

INTEGRATION MECHANISMS: BORROWINGS, HYBRIDS, AND CALQUES IN  
NEOLOGISMS. 'LEXICOGRAPHIC GAP' A QUANTITATIVE ANALYSIS OF  
WORDS ABSENT FROM DICTIONARIES.

Madalova Mohiro'z Navruzovna

Master's student in English Linguistics at Termiz State University

@madalovamohiroz@gmail.com

@m\_navruzovna

**ABSTRACT**

This study investigates integration mechanisms in neologisms—borrowings, hybrids, and calques—and examines the “lexicographic gap.” A corpus of contemporary Uzbek texts from media and digital platforms was compiled and compared with existing dictionaries using quantitative analysis. The results show a notable proportion of frequently used words absent from dictionaries, with borrowings prevailing, followed by hybrids and calques. These findings demonstrate rapid lexical change and reveal limitations in current lexicographic coverage. The study concludes that systematic, data-driven updates are necessary to align dictionaries with evolving language use.

**Keywords:** neologisms, borrowings, hybrid formations, calques, lexicographic gap, corpus linguistics, quantitative analysis, lexicography, language integration, modern Uzbek language, lexical innovation, terminology.

**АННОТАЦИЯ**

Данное исследование рассматривает механизмы интеграции неологизмов — заимствования, гибридные образования и кальки — и анализирует «лексикографический разрыв». На основе корпусного подхода были изучены современные узбекские тексты из СМИ и цифровых платформ с последующим сопоставлением с существующими словарями посредством количественного анализа. Результаты показывают значительную долю часто употребляемых слов, отсутствующих в словарях, при доминировании заимствований, за которыми следуют гибриды и кальки. Полученные данные свидетельствуют о быстром лексическом развитии и указывают на необходимость систематического обновления словарей

**Ключевые слова:** неологизмы, заимствования, гибридные образования, кальки, лексикографический разрыв, корпусная лингвистика, количественный анализ, лексикография, языковая интеграция, современный узбекский язык, лексическое обновление, терминология.

**INTRODUCTION**

The rapid development of globalization, digital communication, and technological innovation has significantly accelerated lexical change in modern languages, including Uzbek. In this context, neologisms emerge as a dynamic layer of vocabulary reflecting ongoing social, cultural, and scientific transformations. These new lexical units enter the language through various integration mechanisms, most notably borrowings, hybrid formations, and calques, which together shape the evolving structure of the lexicon. Borrowings constitute one of the most productive sources of neologisms, particularly in domains such as information technology, business, and media. The increasing influence of English as a global lingua franca has intensified the influx of foreign lexical items into Uzbek, often with minimal phonetic or morphological



adaptation<sup>1</sup>. Alongside direct borrowings, hybrid formations—combinations of native and foreign elements—have become widespread, demonstrating the language’s flexibility in integrating external influences while maintaining internal structural patterns<sup>2</sup>. Similarly, calques, or loan translations, reflect a more conservative strategy, whereby foreign concepts are rendered using native linguistic resources, thus preserving semantic transparency.<sup>3</sup>

Despite the active incorporation of neologisms into everyday usage, a significant discrepancy persists between actual language use and its representation in lexicographic sources. This phenomenon, commonly referred to as the “lexicographic gap,” highlights the lag between lexical innovation and its documentation in dictionaries<sup>4</sup>. The existence of such gaps is particularly evident in rapidly changing communicative environments, such as social media and online discourse, where new terms are coined and disseminated at an unprecedented pace. The study of the lexicographic gap is closely linked to advances in corpus linguistics, which provide empirical methods for analyzing large-scale textual data. By comparing corpus-derived lexical items with entries in existing dictionaries, researchers can quantitatively assess the extent of undocumented vocabulary and identify patterns in neologism formation and usage<sup>5</sup>. This approach enables a more objective evaluation of lexical dynamics and offers valuable insights for lexicographic practice. The relevance of this research lies in its potential to bridge the gap between theoretical linguistics and practical lexicography. Understanding the mechanisms of neologism integration and quantifying lexicographic gaps not only contributes to the study of language change but also supports the development of more comprehensive and up-to-date dictionaries. Furthermore, such analysis is essential for language policy, standardization processes, and the preservation of linguistic identity in the face of global influences. Therefore, this study aims to investigate the integration mechanisms of neologisms in modern Uzbek and to conduct a quantitative analysis of lexicographic gaps by employing a corpus-based methodology. The findings are expected to provide a systematic account of lexical innovation and to highlight the need for adaptive and data-driven lexicographic strategies.

## LITERATURE REVIEW AND METHODOLOGY

Research on neologisms and how they are integrated has been extensively examined in contemporary linguistics, especially in the contexts of lexicology, sociolinguistics, and corpus linguistics. Initial foundational research by Haugen (1950) created a theoretical framework for examining linguistic borrowing, differentiating between importation and substitution.

