pISSN:2349-5707 Volume:13,Issue 2, February -2026 eISSN:2349-5715

PROCESSES OF LEXICAL RENEWAL IN UZBEK-LANGUAGE INTERNET DISCOURSE: A CORPUS-BASED STUDY OF NEOLOGISMS AND BORROWED VOCABULARY

Sherkuziyeva Azizakhon Alijon kizi
Uzbekistan, Fergana Region, Uchkuprik District

Annotation: The article examines mechanisms of lexical renewal in Uzbek-language internet discourse based on corpus and web-corpus data. The object of analysis comprises neologisms and borrowings (primarily Anglicisms) functioning in digital communication. The study proposes an operational corpus-based framework for identifying lexical innovations (frequency and contextual analysis, register comparison, collocational profiling, and markers of morphological adaptation). The main types of innovations are described (direct borrowings, hybrid derivatives, calques, semantic neologisms, and graphic variants), and characteristic patterns of integration of borrowed vocabulary into the agglutinative morphology of Uzbek are demonstrated. The methodological framework relies on standard corpus-linguistic procedures (KWIC concordance, frequency lists, and collocation analysis). The results show that internet discourse constitutes a highly productive environment for the accelerated consolidation of lexical innovations, while the degree of their "naturalization" can be reliably diagnosed through morphological and syntactic evidence observable in corpus contexts.

Keywords: internet discourse, Uzbek language, neologisms, borrowings, Anglicisms, corpus linguistics, KWIC, collocations.

Introduction

Digital communication (social networks, messengers, news websites, and user communities) creates a specific environment of accelerated linguistic change: new realities are named faster than they can be codified in normative dictionaries, and competing designations struggle for usage dominance. In Uzbek, this tendency is particularly evident in technological, media-related, and everyday online communication, where a strong influx of borrowings, hybrid formations, and semantic shifts can be observed.

A corpus-based perspective is crucial for two reasons. First, it allows the researcher to move beyond isolated anecdotal examples toward verifiable data—frequencies, register distributions, contextual profiles, and collocations. Second, corpus methods provide formal criteria for identifying innovation: frequency growth, thematic specialization, orthographic instability, and variability of morphological adaptation. General principles of corpus analysis (concordance, representativeness, register comparability) are thoroughly discussed in foundational works in corpus linguistics.

The empirical basis of the present study includes:

1. large-scale Uzbek web corpora (e.g., Northern Uzbek Web 2019, 87,531,885 tokens), which provide sufficient data for identifying stable innovations;
2. national corpus platforms offering token/lemma search and KWIC concordance tools;
3. methodological descriptions of the structure and annotation principles of the Uzbek National Corpus.

Aim of the study is to develop an academically rigorous and reproducible corpus-based approach to the study of lexical innovations in Uzbek-language internet discourse and to describe the major types of neologisms and borrowings and their patterns of adaptation.

Objectives:

- to operationalize the concepts of *neologism* and *borrowing* for corpus research;
- to describe a methodology for extracting and validating innovative lexical items;
- to classify major types of innovations and identify their structural properties;

pISSN:2349-5707 Volume:13,Issue 2, February -2026 eISSN:2349-5715

- to demonstrate typical mechanisms of morphological integration of borrowings.

Materials and Methods

Materials

(A) Uzbek Web Corpora.

Web corpora are particularly relevant for internet discourse, as they reflect current lexical practices not always represented in traditional print-based registers. As a representative resource, the Northern Uzbek Web 2019 corpus (87.5 million tokens; genre: web texts) is used.

(B) National Corpus Platforms.

Uzbek corpus infrastructures provide token- and lemma-based search as well as KWIC concordances, enabling systematic contextual analysis.

(C) Corpus Methodology Descriptions.

Studies devoted to the structure of the Uzbek National Corpus outline its core components (text base, annotation layers, metadata, user interface, and KWIC tools) and justify its applicability for linguistic research.

Operational Definitions

- **Neologism** is defined as a lexical unit (a new form or a new meaning) exhibiting signs of novelty in the linguistic system or usage and not yet fully codified normatively.
- **Borrowing** is a lexical unit adopted from a donor language (most frequently English in the digital domain) and functioning in Uzbek either in its original or in an adapted form.

From a corpus-analytical perspective, a unit is treated as an *innovation candidate* if it (i) is thematically associated with digital practices, (ii) shows orthographic or morphological variability, and/or (iii) demonstrates a strong internet-register profile (higher relative frequency in web-based discourse compared to more formal registers).

Analytical Procedure

1. **Identification of candidates:**
   - thematic filtering (social media, messengers, streaming, content production, marketing, cybersecurity);
   - formal markers (Latin-script insertions, mixed orthography, non-native consonant clusters, abbreviations);
   - derivational evidence (attachment of Uzbek affixes to foreign stems).
2. **Contextual verification (KWIC):**
   - interpretation of meaning through immediate co-text;
   - exclusion of noise (proper names, brands, URLs);
   - identification of recurrent syntactic frames.
3. **Collocational profiling:**
   - stable combinations (e.g., “X qilmoq”, “X bo‘lmoq”);
   - distribution across semantic domains.
4. **Assessment of integration degree** (“insertion → adaptation → conventionalization”):
   - orthographic stabilization;
   - derivational productivity;
   - normalization of combinatorial behavior.

