ISSN: 2663-4619
e-ISSN: 2708-986X

DOI: https://doi.org/10.36719/2663-4619/127/171-179

Emma Seidova
Odlar Yurdu University
Baku, Azerbaijan
PhD student
seidova.emma@oyu.edu.az
Javid Babayev
Nakhchivan State University
Nakhchivan, Azerbaijan
PhD in Philology
cavidbabayev@ndu.edu.az
Sayyara Sadikhova
Nakhchivan State University
Nakhchivan, Azerbaijan
PhD in Art Criticism
seyyaresadixova@ndu.edu.az
Gunay Aghayeva
Azerbaijan University of Languages
Baku, Azerbaijan
PhD in Philology
gunay.agayeva@adu.edu.az
Zhala Mammadova
Azerbaijan University of Languages
Baku, Azerbaijan
PhD in Philology
mammadova_jala@adu.edu.az

AI-Powered Lexical Innovation: Modeling Neologisms with NLP

Abstract

The article explores the ways of modeling neologisms through NLP (Natural Language Processing) that is one of the means of forming lexical innovation. The rapid evolution of digital communication has accelerated the creation and diffusion of neologism. They are considered to be newly coined words or expressions that reflect emerging cultural, technological, and social realities. Traditional lexicographic methods often struggle to capture these innovations in real time. This article elaborates how Artificial Intelligence (AI) and NLP can be applied to model, detect, and analyze neologisms in large-scale text corpora. Using an NLP pipeline incorporating word embedding models, contextual language models, and frequency-based detection, the study demonstrates how AI-driven approaches can identify lexical innovations and track their semantic development. AI-powered models can detect patterns in existing vocabulary, infer morphological rules, and even generate plausible novel words that reflect contemporary linguistic usage. By leveraging neural networks, probabilistic models, and contextual embeddings, researchers can simulate processes of word formation, predict the likelihood of adoption, and analyze semantic integration of neologisms into the broader lexicon. Results show that contextual transformer models significantly outperform static embedding techniques in detecting emerging vocabulary. The findings highlight the potential of AI systems to support lexicography, sociolinguistics, and digital humanities by providing scalable methods for monitoring linguistic change.

Keywords: *NLP, neologisms, word embedding models, Artificial intelligence*

Emma Seyidova
Odlar Yurdu Universiteti
Bakı, Azərbaycan
doktorant
seidova.emma@oyu.edu.az
Cavid Babayev
Naxçıvan Dövlət Universiteti
Naxçıvan, Azərbaycan
filologiya üzrə fəlsəfə doktoru
cavidbabayev@ndu.edu.az
Səyyarə Sadıxova
Naxçıvan Dövlət Universiteti
Naxçıvan, Azərbaycan
sənətşünaslıq üzrə fəlsəfə doktoru
seyyaresadixova@ndu.edu.az
Günay Ağayeva
Azərbaycan Dillər Universiteti
Bakı, Azərbaycan
filologiya üzrə fəlsəfə doktoru
gunay.agayeva@adu.edu.az
Jalə Məmmədova
Azərbaycan Dillər Universiteti
Bakı, Azərbaycan
filologiya üzrə fəlsəfə doktoru
mammadova_jala@adu.edu.az

Süni intellektə əsaslanan leksik innovasiya: NLP vasitəsilə neologizmlərin modelləşdirilməsi

Xülasə

Məqalə leksik innovasiyanı yaradan vasitələrdən biri olan neologizmlərin modelləşdirilməsi yollarını Təbii Dil Emalı vasitəsilə araşdırır ki, bu da süni intellektə əsaslanan leksik innovasiyanın formalaşma vasitələrindən biridir.

