Journal of Multimedia Information System VOL. 12, NO. 4, December 2025 (pp. 141-148): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2025.12.4.141

Deep Learning-Based Context-Aware NER for Neologism Detection Across Generations

Junhaeng Lee[1], Hyunyoung Oh[1], Young-myoung Kang[1*]

Abstract

The rapid proliferation of neologisms in online and social communication has intensified linguistic disparities across generations, posing challenges for timely and accurate neologism detection. Conventional dictionary- and rule-based methods struggle to capture the dynamic, context-dependent nature of evolving expressions, limiting their practical applicability. This study proposes a deep learning-based, context-aware Named Entity Recognition (NER) model designed to effectively identify neologisms and mitigate generational language gaps. The model leverages KoELECTRA, a pre-trained Korean language model optimized for contextual understanding, and is evaluated on large-scale datasets with varying entity distributions. Experimental results demonstrate that the proposed model achieves robust neologism detection performance, with an F1-score of 0.89 under real-world imbalanced conditions and up to 0.92 in controlled, uniform settings. Furthermore, the model exhibits balanced recognition across diverse entity types, confirming its applicability to both formal and informal language environments. These findings highlight the model's potential as a scalable solution for enhancing linguistic accessibility and supporting real-time adaptation to evolving language, particularly in bridging generational communication gaps.

Key Words: Deep Learning, Natural Language Processing, Named Entity Recognition, Korean Neologisms.

I. INTRODUCTION

In contemporary society, conflicts driven by national, racial, religious, and generational divides are escalating globally. Among these, generational conflict has become a critical social concern in South Korea. According to a 2025 national survey, nearly 80% of respondents identified generational conflict as a serious issue, primarily attributing it to communication barriers and collaboration difficulties across age groups [1]. Notably, these challenges extend beyond differences in value systems, originating from fundamental discrepancies in linguistic expression patterns and communication modalities between generations.

The proliferation of internet-based platforms and social media has significantly accelerated the emergence and diffusion of neologisms, which now serve as integral components of everyday communication, transcending conventional generational boundaries [2]. In the South Korean context, neologisms are extensively utilized across mass media, broadcasting, and public discourse, functioning as key linguistic elements that shape socio-cultural interaction [3]. However, intergenerational disparities in the recognition, comprehension, and utilization of neologisms persist, exacerbating inequalities not only in the transmission but also in the interpretation and accessibility of information [4]. Conventional approaches to neologism detection primarily rely on static dictionary-based methodologies, where unregistered word forms are identified by comparing word frequencies within a target corpus against pre-defined lexicons [5]. While such methods are effective in detecting neologisms with high linguistic stability, they exhibit significant limitations in capturing transient, context-sensitive, and rapidly evolving expressions characteristic of dynamic online language environments [6]. Moreover, these approaches lack adaptability to account for the real-time linguistic innovation cycles prevalent in digital communication spaces.

These methodological limitations directly impede efforts to address generational communication gaps. As neologisms function increasingly as context-dependent linguistic entities rather than stable lexical units, their effective detection requires understanding semantic roles within discourse structures—a capability beyond frequency-based lexical matching. Addressing this challenge necessitates approaches that can capture both contextual nuances and temporal dynamics of evolving expressions, particularly

Manuscript received August 20, 2025; Revised October 05, 2025; Accepted November 04, 2025. (ID No. JMIS-25M-08-041)

Corresponding Author (*): Young-myoung Kang, +82-31-467-8186, ykang@sungkyul.ac.kr
[1]Department of Computer Engineering, Sungkyul University, Anyang, Korea, hkhk9495@sungkyul.ac.kr, ohy414@sungkyul.ac.kr, ykang@sungkyul.ac.kr

within the rapidly shifting linguistic landscape of digital communication.

Accordingly, the present study defines two core limitations underlying existing methodologies: (1) the inability to accommodate the temporal volatility and context dependency inherent in neologism generation and dissemination; and (2) the amplification of intergenerational linguistic asymmetries, which impairs effective communication and perpetuates disparities in information accessibility. To overcome these limitations, we propose a deep learning-based Named Entity Recognition (NER) framework optimized for the detection of neologisms within context-sensitive environments. The proposed model is trained on a large-scale Korean web corpus [7], predominantly composed of news media content, and annotated with seven entity categories: Organization (OG), Artifact (AF), Civilization (CV), Field of Study (FD), Theory/Rule (TR), Term (TM), and Neologism (NW). The integration of neologism detection within a multi-entity recognition architecture enhances the model's generalizability and practical applicability across diverse linguistic settings, contributing to the mitigation of generational communication gaps.

