to technology, social media, economy, and culture.

English, functioning as a global lingua franca, produces and disseminates a vast number of neologisms that quickly spread across languages through borrowing and adaptation. Uzbek, while developing its own internal neologisms through word-formation processes, is also heavily influenced by English borrowings, especially in digital and professional communication.

For bilingual corpora to accurately represent contemporary language use, they must incorporate neologisms as legitimate lexical units. However, the decision to include a new word in a corpus is not straightforward. It requires clear, evidence-based criteria that balance linguistic accuracy, corpus representativeness, and practical usability.

This article aims to establish comprehensive requirements for the inclusion of neologisms from English and Uzbek into bilingual corpora. It examines the semantic and functional characteristics of neologisms in both languages, compares their integration processes, and proposes a systematic framework for corpus compilers.

Methods. The study adopts a comparative, corpus-based, and analytical methodology. Data were collected from English and Uzbek corpora, digital communication platforms, lexicographic sources, and academic literature on neologism studies and bilingual corpus compilation.

Neologisms were analyzed according to multiple parameters: frequency of occurrence, dispersion across text types, semantic stability, collocational behavior, pragmatic function, and

the existence of cross-linguistic equivalents. Special attention was given to the challenges of aligning neologisms across languages with different morphological systems (analytic English vs. agglutinative Uzbek). The analysis also incorporates principles from computational lexicography and the anthropocentric view of language as a reflection of human social and cognitive activity.

Results. In English, neologisms are frequently created through blending (*brunch, infodemic*), clipping (*app, blog*), affixation (*selfie, influencer*), and semantic shift (*ghosting, doomscrolling*). Many of these terms originate in digital and social media contexts and rapidly achieve widespread usage.

In Uzbek, neologisms emerge through internal resources (affixation, compounding, semantic extension) as well as through borrowing and adaptation from English. Examples include direct borrowings such as *kontent, influenser, viral, stori, ril*, and native formations reflecting local realities. Some words undergo semantic extension (e.g., existing Uzbek words acquiring new digital meanings).

Based on the analysis, the following six core criteria are recommended for including neologisms as lexical units in English-Uzbek bilingual corpora:

1. A neologism should demonstrate consistent and repeated use across different genres, registers, and communicative situations over a meaningful period. Isolated or highly context-specific occurrences should not automatically qualify for inclusion. Quantitative thresholds (e.g., minimum frequency per million words) can be established depending on corpus size and domain.

2. The word should possess a relatively stable core meaning, even if it retains some degree of polysemy or contextual variation. Words with highly unstable or rapidly shifting meanings require longer observation before being accepted as established lexical units.

3. Productive use in various syntactic constructions and with diverse collocates indicates that the word has become an active and functional part of the lexicon. High collocational productivity is a strong indicator of lexicalization.

4. For bilingual corpora, there should be a clear or emerging equivalent (or functional correspondence) in the partner language. Purely culture-specific neologisms may require special annotation rather than forced equivalence. The existence of translation equivalents or explanatory paraphrases facilitates integration.

5. The neologism should appear in naturally occurring, authentic texts rather than only in artificial or dictionary examples. Its distribution across formal, informal, professional, and social media registers should be documented.

6. A minimum observation period (typically 2–5 years, depending on the domain) is recommended. This helps distinguish between transient buzzwords and neologisms that have achieved lasting integration into the language. Longitudinal tracking of frequency and usage patterns is particularly valuable.

Discussion. The criteria proposed above reflect both linguistic principles and practical considerations in corpus compilation. Frequency and dispersion ensure representativeness, while semantic stability and collocational productivity confirm that a word has become a functional lexical unit rather than a temporary phenomenon.

In the English-Uzbek bilingual context, several specific challenges arise. English neologisms often enter Uzbek through direct borrowing or adaptation, creating asymmetry in the corpus. Many Uzbek neologisms are English-derived but undergo phonological, morphological, and semantic adaptation. Corpus compilers must decide whether to treat these as foreign borrowings or as integrated Uzbek lexical items.

From an anthropocentric perspective, neologisms embody human creativity and the need to name new realities. Their inclusion in bilingual corpora documents how speakers of both languages respond to technological and social change. At the same time, overly inclusive

policies risk polluting corpora with ephemeral terms, while overly restrictive policies may render corpora outdated.

A balanced, evidence-based approach that combines quantitative measures with qualitative linguistic judgment appears most appropriate. Clear documentation of the decision-making process for each neologism enhances corpus transparency and reusability.

Conclusion. The integration of neologisms from English and Uzbek into bilingual corpora requires well-defined, multidimensional criteria. Frequency, semantic stability, collocational productivity, cross-linguistic relevance, authenticity, and temporal persistence together provide a robust framework for determining when a new word has earned lexical status.

Implementing these criteria will improve the quality, representativeness, and usefulness of English-Uzbek bilingual corpora for linguistic research, lexicography, machine translation, and language technology development. Future work should focus on developing standardized annotation protocols for neologisms and conducting longitudinal studies to better understand the life cycles of new words in both languages.

By establishing clear and principled guidelines, corpus linguistics can more effectively capture the dynamic nature of language while supporting both theoretical inquiry and practical applications in the digital age.

References:

1. Ibragimov J.Y., Formal tilshunoslikda soʻzlar birikuvi masalalarining rganilishi. Oʻzbekiston milliy axborot agentligi OʻZA elektron jurnal. 2022. 10-son. B. 108-112. (10.00.00. OAK rayosatining 2019-yil 28-fevraldagi 262/9.2-son qarori).
2. Ibragimov J.Y., Jahon tilshunosligida collocation atamasi va uning tadqiqi. FarDU. Ilmiy xabarlar. Far ona davlat universiteti ilmiy jurnali. 2022. 4-son. B. 123-126. 20)
3. Ergashev T.T., Localization of neologisms in the uzbek language - adoption and national characteristics of neologisms in the technological field. UZDJTU Xalqaro konferensiya 2025-yil 29-aprel. - Xalqaro ilmiy-amaliy anjuman 2025-yil 29-aprel, 2025-yil.
4. Ergashev T.T., Ingliz va oʻzbek tilidagi neologizmlarning semantik xususiyatlari. ATMU Respublika ilmiy-amaliy anjumani . - Qarshi 2025, 2025-yil.
5. Ergashev T.T., Ingliz va zbek tillaridagi neologizmlarni qiyosiy tipologik tahlili. The Lingua Spectrum Journal . -Volume: 4. April 2025, 2025-yil.