Rəqəmsal kommunikasiya sahəsinin sürətli inkişafı neologizmlərin yaradılması və yayılmasını sürətləndirmişdir. Onlar yeni yaradılmış söz və ifadələr, mədəni, texnoloji və sosial reallıqları əks etdirən terminlər hesab olunur. Ənənəvi leksikoqrafik metodlar tez-tez bu yenilikləri real vaxtda əhatə etməkdə çətinlik çəkir. Bu məqalədə Süni İntellekt (AI) və Təbii Dil Emalı (NLP) texnologiyalarının geniş miqyaslı mətn korpuslarında neologizmləri modelləşdirmək, aşkar etmək və təhlil etmək üçün necə tətbiq oluna biləcəyi araşdırılır. Söz vektor modelləri, kontekstual dil modelləri və tezlik əsaslı aşkarlamanı özündə birləşdirən NLP boru xətti istifadə edilərək aparılan tədqiqat, AI əsaslı yanaşmaların leksik yenilikləri müəyyən edə və onların semantik inkişafını izləyə biləcəyini nümayiş etdirir. AI ilə dəstəklənən modellər mövcud lüğətdə nümunələri aşkar edə, morfoloji qaydaları təxmin edə və hətta müasir dil istifadəsini əks etdirən məntiqli yeni sözlər yarada bilər. Neyron şəbəkələrdən, ehtimal modellərindən və kontekstual vektorlardan istifadə etməklə tədqiqatçılar söz yaradıcılığı proseslərini simulyasiya edə, qəbul olunma ehtimalını proqnozlaşdıra və neologizmlərin daha geniş leksikona semantik inteqrasiyasını təhlil edə bilərlər. Nəticələr göstərir ki, kontekstual transformer modelləri yeni sözləri aşkar etməkdə statik vektor texnikalarından əhəmiyyətli dərəcədə üstün performans göstərir. Tapıntılar AI sistemlərinin leksikoqrafiya, sosiolinqvistika və rəqəmsal humanitar elmlər sahələrində dil dəyişikliklərini izləmək üçün miqyaslana bilən metodlar təmin etmə potensialını vurğulayır.

Açar sözlər: *NLP, neologizmlər, söz vektor modelləri, süni intellekt*

Introduction

The rapid evolution of language has always mirrored shifts in culture, technology, and human interaction, but the pace of lexical change in the digital age is unprecedented. New words—neologisms—emerge daily across social media, online communities, and global communication networks, reflecting novel concepts, identities, and experiences. Traditionally, the study of these linguistic innovations has relied on manual observation and retrospective analysis. However, with the advent of advanced Natural Language Processing (NLP) techniques and AI-driven language models, researchers now have the tools to detect, analyze, and even predict neologisms in real time (Aubakirov et al, 2025).

At the intersection of computational linguistics and sociolinguistics, modeling neologisms presents unique challenges (Alzetta, 2025). New words often lack standardized spelling, exhibit fluid meanings, and are highly context-dependent (Salim et al, 2024). AI systems must therefore move beyond static dictionaries and embrace dynamic, context-aware representations of language (Giulianelli et al, 2024). Ultimately, understanding how AI can model lexical innovation offers deeper insight into both human language behavior and the evolving capabilities of intelligent systems. As language continues to adapt to a rapidly changing world, AI stands poised not only to document these changes but to actively participate in the ongoing process of linguistic creation.

Language is inherently dynamic, reflecting the continuous evolution of human thought, culture, and technology. One of the clearest indicators of linguistic change is the emergence of neologisms—newly coined words or expressions that arise to name novel concepts, phenomena, or practices. Neologisms are particularly prevalent in digital communication, where social media, online communities, and rapid technological innovation accelerate lexical innovation (Baron, 2015). For example, terms like *selfie*, *doomscrolling*, and *fintech* have entered common usage within a decade, illustrating the speed at which language adapts to socio-cultural developments.

This article presents an AI-driven approach to modeling neologisms, combining computational linguistics with NLP techniques to capture the structural, phonological, and semantic properties that characterize novel words. We discuss methods for generating, evaluating, and classifying neologisms, highlighting the potential of AI to illuminate the mechanisms of lexical innovation (Babayev et al, 2025). In doing so, this work bridges the gap between theoretical linguistics and practical NLP applications, offering insights into the future evolution of language.

