

LEXICAL INNOVATION IN DIGITAL DISCOURSE: A CORPUS-BASED STUDY OF NEOLOGISMS ON SOCIAL MEDIA PLATFORMS

Karimbayeva A.Sh.

Teacher of the English Methodology Department,
Uzbekistan State World Languages University.

<https://doi.org/10.5281/zenodo.15192175>

Abstract. *Digital communication has radically transformed language use, with social media platforms acting as catalysts for lexical innovation. This study investigates neologisms emerging in English digital discourse on Twitter, TikTok, Reddit, and Instagram. Using a corpus-based approach, we compiled a 120,000-word sample to identify and classify 83 newly coined lexical items by their morphological features and semantic fields. Our findings reveal the significant role of digital environments in fostering linguistic creativity and driving the gradual standardization of neologisms. The implications for English lexicology, sociolinguistics, and digital linguistics are discussed.*

Keywords: *neologisms, lexical innovation, digital discourse, corpus linguistics, social media, word formation.*

ЛЕКСИЧЕСКИЕ ИННОВАЦИИ В ЦИФРОВОМ ДИСКУРСЕ: КОРПУСНОЕ ИССЛЕДОВАНИЕ НЕОЛОГИЗМОВ НА ПЛАТФОРМАХ СОЦИАЛЬНЫХ СЕТЕЙ

Аннотация. *Цифровая коммуникация радикально изменила использование языка, а платформы социальных сетей выступили в качестве катализаторов лексических инноваций. В этом исследовании изучаются неологизмы, появляющиеся в английском цифровом дискурсе в Twitter, TikTok, Reddit и Instagram. Используя корпусный подход, мы составили выборку из 120 000 слов, чтобы идентифицировать и классифицировать 83 недавно созданных лексических элемента по их морфологическим признакам и семантическим полям. Наши результаты показывают значительную роль цифровой среды в содействии языковому творчеству и постепенной стандартизации неологизмов. Обсуждаются последствия для английской лексикологии, социолингвистики и цифровой лингвистики.*

Ключевые слова: *неологизмы, лексические инновации, цифровой дискурс, корпусная лингвистика, социальные сети, словообразование.*

Introduction

Language is inherently dynamic, continuously adapting to cultural, technological, and social changes. In the digital era, social media platforms have emerged as epicenters for rapid linguistic innovation. Unlike traditional language evolution, which unfolds over extended periods, digital discourse generates new lexical items almost instantaneously. These neologisms often encapsulate the zeitgeist of online communities, reflecting evolving social identities, humor, and socio-political commentaries.

Recent studies have shown that user-driven innovations in online spaces are not merely ephemeral; many of these lexical items eventually find their way into standard language repositories such as dictionaries. However, the processes underlying the emergence, dissemination, and potential standardization of these neologisms remain underexplored, particularly from a morphological and semantic perspective.

The present study fills this gap by employing a corpus-based analysis to examine neologisms in digital discourse. Focusing on English-language content from Twitter, TikTok, Reddit, and Instagram, this research addresses the following questions: What are the predominant mechanisms of word formation in digital neologisms? Which semantic fields are most frequently represented by these new lexical items? How does digital discourse contribute to the transition of neologisms from informal jargon to standardized language? In answering these questions, the paper contributes to a deeper understanding of contemporary lexical innovation and its implications for English lexicology and sociolinguistics.

Methods

A digital corpus was constructed using publicly available posts written in English from Twitter, TikTok, Reddit, and Instagram over a six-month period (August 2024 – February 2025).

The inclusion criteria ensured that only original user-generated content was used, filtering out reposts, advertisements, and bot-generated messages, while covering a variety of topics such as lifestyle, technology, mental health, relationships, and pop culture. The final corpus consisted of approximately 120,000 words. Data extraction involved several digital tools. Tweets were collected using TweetNLP tools and the Twitter API; Reddit threads were gathered via PRAW (Python Reddit API Wrapper); TikTok and Instagram posts were manually collected from trending public profiles, focusing on captions, comments, and hashtags. Text analysis was performed with corpus linguistics software such as AntConc and Sketch Engine, which enabled frequency analysis, keyword extraction, and dictionary cross-referencing.

To identify neologisms, a multistep procedure was implemented. First, keyword frequency analysis was carried out to detect low-frequency terms likely representing novel lexical items. Next, candidate neologisms were cross-referenced with established dictionaries (Oxford English Dictionary, Merriam-Webster) to exclude already institutionalized words. Two independent linguists then manually evaluated each term to confirm its novelty based on morphological innovation and semantic shift. Confirmed neologisms were classified according to their word-formation mechanism (blending, clipping, affixation, acronyms/initialisms, and semantic shift).

All data were anonymized and sourced from public content, and no private or sensitive information was used. The research was conducted in accordance with ethical guidelines for studies involving online data.

