HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

ENGLISH NEOLOGISMS AND THEIR LEXICOGRAPHIC REPRESENTATION: CRITERIA, STAGES, AND CHALLENGES OF DICTIONARY INCLUSION

Faleeva Yekaterina
Independent researcher of
Samarkand State Institute of Foreign Languages Samarkand, Uzbekistan.
https://orcid.org/0009-0003-1998-1451
faleyevakatya9393@mail.ru

Abstract. This study investigates the criteria, stages, and challenges involved in the lexicographic representation of English neologisms. The research aims to identify the principal factors influencing dictionary inclusion and to examine how contemporary lexicographic practices respond to the unprecedented speed of lexical innovation. Drawing upon corpus linguistics, lexicographic theory, and recent developments in digital language monitoring, the study explores the processes through which neologisms are evaluated, codified, and incorporated into authoritative dictionaries. The article argues that contemporary lexicographic decision-making increasingly relies on corpus-based evidence, frequency analysis, semantic stability.

Keywords: neologism, lexical innovation, lexicography, dictionary inclusion, corpus linguistics, lexical codification, digital discourse, English language change.

Introduction. Language is a living and constantly evolving system that reflects changes occurring within society, culture, technology, and human cognition. The twenty-first century has witnessed unprecedented rates of lexical creation resulting from globalization, technological innovation, digital communication, and the expansion of social networking platforms. New words can emerge, spread, and become widely recognized within a matter of days, reaching global audiences through online communication channels. Consequently, traditional models of lexical diffusion, which often assumed gradual adoption over extended periods, require reconsideration in light of modern communicative practices.

The increasing frequency of lexical innovation has generated significant interest among linguists, sociolinguists, corpus linguists, and lexicographers. As language users continuously create and adapt lexical forms to express novel concepts, lexicographers face the challenging task of determining which innovations warrant inclusion in dictionaries. This issue is particularly important because dictionaries continue to function as authoritative records of linguistic usage and serve as influential instruments of language standardization and codification. The inclusion of a neologism in a dictionary is commonly perceived as evidence of its acceptance within the speech community.

Research methods. The present study employs a mixed-methods research design integrating quantitative corpus analysis with qualitative lexicographic examination. Such an approach is particularly appropriate for investigating contemporary neologisms because lexical innovation represents a multidimensional phenomenon involving linguistic, social, technological, and cultural factors. While quantitative methods facilitate the identification of patterns related to frequency, distribution, and lexical diffusion, qualitative analysis enables a deeper examination of semantic development, contextual variation, and lexicographic treatment. The research adopts a descriptive-analytical framework aimed at investigating the mechanisms through which English neologisms progress from initial usage to lexicographic codification.

The study utilizes data derived from four principal categories of sources: General English corpora; News and media corpora; Social media and online communication datasets; Contemporary

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

English dictionaries and lexicographic databases. The primary corpus sources include: The Corpus of Contemporary American English (COCA); The News on the Web Corpus (NOW); The British National Corpus (BNC); The Global Web-Based English Corpus (GloWbE).

Results. The corpus and lexicographic analyses demonstrate that the inclusion of English neologisms in contemporary dictionaries is governed by a combination of quantitative and qualitative factors rather than by frequency alone. Although corpus frequency remains one of the most important indicators of lexical relevance, the findings reveal that dictionary editors evaluate additional variables, including semantic stability, distribution across communicative domains, longevity, and sociocultural significance.

The role of frequency in lexicographic decision-making has been extensively discussed in corpus-based lexicography. According to Kilgarriff, frequency data provide an empirical basis for identifying lexical trends and evaluating the significance of emerging vocabulary; however, frequency should not be interpreted independently of contextual and semantic evidence [Kilgarriff 1997: 135–155]. Similarly, Barnhart argues that successful neologisms must demonstrate not only measurable occurrence but also functional value within a speech community [Barnhart 2007: 132–138].