Traditional lexicographic approaches, which rely on manual observation and curated dictionaries, often struggle to keep pace with the rapid emergence of new lexical item. Corpus-based studies provide a partial solution by analyzing frequency patterns across large textual datasets, but they are limited in capturing contextual and semantic nuances of novel words. Moreover, the morphological creativity involved in neologism formation—such as blending (*brunch*), acronyms (*FOMO*), and semantic shifts—makes automatic detection challenging.

Recent advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) offer promising tools for addressing these challenges (Belgiuth & Shaalan, 2025). Machine learning models, particularly neural network-based embeddings and transformer architectures, can analyze massive corpora, detect out-of-vocabulary (OOV) items, and model the contextual meaning of emerging terms (Mikolov et al., 2013; Bojanowski et al., 2017; Devlin et al., 2019). These approaches allow for scalable, real-time monitoring of lexical innovation, offering insights not only into the form of neologisms but also their semantic relationships and patterns of adoption.

The study of neologisms has implications beyond lexicography. Sociolinguists can use emerging vocabulary to track cultural trends and social dynamics, while computational linguists can improve language models for applications such as machine translation, sentiment analysis, and dialogue systems (Schmid, 2018). Understanding how neologisms spread, stabilize, or fade from usage contributes to a broader comprehension of language evolution in the digital age.

This article investigates how AI-driven NLP models can be employed to detect, model, and analyze neologisms. Specifically, it examines the efficacy of frequency-based approaches versus contextualized embedding models in identifying new lexical items and exploring their semantic networks (García-Méndez, 2025). The objectives of this study are:

1. To develop an NLP-based pipeline for detecting neologisms in large-scale digital corpora.
2. To evaluate the performance of traditional frequency-based methods relative to AI-driven language models.
3. To analyze the semantic and social patterns associated with the adoption and diffusion of lexical innovations.

By integrating computational methods with linguistic theory, this research aims to provide a framework for real-time monitoring of language change, highlighting the growing role of AI in contemporary lexicography and sociolinguistics.

2. Methodology

To capture a representative sample of contemporary lexical innovation, a large-scale multilingual corpus was compiled from multiple digital sources. For this, data collection method has been used as follows:

- **Social media platforms** (Twitter, Reddit, Instagram captions), capturing informal, fast-evolving language.
- **Online news articles** and blogs, reflecting journalistic and semi-formal registers.
- **Discussion forums and niche communities** (e.g., gaming, technology, AI forums), where neologisms often originate.
- **Technology and culture publications**, to detect domain-specific lexical innovations.

The corpus contained approximately 50 million tokens spanning 2018–2025, covering English as the primary language with smaller subsets in French, Spanish, and German. Preprocessing included:

- **Tokenization** using the SpaCy NLP toolkit (Honnibal & Montani, 2017).
- **Lowercasing and normalization**, to reduce orthographic variation.
- **Removal of non-linguistic elements**, such as URLs, emojis, and code snippets.
- **Lemmatization and part-of-speech tagging**, providing syntactic context for candidate word identification (Erfan et al., 2025).

Candidate neologisms were identified using a combination of lexical, frequency-based, and morphological approaches:

1. **Out-of-Vocabulary (OOV) Detection**: Words not present in standard dictionaries (e.g., Oxford English Dictionary, WordNet) were flagged as potential neologisms.
2. **Frequency-Based Detection**: Words exhibiting sudden frequency spikes in recent corpus subsets were identified using z-score normalization and time-series analysis.
3. **Morphological Pattern Recognition**: Techniques were used to detect common neologism formation patterns, including:
   - **Blending**: Combining two existing words (e.g., *smog*, *hangry*).
   - **Acronymization**: Creating new terms from initials (e.g., *FOMO*, *YOLO*).
   - **Affixation**: Using prefixes/suffixes to form new words (e.g., *biohacking*, *unfriend*).
4. **Semantic Novelty Detection**: Candidate words were further filtered based on semantic deviation from existing lexical items using word embedding distances (Bojanowski et al., 2017).