To evaluate the robustness and performance consistency of the proposed model under varying data distributions, three dataset variants were constructed: Original, Balanced, and Uniform. The model architecture employs KoELECTRA [8], a Korean language-adapted version of the Transformer-based ELECTRA framework [9], pre-trained primarily on news corpora to optimize the recognition of structured linguistic patterns and context-dependent expressions. Experimental results indicate that the proposed model achieves a macro-averaged F1-score of 0.88 and a neologism-specific F1-score of 0.91 under the Balanced dataset configuration, demonstrating superior performance in detecting evolving and contextually grounded linguistic forms. These findings validate the model's capacity to address the limitations of conventional dictionary- or rule-based approaches and highlight its potential to facilitate intergenerational linguistic integration, thereby promoting equitable information access and enhancing efficiency in social communication.

II. RELATED WORK

2.1. Neologism Collection

Traditional approaches to neologism collection have primarily relied on extracting candidate terms from various online platforms, including internet communities, blogs, social media, and collaborative knowledge bases such as Wikipedia [4]. In this process, words that exhibited low frequency within the collected corpus and were absent from existing dictionaries were flagged as potential neologisms. These candidates were then compared against curated lexicons comprising well-defined terms to finalize the neologism list [10]. Through iterative refinement, existing dictionaries were either expanded or updated to incorporate newly identified expressions.

However, neologisms in contemporary digital environments exhibit inherently unstable life cycles, characterized by rapid emergence, dissemination, and disappearance. Their meanings and usage patterns are highly fluid, often varying across timeframes, online subcultures, and communities. Consequently, static lexicon- or rule-based methods struggle to effectively capture this linguistic volatility. Furthermore, neologisms frequently manifest in diverse forms—including low-frequency tokens, intentional misspellings, abbreviations, hybrid structures, and combinations with emojis—complicating their detection through standardized pattern recognition techniques. The semantic ambiguity or polysemy associated with many neologisms presents additional challenges for deep learning models, particularly concerning the acquisition of sufficient, high-quality annotated data required for accurate embedding learning [10].

2.2. Named Entity Recognition

Named Entity Recognition (NER) is a fundamental subtask in natural language processing (NLP) that involves identifying and classifying entities within unstructured text, including persons, locations, organizations, temporal expressions, and domain-specific entities such as chemical compounds or pharmaceutical products [11]. NER serves as a critical preprocessing step for a wide range of downstream NLP tasks, including question answering, information retrieval, coreference resolution, and topic modeling.

Initial research in NER predominantly utilized rule-based approaches, leveraging handcrafted semantic, syntactic, and pattern-based rules in conjunction with domain-specific dictionaries. While these methods achieved high precision within narrow domains, they suffered from limited scalability and low recall, primarily due to the incompleteness of lexicons and the rigidity of manually crafted rules [12].

The advent of deep learning has significantly advanced NER methodologies, with neural network-based models demonstrating superior performance over traditional rule-based systems [12]. In particular, the Bidirectional Long Short-Term Memory with Conditional Random Fields (BiLSTM-CRF) architecture has become a widely adopted framework, effectively capturing bidirectional contextual dependencies within sentences and improving sequential labeling accuracy [13].

As illustrated in Fig. 1, deep learning-based NER models generally consist of three core components: input

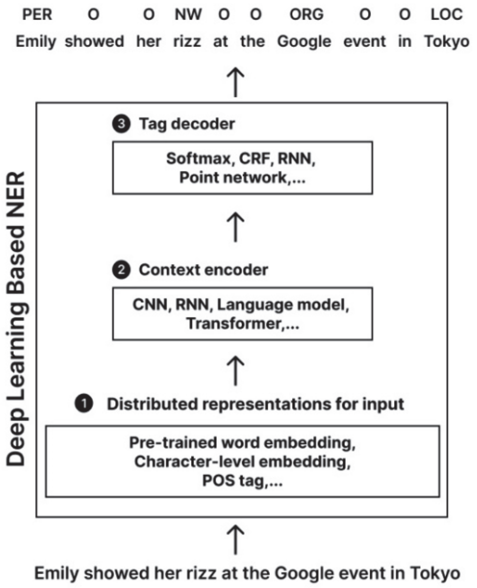

Fig. 1. Architecture of deep learning-based NER models.