---

<sup>1</sup> Haspelmath, M. (2009). *Lexical Borrowing: Concepts and Issues*.

<sup>2</sup> Haugen, E. (1950). *The Analysis of Linguistic Borrowing*.

<sup>3</sup> Vinay, J.-P., & Darbelnet, J. (1995). *Comparative Stylistics of French and English*.

<sup>4</sup> Svensén, B. (2009). *A Handbook of Lexicography*.

<sup>5</sup> McEnery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice*.



This model continues to be applicable in modern research, particularly in examining borrowings in Uzbek, where terms like internet (1990s), marketing (early 2000s), and startup (2010s) have integrated into the language alongside socio-economic transformations.<sup>6</sup>

Likewise, Crystal (2003) underscored English's function as a worldwide language, stressing its influence on vocabulary growth in non-English languages, particularly in Central Asian linguistic settings.

In the 21st century, researchers have progressively concentrated on hybrid formations and calques as effective means of lexical creativity. Haspelmath (2009) indicates that hybridization demonstrates linguistic efficiency and cultural adaptation, enabling speakers to effectively merge native and foreign components<sup>7</sup>.

Uzbek hybrid terms like onlayn-dars (commonly used since 2020 during the COVID-19 pandemic) and internet-do'kon (2010s) demonstrate the localization of global ideas. Calques, conversely, have been thoroughly examined in translation theory (Vinay & Darbelnet, 1995), where they are characterized as structural or semantic reproductions of foreign phrases. The idea of the "lexicographic gap" has become more significant in recent years, especially due to the progress of corpus-based studies.

Svensén (2009) describes this gap as the difference between real language use and how it's depicted in dictionaries, highlighting that conventional lexicographic methods usually do not keep up with swift lexical changes.

This problem is particularly noticeable in languages that are undergoing significant modernization, like Uzbek, where numerous commonly used neologisms exist. To address these challenges, modern research increasingly relies on corpus linguistics methodologies. McEnery and Hardie (2012) highlight the importance of large, representative text corpora for identifying frequency patterns and emerging lexical trends. In this study, a corpus of approximately 50,000 words was compiled from Uzbek online media, social networks, and news platforms published between 2015 and 2025.

The methodological framework consists of several stages. First, data collection was carried out by extracting texts from digital sources, ensuring diversity in genres such as news articles, blogs, and informal discourse. Second, lexical units were processed using frequency analysis to identify high-frequency items potentially representing neologisms. Third, these items were compared against existing Uzbek explanatory dictionaries published before 2020 to determine their lexicographic status. Words absent from these dictionaries were classified as elements of the lexicographic gap.

Furthermore, the identified neologisms were categorized into three groups: borrowings, hybrids, and calques. Quantitative analysis was then applied to determine their relative proportions. Preliminary findings indicate that borrowings account for approximately 55–60% of the identified neologisms, hybrids for 25–30%, and calques for 15–20%. These results align with global linguistic trends, where direct borrowing remains the dominant strategy in lexical expansion.

---

**Haugen (1950):** Refers to Einar Haugen's foundational linguistic work, "*The Analysis of Linguistic Borrowing*", published in the journal *Language* (Vol. 26, No. 2).

<sup>7</sup> **Haspelmath (2009):** Refers to Martin Haspelmath's research on loanwords and linguistic borrowing, prominently featured in the Loanword Typology Project and the book *Loanwords in the World's Languages: A Comparative Handbook*.



## RESULTS

The empirical analysis of the 50,000-word corpus, extracted from Uzbek digital media and social networks between 2015 and 2025, revealed a total of 342 distinct lexical units functioning as neologisms. Quantitative processing of these units demonstrated a clear hierarchy in the integration strategies of new concepts. The primary findings indicate that **direct borrowings** constitute the largest category, accounting for exactly 58.2% (199 units) of the identified dataset. This group is heavily concentrated in domains such as digital technology and entrepreneurship, featuring frequently used terms such as *bloger*, *vlog*, and *podkast*. The second most frequent category is **hybrid formations**, which represent 26.3% (90 units) of the total extracted neologisms. High-frequency units in this group include structurally combined terms like *onlayn-dars* and *internet-do'kon*. Conversely, **calques** (loan translations) yielded the smallest proportion, comprising only 15.5% (53 units) of the data. Prominent examples in this segment include structural replications like *sun'iy intellekt* and *ijtimoiy tarmoq*. Furthermore, the study identified a significant negative result regarding lexicographic documentation. When comparing the 342 identified neologisms against official Uzbek explanatory dictionaries published before 2020, 87.4% of these active lexical units were completely absent. This quantifiable gap confirms a substantial discrepancy between real-time digital communication and standardized academic records.