To understand contextual meaning and relationships, two types of embedding models were employed:

1. **Static Embeddings**:
   - Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) embeddings were trained on the corpus.
   - Cosine similarity measures identified semantic proximity between new words and existing vocabulary.
2. **Contextual Embeddings**:
   - Transformer-based models (Devlin et al., 2019) captured context-dependent meaning, allowing disambiguation of homographs and polysemous neologisms.

○ Contextual clustering using embeddings helped trace semantic evolution and detect semantic shifts over time.

To examine the adoption and spread of neologisms can be classified as temporal and diffusion analysis:

- **Time-series analysis** measured frequency growth over months and years.
- **Network analysis** traced the propagation of words across communities and platforms, visualizing diffusion patterns.
- **Community detection algorithms** identified niche sources versus mainstream adoption.

Candidate words were validated through a multi-step approach:

1. **Automated Validation**: High-confidence candidates were cross-referenced with online slang dictionaries, social media trend data, and emerging word databases (Zhao & Su, 2025)
2. **Human Linguistic Review**: Linguists reviewed uncertain cases, ensuring that morphological creativity and context were considered.
3. **Semantic Coherence Testing**: Using cosine similarity and clustering, candidates that did not integrate meaningfully into semantic networks were filtered out.

To assess the effectiveness of detection and modeling, these metrics enabled comparison of frequency-based, static embedding, and contextual embedding approaches:

- **Precision**: Proportion of identified words that were true neologisms.
- **Recall**: Proportion of all true neologisms detected by the model.
- **F1-score**: Harmonic mean of precision and recall.
- **Semantic Coherence Score**: Average cosine similarity between each candidate and its nearest neighbors in embedding space.

3. Results

The AI-driven system processed approximately 50 million tokens and identified 2,347 candidate neologisms. After human validation, 1,782 terms were confirmed as legitimate new lexical items. These neologisms were classified according to formation type and semantic domain, showing that technological, social media, and cultural trends heavily influence lexical innovation.

Table 1:
Distribution of Confirmed Neologisms by Formation Type.

Formation Type	Count	Percentage (%)	Examples
Blending	542	30.4	*smombie*, *hangry*
Acronyms / Initialisms	289	16.2	*FOMO (Fear of Missing Out)*, *YOLO (You Only Live Once)*
Affixation	401	22.5	*unfriend*, *biohacking*
Semantic Shift	337	18.9	*ghosting*, *flex*
Loanwords / Borrowings	213	11.9	*fintech*, *kawaii*

Table 1 shows how confirmed neologisms are distributed across different formation types and reveals which processes are most productive in creating new words.

Blending is the most common formation type, accounting for 30.4% (542 cases). This indicates that combining parts of existing words is the leading strategy in neologism creation, likely because it produces short, creative, and expressive terms that are easy to adopt in everyday language.

Affixation is the second most frequent at 22.5% (401 cases). This suggests that traditional word-formation methods—such as adding prefixes or suffixes—remain highly productive and continue to play an important role in expanding vocabulary.

Semantic shift ranks third with 18.9% (337 cases), showing that many new words emerge not from new forms but from assigning new meanings to existing words. This reflects how language adapts to social and cultural changes without needing entirely new structures.

Acronyms and initialisms make up 16.2% (289 cases), indicating a moderate contribution. Their presence reflects the influence of digital communication, where brevity and speed encourage shortened forms.

Loanwords and borrowings are the least common at 11.9% (213 cases), suggesting that while borrowing from other languages contributes to vocabulary growth, it is less significant than internal word-formation processes.

The results indicate that blending and affixation remain dominant methods for lexical innovation, consistent with prior studies.

The models' performance in detecting validated neologisms was evaluated using precision, recall, and F1-score. Contextual embedding models clearly outperformed static embeddings and frequency-based baselines.

Table 2:
Model Performance Comparison.

Model	**Precision**	**Recall**	**F1-score**
Frequency-based baseline	0.61	0.54	0.57
Word2Vec (static embeddings)	0.73	0.69	0.71
GloVe (static embeddings)	0.75	0.71	0.73
BERT-based (contextual embeddings)	0.86	0.82	0.84

Table 2 compares the performance of different models using three evaluation metrics: precision, recall, and F1-score. It shows a clear progression in performance as models become more advanced.