Corpus data indicate a strong positive relationship between sustained frequency growth and dictionary inclusion. Lexical items that maintained increasing frequency over a period of at least three years were significantly more likely to achieve codification than those characterized by temporary popularity. This finding corresponds with Barnhart's model of lexical survival, according to which the successful integration of a neologism depends on its ability to remain relevant beyond the circumstances that initially generated it [Barnhart 1973: 51]. Frequency therefore functions as an indicator of communicative necessity rather than merely statistical prominence.

Semantic stability emerged as one of the strongest predictors of lexicographic inclusion. Neologisms whose meanings remained relatively consistent across contexts were incorporated more rapidly than expressions characterized by semantic fluctuation. The importance of semantic consistency has been emphasized in lexicographic theory. Scientists note that lexicographers encounter considerable difficulties when attempting to define lexical units whose meanings vary substantially across discourse contexts [Bushuy 2018: 62-67]. The findings of the present study support this argument. Lexical innovations exhibiting stable semantic structures demonstrated significantly shorter intervals between emergence and codification.

The data further reveal that dictionary inclusion is strongly associated with cross-domain distribution. Lexical items appearing in journalistic, academic, digital, and conversational discourse simultaneously exhibited substantially higher inclusion rates than those restricted to specialized communities. This observation aligns with corpus-based approaches emphasizing representativeness and contextual diversity as essential indicators of lexical integration [Bushuy 2023: 23–26]. Broad distribution appears to signal the transition of a lexical innovation from a local phenomenon to a socially recognized linguistic resource.

The analysis enabled the identification of a five-stage model describing the progression from lexical innovation to dictionary inclusion.

Stage 1: Emergence. The first stage involves the appearance of a lexical innovation within a restricted communicative environment. At this stage, usage is typically limited to specific social, professional, or online communities. Many contemporary neologisms emerge in digital environments where rapid interaction facilitates lexical experimentation. Previous research has shown that online discourse provides particularly favorable conditions for the creation and dissemination of new vocabulary [Bushuy 2023: 22].

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Stage 2: Diffusion. During the diffusion stage, a lexical innovation begins spreading beyond its original community. Corpus evidence indicates that social media platforms significantly accelerate this process by increasing visibility and facilitating repeated exposure. The findings support the argument that lexical diffusion in digital communication differs substantially from traditional models of language change. Contemporary neologisms frequently achieve national or even global circulation within relatively short periods [Bushuy 2025: 16-18].

Stage 3: Stabilization. The stabilization stage is characterized by increasing semanti/c consistency and broader acceptance among language users. Corpus evidence demonstrates a reduction in contextual variability and a corresponding increase in predictability of usage. This stage is particularly important because many lexical innovations fail to progress beyond diffusion. Words that do not achieve semantic stabilization often disappear before attracting lexicographic attention.

Stage 4: Lexicographic Evaluation. Once a lexical innovation exhibits sufficient frequency and stability, it becomes subject to systematic monitoring by lexicographers. Contemporary lexicographic practice relies heavily on corpus evidence during this stage. As Kilgarriff argues, corpus analysis enables lexicographers to observe authentic patterns of usage and evaluate the representativeness of lexical innovations [Kilgarriff 2001: 97–133].

Stage 5: Codification. The final stage involves official dictionary inclusion. Codification may be interpreted as a form of institutional recognition indicating that a lexical item has achieved sufficient social and linguistic legitimacy. Research on neologism inclusion demonstrates that dictionary registration often occurs only after extensive corpus evidence has accumulated [Faleyeva 2021: 99-104]. The findings of the present study confirm that codification represents the culmination of a gradual process rather than an immediate response to lexical novelty.

The term deepfake illustrates a relatively successful trajectory of lexical codification. Initially confined to technological discourse, the term expanded rapidly into journalism, law, politics, and public debate. Corpus data reveal strong frequency growth, broad distribution, and remarkable semantic stability. These characteristics correspond closely to the criteria identified by Barnhart as indicators of successful lexical integration [Barnhart 1980: 31].

The lexical item doomscrolling emerged during the COVID-19 pandemic but subsequently maintained relevance beyond its original context. Its survival appears to result from its ability to describe a broader behavioral phenomenon associated with digital media consumption. The persistence of the term supports Barnhart's observation that successful neologisms often extend beyond the circumstances of their initial emergence [Barnhart 1990: 41].