Results

Analysis of the corpus yielded 83 unique neologisms, which were distributed across several word formation types. Blended words, which merge parts of two or more words (e.g., “situationship,” a blend of “situation” and “relationship”; “bromance,” combining “brother” and “romance”), comprised approximately 31% of the identified neologisms. Clipping, the shortening of longer words (e.g., “sus” from “suspicious”; “delulu” from “delusional”), represented about 17% of the total. Affixation, involving the addition of prefixes or suffixes (e.g., “doomscrolling,” the act of endlessly scrolling through negative news; “unfriend,” to remove someone from a social network), accounted for roughly 22%. Acronyms and initialisms (e.g., “FOMO” for “Fear Of Missing Out”; “YOLO” for “You Only Live Once”) made up around 13% of the neologisms, while cases of semantic shift (e.g., “ghosting,” which now describes the act of abruptly ceasing all communication in personal relationships; “clickbait,” originally a neutral term now associated with misleading headlines) contributed approximately 17%. The analysis further revealed that these neologisms clustered in several semantic fields.

Terms related to identity and relationships (e.g., “situationship” and “ghosting”) reflected changing interpersonal dynamics. A significant number of neologisms expressed contemporary emotional experiences and mental health challenges (e.g., “doomscrolling” and “delulu”).

Others commented on technological influences and media consumption (e.g., “algorithmic rage” and “TikTok brain”), while some served as tools for sociocultural commentary (e.g., variations of “girlboss”).

This distribution indicates that digital discourse not only generates new words for practical communication but also creates precise labels for emerging social phenomena.

Discussion

The findings of this study underscore the transformative impact of digital media on the English lexicon. Social media platforms have democratized the process of word formation, enabling users from diverse backgrounds to contribute to lexical innovation. The rapid creation and dissemination of neologisms through digital channels challenge traditional lexicographical processes and highlight the fluid nature of contemporary language.

Many of the neologisms identified in this study began as informal, context-specific expressions but have the potential to become standardized over time. For instance, terms such as “ghosting” and “doomscrolling” have already made their way into reputable dictionaries, illustrating the gradual process of lexical institutionalization. This trend emphasizes the need for lexicologists to continuously monitor digital discourse as a significant source of language change.

From a sociolinguistic perspective, neologisms serve multiple functions in digital communication. They facilitate identity formation by allowing speakers to express nuanced aspects of their social and interpersonal lives. The use of specific neologisms can also act as a marker of in-group identity, reinforcing social bonds within digital communities. Moreover, many of the identified lexical innovations encapsulate modern emotional experiences and collective sentiments, thus providing valuable insight into the psychological landscape of contemporary society.

The technological infrastructure of social media further accelerates the spread of new words. Limited character spaces, visual elements, and algorithm-driven content delivery all contribute to the rapid propagation of neologisms. The ephemeral nature of digital content means that while some neologisms may fade quickly, others gain traction and eventually enter standard language. This dynamic cycle of creation, adaptation, and possible obsolescence challenges traditional views on language evolution and calls for new approaches in lexicographic research.

Despite its contributions, this study has certain limitations.

The corpus represents a snapshot of digital discourse over six months, and longer-term studies might reveal additional patterns or changes in neologism usage. Additionally, while the selected platforms offer a diverse range of content, they may not capture the full spectrum of

digital communication, particularly in less mainstream or regional online communities. Finally, the manual annotation process, while rigorous, inevitably involves a degree of subjectivity.

Future research could benefit from advanced natural language processing techniques to further refine the identification and classification of neologisms.

Looking forward, cross-linguistic comparisons and studies that incorporate additional digital platforms would be valuable. Understanding how similar lexical innovations occur in other languages could provide broader insights into the global impact of digital media on language. Moreover, exploring the sociocultural implications of neologisms—such as their role in identity construction and group dynamics—could deepen our understanding of modern communication practices.

Conclusion

This study examined lexical innovation in digital discourse by analyzing neologisms emerging from English-language social media. Through a corpus-based approach, 83 unique neologisms were identified and classified by their word formation mechanisms. The findings highlight the crucial role of digital platforms in fostering rapid linguistic creativity and demonstrate how new lexical items serve to label emerging social phenomena. The study emphasizes that digital neologisms, while often informal and transient at first, have the potential to achieve standardization and contribute significantly to the evolution of the English language.

Future research should continue to explore the interplay between digital communication and language change, considering broader datasets and cross-cultural perspectives.

REFERENCES

1. Algeo, J. (1991). *Fifty Years among the New Words: A Dictionary of Neologisms*. Cambridge University Press.
2. Crystal, D. (2006). *Language and the Internet*. Cambridge University Press.
3. Tagliamonte, S. (2016). *Teen Talk: The Language of Adolescents*. Cambridge University Press.
4. Oxford English Dictionary. (2024). [Online resource].
5. Sketch Engine. (2023). [Corpus tool for linguistic analysis].
6. Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of Lexical Change in Social Media. *PLOS ONE*, 9(11), e113114.