



USING CORPORA TO STUDY NEOLOGISMS AND LEXICAL INNOVATION

Sindorova Muxlisa

Qushaqova Durдона

Abdullajonova Hakima

Jizzakh State Pedagogical University

Corresponding author email: turayevadurdona89@gmail.com

<https://doi.org/10.5281/zenodo.17971502>

Abstract

This article explores the role of corpus linguistics in the study of neologisms and lexical innovation in contemporary English. As language continuously adapts to social, technological, and cultural change, new words and meanings emerge to meet evolving communicative needs. Traditional methods of studying lexical change often rely on dictionaries and intuition, which may not accurately reflect real language use. In contrast, corpus-based approaches provide empirical evidence drawn from large collections of authentic texts. The study discusses how corpora are used to identify neologisms, analyse their frequency, contextual usage, and patterns of dissemination across different registers and time periods. It also examines how corpus data contributes to the analysis of broader lexical innovations, including semantic change and productive word-formation processes. Finally, the article highlights the advantages and limitations of corpus-based methods, arguing that despite certain challenges, corpora remain an indispensable tool for understanding lexical development in modern linguistics.

Keywords

corpus linguistics; neologisms; lexical innovation; language change; vocabulary development; English linguistics

Introduction

Language is not a fixed system; it continuously changes in response to social, technological, and cultural developments. One of the most visible signs of linguistic change is the appearance of new words and expressions, commonly referred to as neologisms. Alongside these, languages also experience lexical innovation, which includes new meanings of existing words, novel word-formation patterns, and creative combinations of lexical items. In recent decades, corpus linguistics has become an essential methodological approach for studying these phenomena. By analysing large collections of authentic texts, corpora provide empirical evidence that allows linguists to observe how new lexical items emerge, spread, and become established in real language use. This article aims to examine how linguistic corpora are used to study neologisms





and lexical innovation. It discusses the theoretical background of neologisms, explains the role of corpora in linguistic research, and highlights the advantages and limitations of corpus-based approaches in analysing lexical change.

Neologisms and Lexical Innovation: A Theoretical Overview

Neologisms are generally defined as newly created lexical items or existing words that have acquired new meanings within a relatively short period of time. They often arise in response to innovations in technology, politics, science, and popular culture. Words such as selfie, hashtag, and streaming illustrate how new concepts require new linguistic forms.

Lexical innovation, however, is a broader concept than neologism alone. It includes semantic change, such as when an existing word develops a new meaning, as well as morphological innovation, for example through productive affixes or compounding. From a linguistic perspective, studying lexical innovation helps explain how languages expand their expressive capacity and adapt to changing communicative needs.

Traditional approaches to studying neologisms relied heavily on dictionaries, literary texts, and linguistic intuition. While useful, these methods often fail to capture real-time language change and everyday usage. This limitation has led to the growing importance of corpus-based research.

What Is a Linguistic Corpus?

A linguistic corpus is a large, structured, and electronically stored collection of authentic texts. These texts may include written sources such as newspapers, academic articles, blogs, and fiction, as well as spoken data like interviews and conversations. Corpora are usually annotated and searchable, allowing researchers to analyse frequency, collocation, grammatical patterns, and contextual usage.

One important distinction is between synchronic corpora, which represent language use at a specific point in time, and diachronic corpora, which include texts from different historical periods. Diachronic corpora are particularly valuable for studying neologisms because they make it possible to trace the emergence and development of new words over time.

Identifying Neologisms Through Corpora

One of the primary uses of corpora in lexical studies is the identification of neologisms. By comparing frequency data across different time periods, researchers can detect words that suddenly increase in usage or appear for the first time. For instance, words related to digital communication often show a rapid rise in frequency in contemporary corpora.





Corpora also allow linguists to examine the contexts in which neologisms occur. Concordance lines reveal how a new word is used in sentences, what words it commonly co-occurs with, and whether its meaning is stable or still developing. This contextual analysis is crucial for distinguishing between temporary buzzwords and lexical items that are likely to become permanent parts of the language. Furthermore, corpus data can show variation in usage across genres and registers. A neologism may first appear in informal online communication before spreading to newspapers and academic writing. Such patterns are difficult to observe without large-scale corpus evidence.

Analysing Lexical Innovation Using Corpus Data

Beyond identifying new words, corpora are highly effective in analysing broader patterns of lexical innovation. Corpus-based studies can reveal productive word-formation processes, such as the increasing use of prefixes like eco- or cyber-. By analysing large datasets, linguists can determine which patterns are genuinely productive rather than based on isolated examples.

Semantic innovation can also be studied through corpora. When a word develops a new meaning, corpus analysis helps identify the transitional stages of that change. For example, shifts in collocational patterns often indicate semantic expansion or narrowing. This empirical approach provides stronger evidence than intuition-based analysis.

Another important aspect is the global spread of lexical innovations. English, as a

global language, absorbs new words from different cultures and regions. Corpora that include international varieties of English allow researchers to compare how innovations emerge and spread across different speech communities.

Advantages and Limitations of Corpus-Based Approaches

The main advantage of using corpora to study neologisms and lexical innovation is objectivity. Corpus data is based on real language use, which reduces reliance on subjective judgments. In addition, the large size of corpora allows researchers to analyse trends and patterns that would be invisible in smaller datasets.

However, corpus-based approaches also have limitations. Corpora may lag behind real-time language change, especially when it comes to very recent neologisms. Informal spoken language and emerging online slang may be underrepresented in some corpora. Moreover, corpus data alone cannot fully explain the social





motivations behind lexical innovation, which often require qualitative and sociolinguistic analysis.

Conclusion

In conclusion, corpora play a crucial role in the study of neologisms and lexical innovation. They provide empirical, large-scale evidence of how new words and meanings emerge and develop in real language use. By enabling detailed analysis of frequency, context, and variation, corpora offer insights that traditional methods cannot provide. Despite certain limitations, corpus-based approaches remain one of the most effective tools for understanding lexical change in modern linguistics. As language continues to evolve rapidly in response to global and technological influences, the importance of corpus linguistics in studying lexical innovation is likely to increase.

References

- Baker, P. (2006). Using corpora in discourse analysis. London: Continuum.
- Bauer, L. (2001). Morphological productivity. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CB09780511486210>
- Crystal, D. (2003). The Cambridge encyclopedia of the English language (2nd ed.). Cambridge: Cambridge University Press.
- Hoffmann, S., Evert, S., Smith, N., Lee, D., & Berglund Prytz, Y. (2008). Corpus linguistics with BNCweb – A practical guide. Frankfurt am Main: Peter Lang.
- Kerremans, D. (2015). A web of new words: A corpus-based study of the conventionalization process of English neologisms. Frankfurt am Main: Peter Lang.
- McEnery, T., & Hardie, A. (2012). Corpus linguistics: Method, theory and practice. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/CB09780511981395>
- Renouf, A. (2007). Tracing lexical productivity and creativity in the British media: The role of corpus data. In P. Skandera (Ed.), *Phraseology and culture in English* (pp. 61–79). Berlin: Mouton de Gruyter.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