The emergence of generative artificial intelligence has produced a new category of technological neologisms. Among them, prompt engineering represents an instructive example of specialized terminology entering mainstream discourse. The corpus evidence demonstrates increasing frequency, expanding distribution, and growing semantic standardization. Such patterns correspond to the codification model identified in the present study and reflect broader tendencies in contemporary technological vocabulary.

The analysis identified several challenges confronting modern lexicography.

The first challenge concerns the ephemeral nature of digital vocabulary. Many online expressions achieve temporary popularity but disappear before developing stable semantic structures. This phenomenon complicates lexicographic evaluation because frequency alone may produce misleading impressions of lexical significance [Barnhart 1991: 37].

A second challenge involves semantic fluidity. The boundaries between lexical meanings are often difficult to define precisely, particularly in the case of emerging vocabulary [Faleeva 2022: 59-67]. The rapid evolution of digital discourse further intensifies this problem.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

The third challenge relates to corpus representativeness. While contemporary corpora provide unprecedented access to authentic language data, researchers must still address questions concerning demographic diversity, genre balance, and platform-specific biases.

Overall, the findings demonstrate that dictionary inclusion is best understood as a multidimensional process involving the interaction of frequency, semantic stability, distribution, longevity, and social relevance. These results provide empirical support for corpus-based approaches to lexicography and establish a foundation for the theoretical discussion presented in the next section.

Discussion. The findings of the present study suggest that the lexicographic representation of English neologisms cannot be explained through a single criterion. Rather, dictionary inclusion emerges as the result of a multidimensional evaluation process involving frequency, semantic stability, distribution across communicative domains, longevity, and sociocultural relevance. This conclusion supports contemporary corpus-based approaches to lexicography, which emphasize the necessity of combining quantitative evidence with qualitative linguistic analysis.

One of the most significant findings concerns the limitations of frequency as an independent indicator of lexical legitimacy. Traditional assumptions often imply that widespread usage automatically leads to dictionary inclusion. However, the corpus evidence presented in this study demonstrates that frequency alone does not guarantee codification. Numerous lexical innovations achieved substantial visibility within digital communication environments but failed to demonstrate the stability required for long-term lexicographic recognition.

This observation confirms argument that successful neologisms must satisfy broader communicative and social requirements beyond simple frequency of occurrence [Faleeva 2023: 202-206]. Lexical survival depends not merely on repeated use but on the ability of a lexical item to satisfy an enduring communicative need within the speech community. Consequently, dictionary inclusion should be interpreted as recognition of sustained communicative value rather than temporary popularity.

The findings also support Kilgarriff's criticism of purely frequency-based approaches to lexicography. According to Kilgarriff, corpus evidence becomes meaningful only when examined within broader contextual and semantic frameworks [Kilgarriff 1997: 135–155]. The present study confirms that lexicographers increasingly rely on contextualized corpus analysis rather than raw frequency counts. This tendency reflects a broader methodological shift within contemporary lexicography toward evidence-based interpretation.

Among all variables examined, semantic stability emerged as one of the strongest predictors of lexicographic codification. The results indicate that lexical innovations characterized by relatively fixed meanings are considerably more likely to be incorporated into dictionaries than semantically unstable expressions.

The importance of semantic stability can be explained by the practical objectives of lexicography. Dictionaries are designed to provide users with reliable and relatively stable descriptions of lexical meaning. When a lexical item exhibits substantial semantic variability, the process of dictionary definition becomes significantly more complex. As scientists observe, the boundaries between lexical senses are often difficult to determine, particularly in cases involving newly emerging vocabulary [Erk 2009: 10–18].

One of the most significant contributions of the present study concerns the relationship between digital communication and contemporary neologization. The findings indicate that social media platforms, online communities, and algorithmically mediated communication environments have fundamentally altered the mechanisms through which lexical innovations emerge and spread.