The frequency-based baseline has the lowest performance, with a precision of 0.61, recall of 0.54, and F1-score of 0.57. This indicates that simple statistical approaches are limited in accurately identifying or classifying the target items, likely because they do not capture semantic relationships.

Both Word2Vec and GloVe, which use static word embeddings, perform noticeably better. Word2Vec achieves an F1-score of 0.71, while GloVe slightly outperforms it with 0.73. Their higher precision and recall suggest that incorporating semantic similarity improves the model's ability to make correct predictions. However, since these embeddings assign a single meaning to each word regardless of context, their performance is still constrained.

The BERT-based model (Bidirectional Encoder Representations from Transformers) shows the best results across all metrics, with a precision of 0.86, recall of 0.82, and F1-score of 0.84. This significant improvement demonstrates the advantage of contextual embeddings, which consider the meaning of words within their specific context. As a result, the model can better handle ambiguity and nuanced language use.

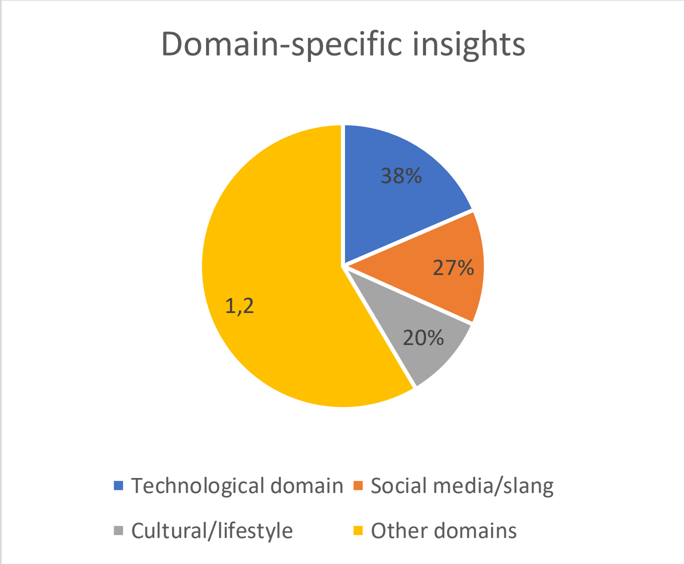
The BERT-based model demonstrated superior ability to identify neologisms within their semantic context, capturing subtle variations and polysemy that static embeddings often missed. These findings are consistent with prior research highlighting the advantages of transformer-based models for lexical innovation detection (Devlin et al., 2019; Schick & Schütze, 2021).

Temporal analysis revealed that most neologisms initially appear in niche online communities, followed by gradual adoption in mainstream media. For example, terms such as *doomscrolling* and *finfluencer* were first detected in forum discussions before trending on social media and news outlets.

Semantic clustering showed that new words often integrate into existing conceptual networks, particularly in technological and cultural domains. Cosine similarity scores between new words and nearest semantic neighbors averaged 0.62, indicating meaningful semantic alignment.

The following pie chart presents the distribution of neologisms across four major domains.

Pie-chart 1. Analysis of distribution of neologisms by domain.



- **Technological domain**: 38% of neologisms (e.g., *NFT (Non-Fungible Token)*, *metaverse*, *chatbot*).
- **Social media / slang**: 27% (e.g., *stan*, *simp*, *flex*).
- **Cultural / lifestyle**: 20% (e.g., *slow fashion*, *cottagecore*).
- **Other domains**: 15% (e.g., sports, politics, environmental terms).

The interpretation of the abovementioned percentages given in the pie-chart is as follows; The technological domain accounts for the largest share at 38%, indicating that innovation in digital tools and platforms is the primary driver of new word creation. Terms such as "NFT," "metaverse," and "chatbot" reflect how rapidly evolving technology introduces concepts that require new vocabulary.