This combination of frequency, context, and collocation analysis corresponds to standard corpus-linguistic methodology.

Results

1. Types of Lexical Innovations in Uzbek Internet Discourse

Corpus analysis reveals several dominant classes of innovations:

1. **Direct borrowings (minimal adaptation).**

pISSN:2349-5707 Volume:13,Issue 2, February -2026 eISSN:2349-5715

Examples typical of interface and media vocabulary include *online/onlayn*, *blog*, *post*, *story*, *stream*, *podcast*, *trend*, *content*, *fake*, *like*.

These items often function as uninflected insertions or show limited orthographic adaptation.

2. Adapted borrowings (phonographic and morphological integration).

Such items exhibit Uzbek-oriented spelling (e.g., *onlayn*) and participate in regular morphological processes.

3. Hybrid derivatives (foreign stem + Uzbek affixes or function words).

This group provides strong evidence of integration into Uzbek agglutinative morphology. Typical models include:

- **[stem] + -lash / -lashmoq:** *layklashmoq* 'to like', *skrinshotlash* 'to take screenshots';
- **[stem] + -chi** (agentive): *blogerchi* (competing with *bloger*);
- **analytic constructions with *qilmoq*:** *post qilmoq*, *repost qilmoq*, *skrinshot qilmoq*, *strim qilmoq*;
- **hybrid noun phrases:** *kontent rejasi*, *SMM mutaxassisi*, *onlayn dars*.

Regular affixation and participation in standard syntactic patterns indicate that such units are no longer foreign insertions but elements of Uzbek lexical structure.

4. Calques and semi-calques.

These include literal or partial translations of English multiword expressions and semantic replication of foreign collocational patterns.

5. Semantic neologisms.

Existing Uzbek words acquire new meanings in digital contexts (e.g., *kanal*, *platforma*, *tarmoq*, *sahifa*). These shifts are detected through changes in collocational behavior rather than form.

2. Mechanisms of Morphological Adaptation

Two main strategies dominate:

(A) Analytic verbalization via *qilmoq*.

This highly productive pattern allows borrowed nouns to function as predicates (*post qilmoq*, *repost qilmoq*). It is grammatically transparent and easily extendable.

(B) Suffixal derivation.

Verbal and nominal derivation using Uzbek affixes (*-lash*, *-lan*, etc.) integrates borrowings into word-formation networks, enabling further grammatical inflection.

3. Internet Discourse as an Accelerator of Innovation

The digital environment is characterized by:

- high repetition of formulaic patterns;
- strong dependence on global platform languages;
- rapid circulation of trends and memes, producing waves of neologization.

These observations align with existing research on internet discourse as a primary source of lexical innovation.

Discussion

1. Advantages of the Corpus Approach

Unlike impressionistic lists of neologisms, corpus analysis enables:

- differentiation between stable innovations and occasionalisms;
- tracing of orthographic competition and stabilization;
- measurement of register specificity.

pISSN:2349-5707 Volume:13,Issue 2, February -2026 eISSN:2349-5715

2. Norm vs. Usage

Corpus evidence helps neutralize ideological debates about borrowings. Units exhibiting morphological adaptation, stable collocations, and cross-register diffusion can be objectively identified as part of actual usage, regardless of prescriptive attitudes.

3. Cross-Linguistic Parallels

Similar processes have been documented in corpus-based studies of other languages, including Russian internet discourse, confirming the general role of digital communication as an accelerator of lexical change.

Conclusion

Corpus-based analysis of Uzbek-language internet discourse demonstrates that lexical renewal follows stable patterns. Direct borrowings and graphic variants form an initial layer of innovation, from which some units undergo morphological and combinatorial integration and become conventionalized. Internet discourse functions as a key domain of contemporary lexical expansion, while corpus methodology provides a reliable empirical basis for documenting and explaining this process.

The proposed framework is reproducible and may be extended through diachronic comparison of web corpora, register-based analysis, and automated extraction of neologisms.

Notes

1. Corpus linguistics is understood here as computer-assisted analysis of large collections of authentic texts using concordance, frequency, and collocation procedures.
2. Internet discourse is recognized as a productive source of neologisms requiring contextual interpretation.
3. Micro-diachronic corpus studies demonstrate similar mechanisms of rapid lexical change in online environments.

References.

1. McEnery, T., & Hardie, A. *Corpus Linguistics: Method, Theory and Practice.* Cambridge University Press, 2012.
2. Biber, D., Conrad, S., & Reppen, R. *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge University Press, 1998.
3. Sinclair, J. *Corpus, Concordance, Collocation.* Oxford University Press, 1991.
4. Primova, M. H. *The Structure of the Uzbek National Corpus: Core Components and Information Types.* Proceedings of the 5th International Conference “Computational Linguistics: Problems, Solutions, Prospects”, 2025.
5. Leipzig Corpora Collection: Northern Uzbek Web 2019 Dataset.
6. Irgasheva, S. *Translation Strategies for English Neologisms in Internet Discourse.* Translation Forum, 2022.