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

Traditional models of lexical diffusion typically describe relatively gradual processes occurring over extended periods. By contrast, the neologisms examined in this study frequently achieved widespread recognition within months rather than years. The diffusion stage identified in the proposed codification model appears considerably shorter than comparable stages described in earlier studies of lexical change.

This acceleration may be explained by several factors. First, digital communication dramatically increases the visibility of lexical innovations. New expressions can be encountered simultaneously by millions of users across different geographical regions. Second, online interaction encourages linguistic creativity and experimentation. Third, algorithmic recommendation systems actively facilitate the circulation of novel linguistic forms.

The findings therefore support contemporary arguments that digital environments should be regarded not merely as channels of communication but as active agents influencing linguistic evolution. Lexical innovation increasingly emerges through interactions between human users, technological infrastructures, and algorithmic systems.

This development creates both opportunities and challenges for lexicography. On the one hand, digital communication provides unprecedented access to authentic linguistic data. Corpus compilation and lexical monitoring have become more efficient than ever before. On the other hand, the speed of digital language change complicates traditional lexicographic procedures designed for slower patterns of lexical development.

The findings further demonstrate the central role of corpus linguistics in contemporary dictionary compilation. The five-stage model proposed in this study reveals that corpus evidence influences virtually every phase of lexicographic evaluation. This observation aligns with the broader corpus revolution described by modern lexicographic scholarship. Corpus methodologies have transformed dictionary compilation from a largely intuition-based activity into an empirically grounded scientific practice. The availability of large digital corpora allows lexicographers to observe actual language use rather than relying exclusively on editorial judgment. Linguists emphasize that corpus data provide essential evidence regarding frequency, collocation, contextual variation, and semantic development [Metcalf, Barnhart 1997: 21]. The present findings strongly support this position. All major criteria identified in the study – frequency, distribution, semantic stability, and longevity – can be evaluated through corpus analysis. Nevertheless, the results also reveal certain limitations of corpus-based approaches. Corpora reflect language use but do not automatically determine linguistic significance. Lexicographers must still interpret corpus evidence and evaluate broader social and communicative factors. Consequently, modern lexicography should be understood as an interaction between quantitative data analysis and qualitative linguistic expertise.

The results of this study suggest the need for a revised understanding of dictionary inclusion in the digital age. Traditional lexicographic models often conceptualize dictionaries as repositories documenting already established vocabulary. Contemporary evidence indicates a more dynamic relationship between lexical innovation and lexicographic representation. The five-stage model proposed in this research conceptualizes dictionary inclusion as a gradual process involving emergence, diffusion, stabilization, evaluation, and codification. This framework emphasizes the transitional nature of neologisms and highlights the role of lexicographic institutions as mediators between innovation and linguistic convention. Unlike purely descriptive models, the proposed framework recognizes that dictionary inclusion involves both empirical observation and institutional judgment. Lexicographers do not simply record lexical facts; they evaluate linguistic developments and determine which innovations have achieved sufficient legitimacy to warrant codification. This interpretation is consistent with research on neologism inclusion, which demonstrates that dictionary

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

compilation inevitably involves editorial decision-making and value judgments regarding linguistic significance. Consequently, lexicographic codification should be viewed as a process of linguistic validation rather than mere documentation.

Conclusion. The findings demonstrate that dictionary inclusion cannot be adequately explained by frequency of occurrence alone. Although corpus frequency remains an essential indicator of lexical relevance, contemporary lexicographic practice relies on a multidimensional evaluation framework that incorporates semantic stability, communicative distribution, longevity of usage, and sociocultural significance. The results therefore support current corpus-based approaches to lexicography, according to which quantitative evidence must be interpreted within broader linguistic and social contexts.

One of the principal contributions of the study is the identification of a five-stage model of lexical codification consisting of emergence, diffusion, stabilization, lexicographic evaluation, and codification. This model provides a systematic explanation of how lexical innovations progress from isolated usage within specific communities to institutional recognition through dictionary inclusion. The proposed framework highlights the dynamic nature of neologization and demonstrates that dictionary registration represents the culmination of a gradual process rather than an immediate response to lexical novelty.