representation, contextual encoding, and tag decoding [12]. The input representation layer generates high-dimensional vector representations of words by integrating multiple features, including word embeddings, character-level embeddings, part-of-speech (POS) tags, and external gazetteer information. The context encoding layer, typically implemented using Recurrent Neural Networks (RNNs) or Convolutional Neural Networks (CNNs), employs mechanisms such as Long Short-Term Memory (LSTM) or Bidirectional LSTM (BiLSTM) networks to effectively capture long-range dependencies within the text [13-15]. To further enhance sequence-level consistency in label predictions, a Conditional Random Field (CRF) layer is incorporated, modeling transition probabilities between tags [13].

More recently, the introduction of pre-trained language models and Transformer-based architectures has revolutionized NER by enabling context-aware representation learning without the need for manual feature engineering. These models, trained on large-scale corpora, acquire generalized linguistic knowledge that facilitates high-performance NER, even in scenarios with limited labeled data [16]. Their ability to effectively capture complex semantic relationships has significantly improved NER accuracy across diverse languages and domains.

III. METHODOLOGY

3.1. Data Set

The dataset utilized in this study is a large-scale, web-based Korean corpus comprising approximately one billion words, released in 2021. The corpus primarily consists of news articles spanning 17 thematic categories, including society, culture, science, and economics. To facilitate entity recognition tasks, the dataset was annotated with seven entity types: Organization (OG), Artifact (AF), Civilization (CV), Field of Study (FD), Theory/Rule (TR), Term (TM), and Neologism (NW). The annotation process was conducted using an automated authoring tool to ensure large-scale consistency. For neologism tagging, only newly emerging terms that appeared with a certain frequency but were absent from standardized Korean dictionaries were labeled accordingly.

Given the inherent ambiguity and inconsistency in the definition of neologisms in prior works, this study established explicit classification criteria to ensure consistency and clarity in neologism identification. Specifically, neologisms were strictly limited to the following categories:

- Compound expressions formed by the combination of two or more existing words,
- Hybrid terms resulting from partial word combinations or morphological blending,
- Loanwords or semantic extensions of pre-existing terms introduced into Korean.

Any expressions falling outside these categories were excluded from the neologism label set [17]. To maintain alignment between the refined neologism criteria and the overall dataset structure, a proportional downsampling strategy was applied, ensuring that the entity distributions within the dataset reflected the updated classification boundaries. The complete process of neologism selection, filtering, and downsampling is illustrated in Fig. 2.

An analysis of the dataset revealed significant distributional imbalance across entity types. In particular, Organization (OG) and Civilization (CV) entities dominated the corpus, which is attributable to the corpus's content composition, heavily focused on socio-cultural narratives, government entities, corporate organizations, and popular

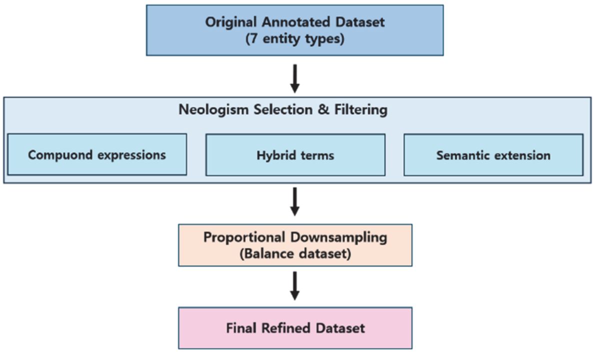

Fig. 2. Dataset refinement pipeline for neologisms.

culture. Conversely, entities associated with specialized or domain-specific expressions, such as Neologism (NW) and Theory/Rule (TR), were comparatively scarce, reflecting their lower natural occurrence within mainstream media texts. Table 1 presents a summary of the entity type distribution, illustrating both the original and post-filtering instance counts for each category.

These distributional patterns highlight the challenges associated with low-resource entity types such as neologisms, underscoring the importance of developing robust models capable of accurately detecting rare and context-sensitive expressions within imbalanced datasets.