## DISCUSSION

The findings of this study demonstrate that direct borrowings are the absolute dominant strategy in adopting new concepts in modern Uzbek, accounting for 55–60% of identified neologisms. This strong preference aligns perfectly with the principle of linguistic economy proposed by Martin Haspelmath (2009)<sup>8</sup>, suggesting that native speakers favor adopting ready-made international forms over creating native equivalents. For example, terms like *bloger*, *vlog*, and *podkast* have transitioned into the Uzbek lexicon without undergoing any major phonetic or morphological alterations. The sheer dominance of direct borrowings highlights just how rapidly internet access and digital culture have permeated Uzbek society. The second largest group is represented by hybrid formations, making up 25–30% of the recorded data. This phenomenon highlights the language's natural defense mechanism and adaptability. While the Uzbek language actively accepts foreign root words, it seamlessly merges them with its internal grammatical and lexical elements. For instance, in localized terms like *onlayn-dars* and *internet-do'kon*, the foreign elements are directly combined with native Uzbek words like "dars" (lesson) and "do'kon" (store). These hybrids provide a balanced way to localize global concepts according to the structural rules of the Uzbek language, making them much easier for native speakers to understand in everyday speech. Conversely, calques (loan translations) make up the smallest share of the extracted corpus at only 15–20%. Prominent examples include compound terms like *sun'iy intellekt* (artificial intelligence) and *ijtimoiy tarmoq* (social network). Because calques require translating a foreign expression word-for-word, they force the language to assign fresh semantic meanings to pre-existing native words. The relatively low frequency of calques suggests that the current stream of global information is moving too fast for speakers to actively search for exact translations or coin accurate localized calques.

- 
- <sup>8</sup> Haspelmath, M., & Tadmor, U. (2009). *Loanwords in the World's Languages: A Comparative Handbook*. De Gruyter Mouton.



In conclusion, as noted by Svensén (2009)<sup>9</sup>, a prominent "lexicographic gap" has formed between actual spoken use and formal documentation. Academic dictionaries are failing to keep pace with this rapid evolution. Therefore, integrating corpus linguistics directly into modern lexicography is absolutely necessary. Ultimately, Uzbek is not behaving as a passive recipient of foreign loanwords, but rather as an active, dynamic system capable of processing global concepts through its own structural patterns.

## CONCLUSION

The comprehensive analysis of the 50,000-word corpus extracted from Uzbek digital media platforms between 2015 and 2025 provides critical insights into the mechanisms of neologization in the modern Uzbek language. The empirical results of this study clearly indicate that the integration of new concepts is a highly dynamic process, heavily influenced by the rapid pace of global communication and digital transformation. Quantitative data processing revealed that direct borrowings are the absolute dominant strategy for lexical expansion, accounting for 58.2% of the identified neologisms. This finding confirms the principle of linguistic economy, demonstrating that native speakers actively prefer adopting ready-made international forms in highly specialized and rapidly evolving domains such as digital technology, social media, and modern entrepreneurship.

Furthermore, the significant proportion of hybrid formations, standing at 26.3%, reveals the language's natural defense mechanism and structural adaptability. By combining foreign elements with native roots, the Uzbek language demonstrates a strong capacity for localizing global concepts while maintaining its core grammatical integrity. On the other hand, the minimal percentage of calques, yielding only 15.5%, suggests that current information flows are moving too fast for speakers to engage in continuous translation processes or assign new semantic meanings to existing native words. Perhaps the most critical outcome of this research is the quantifiable measurement of the lexicographic gap. The fact that 87.4% of the identified active neologisms are completely undocumented in official explanatory dictionaries published before 2020 confirms that traditional lexicographic practices are failing to keep pace with rapid linguistic changes. This discrepancy highlights the urgent need for a systematic paradigm shift in how dictionaries are compiled and updated.

In conclusion, this study proves that the Uzbek language is not merely a passive recipient of external linguistic influence, but an active, adaptive system capable of processing global concepts through diverse structural mechanisms. To prevent further academic marginalization of everyday digital communication, future research must prioritize corpus linguistics and continuous data monitoring. Integrating advanced computational tools and corpus-based methodologies into standardized lexicography is no longer just an alternative, but an absolute necessity for preserving and effectively documenting the continuous evolution of the modern Uzbek lexicon.

- 
1. <sup>9</sup> Svensén, Bo. (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press.



REFERENCES

1. **Crystal, D.** (2003). *English as a Global Language* (2nd ed.). Cambridge University Press.
2. **Haspelmath, M.** (2009). *Loanwords in the World's Languages: A Comparative Handbook*. De Gruyter Mouton.
3. **Haugen, E.** (1950). The analysis of linguistic borrowing. *Language*, 26(2), 210–231.
4. **McEnery, T., & Hardie, A.** (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press.
5. **Svensén, B.** (2009). *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge University Press.
6. **Vinay, J.-P., & Darbelnet, J.** (1995). *Comparative Stylistics of French and English: A Methodology for Translation*. John Benjamins Publishing Company.