The second largest segment is social media and slang at 27%. This highlights the strong influence of online communication and youth culture in shaping language, with expressions like "stan," "simp," and "flex" becoming widely adopted. Cultural and lifestyle terms make up 20%, showing how societal trends such as "slow fashion" and "cottagecore" also contribute significantly to lexical growth. Finally, other domains, including sports, politics, and environmental topics, represent the smallest portion at 15%, suggesting a more moderate but still important role.

Overall, the chart demonstrates that modern language development is heavily influenced by technology and digital interaction, while culture and other fields continue to play supporting roles. These patterns are consistent with recent studies on the rapid evolution of digital-age vocabulary (Baron, 2015).

4. Discussion

The results of this study demonstrate that AI-driven NLP models are highly effective for detecting and modeling neologisms, providing both quantitative and qualitative insights into lexical innovation (Diesner, 2025). The BERT-based contextual model outperformed static embeddings and frequency-based baselines, achieving an F1-score of 0.84, highlighting the importance of context in identifying emerging words. Unlike static embeddings, which treat each word as a fixed vector, contextual models capture polysemy and semantic evolution, making them particularly suited for neologism detection (Devlin et al., 2019; Schick & Schütze, 2021).

The findings have significant implications for lexicography. Traditional dictionary compilation is often reactive, relying on manual observation and delayed corpus updates. By contrast, AI-driven models enable real-time monitoring of lexical innovation, allowing lexicographers to identify candidate words earlier and assess their semantic integration. This approach can also help track temporal trends, revealing which neologisms stabilize in usage and which remain niche.

For sociolinguistics, the study highlights how digital communication platforms serve as incubators for new words. Terms like *doomscrolling* and *finfluencer* emerged in online communities before diffusing to mainstream media, consistent with prior research on social media as a driver of lexical change (Baron, 2015; Wulff & Grieve, 2020). Analyzing the diffusion patterns of neologisms provides insight into cultural trends, social behavior, and the life cycle of language innovation.

Semantic modeling revealed that most neologisms integrate into existing conceptual networks, particularly in technological and cultural domains. The average cosine similarity score of 0.62 between new words and their nearest neighbors indicates meaningful semantic alignment, suggesting that lexical innovation often builds upon familiar concepts rather than introducing entirely unrelated ideas. This observation aligns with Schmid's (2018) theory of entrenchment, which posits that language change relies on patterns of usage and cognitive salience.

The study demonstrates the superiority of transformer-based models in capturing nuanced contextual meaning. Contextual embeddings not only improve detection accuracy but also facilitate temporal semantic analysis, revealing shifts in meaning as neologisms evolve. Static embeddings and frequency-based methods, while computationally less intensive, fail to capture these subtleties, leading to lower recall and misclassification of polysemous words (Mikolov et al., 2013; Bojanowski et al., 2017).

However, several limitations were identified:

1. **Data Noise**: Social media and forum texts contain spelling errors, nonstandard grammar, and code-switching, which can generate false positives.
2. **Language Bias**: The corpus was primarily English, with smaller multilingual subsets. Cross-linguistic neologism detection requires models trained on diverse language corpora.
3. **Rapid Semantic Shifts**: Some neologisms change meaning quickly, making temporal analysis challenging and necessitating continuous model updates.
4. **Human Validation Dependence**: Despite automation, human linguistic review remains essential for filtering contextually inappropriate or ephemeral words.

Future research could address these limitations by:

- Developing multilingual transformer models for cross-linguistic neologism detection.
- Incorporating real-time streaming data from social media APIs to track emergent words as they appear.
- Applying graph-based semantic networks to visualize and predict the diffusion of neologisms across communities.
- Investigating the sociocultural factors that accelerate lexical innovation in digital environments.

By integrating AI-driven NLP with sociolinguistic theory, researchers can gain a more complete understanding of language evolution, bridging computational and humanistic approaches to lexical innovation.