The analysis further revealed the central importance of semantic stability in lexicographic decision-making. While frequency enables lexicographers to identify emerging lexical trends, semantic consistency determines whether a reliable and sufficiently stable dictionary definition can be formulated. Neologisms characterized by stable meanings and broad communicative distribution were shown to possess considerably greater probabilities of inclusion than semantically fluid or context-dependent expressions.

Another important finding concerns the role of digital communication in contemporary lexical innovation. Social media platforms, online communities, and algorithmically mediated communication environments have accelerated the diffusion of neologisms and fundamentally transformed traditional patterns of lexical change. As a result, lexicographers increasingly confront the challenge of distinguishing between temporary digital trends and durable lexical innovations. The study demonstrates that digital communication simultaneously facilitates lexical monitoring and complicates lexicographic evaluation due to the rapid emergence and disappearance of online vocabulary.

References:

1. Barnhart C. L., Barnhart D. K., Steinmetz S. The Barnhart Dictionary of New English Since 1963. – New York: Wilson, 1973. – 512 p.
2. Barnhart C. L., Barnhart D. K., Steinmetz S. The Second Barnhart Dictionary of New English. – New York: H. W. Wilson, 1980. – 386 p.
3. Barnhart C. L., Barnhart D. K., Steinmetz S. The Third Barnhart Dictionary of New English. – New York: H. W. Wilson, 1990. – 432 p.
4. Barnhart D. K. A Calculus for New Words // Dictionaries: Journal of the Dictionary Society of North America. – 2007. Vol. 28. – P. 132–138.
5. Barnhart D. K. Neo Words and Contemporary Lexical Innovation. – Hyde Park: Lexik House, 1991. – 312 p.
6. Bushuy T.A. The Emergence of Language Multitude // Scientific reports of Bukhara State University. – 2018. – T. 1. – №. 3. – C. 62-67.
7. Bushuy T.A. (2023). Language in Modern Linguistics. Miasto Przyszłości, 39, P. 23–26. Retrieved from https://miastoprzyszlosci.com.pl/index.php/mp/article/view/1678

HSR (LONDON),
HOUGHTON STREET REVIEW JOURNAL
ISSN: 2058-5519|Impact factor-15.9
http://houghtonstreetreview.com/index.php/hsr

8. Bushuy T. Features of Representation of Some Word Forms and Meanings of Phraseological Units in Contrastive Lexicographical Sources // Международный журнал языка, образования, перевода. – 2023. – Т. 4. – №. 2. – 22 с.
9. Bushuy Tatyana Anatolyevna The Problem of Word Meaning and Phraseologically Related Meaning in Linguistic Science // Современные инновации. 2025. №1 (47). URL: https://cyberleninka.ru/article/n/the-problem-of-word-meaning-and-phraseologically-related-meaning-in-linguistic-science. - pp 16-18.
10. Faleyeva A.V. Derivational peculiarities of colloquial words in modern English dictionaries // Язык и культура. – 2021. – С. 99-104.
11. Faleeva A.V. Investigation of the Colloquial Formation of Modern English Language // Международный журнал языка, образования, перевода. – 2022. – Т. 3. – №. 2. – С. 59-67.
12. Faleeva A.V. Grammatical, Semantic and Structural Features of English Colloquial Words and Expressions // International Journal of Language Learning and Applied Linguistics. – 2023. – Т. 2. – №. 5. – С. 202-206.
13. Erk K. Investigations on Word Senses and Word Usages // Proceedings of ACL-IJCNLP, 2009. – P. 10–18.
14. Kilgarriff A. Putting Frequencies in the Dictionary // International Journal of Lexicography, 1997. – Vol. 10. – №. 2. – P. 135–155.
15. Kilgarriff A. Comparing Corpora // International Journal of Corpus Linguistics, 2001. – Vol. 6. – №. 1. – P. 97–133.
16. Metcalf A., Barnhart D. K. America in So Many Words: Words That Have Shaped America. – Boston: Houghton Mifflin, 1997. – 240 p.