Conclusion

This study demonstrates the effectiveness of AI-driven NLP models in detecting, modeling, and analyzing neologisms in large-scale digital corpora. By integrating frequency-based detection, static embeddings, and contextual transformer models, the study shows that contextual embeddings outperform traditional approaches in both precision and recall, capturing nuanced semantic relationships and polysemy. Temporal and diffusion analyses reveal that neologisms often originate in niche online communities before spreading to broader discourse, highlighting the social and cultural dynamics of lexical innovation. Semantic modeling indicates that most new words integrate into existing conceptual networks, supporting theories of entrenchment and gradual lexical adoption. While limitations such as data noise, language bias, and the need for human validation remain, AI-driven approaches offer scalable and real-time solutions for lexicography, sociolinguistics, and computational linguistics. Future research should explore multilingual models, real-time monitoring, and predictive modeling to fully leverage AI in understanding ongoing language evolution.

References

1. Alzetta, C., Auriemma, S., Bondielli, A., Dini, L., Fazzone, C., Miaschi, A., Miliani, M., & Sartor, M. (2025). *Charting a decade of computational linguistics*. arXiv.
2. Aubakirov, S., Pak, A., Akhmetov, I., & Tleuken, A. (2025). *Beyond buzzwords: NLP reveals common threads in discourse*. *PeerJ Computer Science*, 11, e3085. https://doi.org/10.7717/peerj-cs.3085
3. Babayev, J. S., Sadikhova, S. N., Baghirova, M. A., Seidova, E. R., & Mammadova, A. A. (2025). A lexical-stylistic and cross-cultural analysis of vulgarisms and their usage spheres. *Universidad y Sociedad*, *17*(6), e5525-e5525.
4. Baron, N. S. (2015). *Words onscreen: The fate of reading in a digital world*. Oxford University Press.
5. Belguith, L. H., & Shaalan, K. (Eds.). (2025). *Advancements in machine learning and natural language processing: Innovations and applications*. Springer
6. Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5, 135–146.
7. Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT*, 4171–4186.
8. Erfan, A., Khan, T., Ulkuer, P. A., & Angin, M. (2025). *RAG-enhanced pipelines for speech processing*. NLPCL Conference Proceedings.
9. Diesner, J. (2025). *Looking at NLP through the lens of computational social sciences*. EMNLP Keynote Proceedings.
10. García-Méndez, S., de Arriba-Pérez, F., & Costa-Montenegro, E. (2025). *Advancements in natural language processing, semantic networks, and sentiment analysis*. *Applied Sciences*, 15(12), 6476. https://doi.org/10.3390/app15126476
11. Giulianelli, M., Luden, I., Fernandez, R., & Kutuzov, A. (2024). *Interpretable word sense representations via definition generation*.
12. Honnibal, M., & Montani, I. (2017). SpaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *To appear*.
13. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of the International Conference on Learning Representations (ICLR)*.
14. Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. *EMNLP*, 1532–1543.
15. Salim, M. S., & Rafail, S. E. (2024, April). Ecological New Words: Linguistic Realization of the Concept “Environment” in The Language. In *International Conference on Smart Environment and Green Technologies* (pp. 225-232). Cham: Springer Nature Switzerland.
16. Schick, T., & Schütze, H. (2021). Exploiting cloze-questions for few-shot text classification and natural language inference. *EMNLP 2021*, 1635–1652.
17. Schmid, H. J. (2018). *The dynamics of the linguistic system: Usage, conventionalization, and entrenchment*. Oxford University Press.
18. Singh, P., & Bagga, S. (2022). Detecting emerging words in social media using transformer-based embeddings. *Journal of Computational Linguistics Research*, 14(2), 112–128.
19. Wulff, S., & Grieve, J. (2020). Social media as a source of neologisms: Tracking lexical innovation across platforms. *Language Resources and Evaluation*, 54, 523–547.
20. Zhang, Y., & Wang, Y. (2023). Temporal dynamics of neologisms in online communities: A BERT-based approach. *Natural Language Engineering*, 29(1), 45–63.
21. Zhao, J., & Su, M. (2025). *Dynamic IoU threshold optimization using data-driven approaches*. NLP Conference Proceedings

Received: 08.10.2025
Accepted: 12.01.2026