3.2. Model

The widely used BERT model [18], based on a masking mechanism, conducts pre-training through the Masked Language Modeling (MLM) strategy, where a subset of tokens within an input sentence is replaced with special mask tokens, and the model is trained to predict the original tokens. While effective for general language understanding, this approach inherently suffers from two limitations. First, as training signals are provided only for the masked tokens, data utilization efficiency remains suboptimal. Second, masked tokens do not appear during the downstream fine-tuning or inference stages, resulting in a distributional mismatch between pre-training and task-specific deployment [9].

To address these limitations, the ELECTRA model introduces a more efficient pre-training paradigm known as Replaced Token Detection (RTD). ELECTRA employs a dual-Transformer architecture consisting of a Generator and a Discriminator. The Generator produces plausible replacements for randomly selected tokens, while the Discriminator performs token-level binary classification to distinguish between original and replaced tokens. Unlike MLM, this approach eliminates the need for artificial mask tokens, preserving the natural contextual flow of input sequences while enabling loss computation across all tokens. As a result, ELECTRA achieves superior learning efficiency and improved alignment between pre-training and downstream tasks. The overall model architecture and training mechanism are depicted in Fig. 3.

Table 1. Distribution of named entities before and after filtering.

Entity	Distribution (%)	Original count	Filtered count
OG	37.8	38,143,982	3,137,400
AF	6.1	6,153,459	522,900
CV	45.8	46,215,902	3,801,400
FD	1.6	1,614,704	132,800
TR	1.0	1,009,190	83,000
TM	7.6	7,669,844	630,800
NW	0.1	100,919	8,331

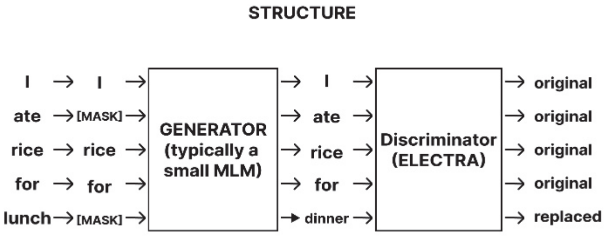

Fig. 3. ELECTRA pre-training architecture.

In this study, we adopted KoELECTRA, a Korean language-optimized variant of ELECTRA. KoELECTRA was pre-trained on a large-scale Korean corpus comprising approximately 14 GB of text, including diverse sources such as news articles, encyclopedic entries, and Namu Wiki content. This corpus composition emphasizes formal, structured language, enhancing the model's capacity for grammatical correctness and nuanced contextual understanding. Given the high domain similarity between the KoELECTRA pre-training corpus and the dataset used in this study, KoELECTRA was selected as the backbone model, providing a robust foundation for Korean-language entity recognition tasks.

Building upon this backbone, our study proposes an integrated training pipeline designed to enhance contextual understanding and task adaptability. The overall architecture of the proposed system is illustrated in Fig. 4.

3.3. Evaluation Metrics

In Named Entity Recognition (NER) tasks, model performance is primarily evaluated using the F1-score, which represents the harmonic mean of precision and recall, providing a balanced assessment of both error suppression and detection capability [19]. Precision quantifies the proportion of correctly identified entity mentions among all entity predictions made by the model. It reflects the model's ability to minimize false positives and is defined as:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Positives}}. \tag{1}$$

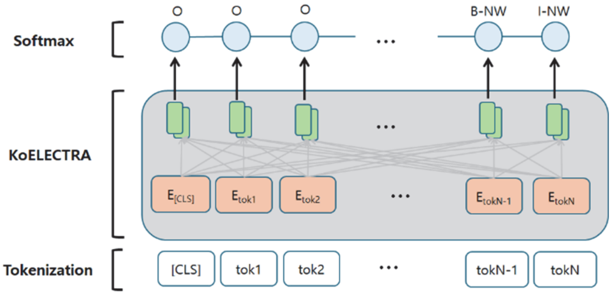

Fig. 4. Overall architecture of the proposed training pipeline.

Recall measures the proportion of actual entity mentions that the model correctly identifies, indicating the model's sensitivity and its ability to minimize false negatives. Recall is defined as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives}+\text{False Negatives}}. \quad (2)$$

The F1-score is calculated as the harmonic mean of precision and recall, providing a single metric that captures the trade-off between these two aspects:

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision}+\text{Recall}}. \quad (3)$$

In NER tasks, evaluation methodologies are typically classified into strict and relaxed criteria based on the matching granularity between predicted and ground truth entities. Under strict evaluation, a prediction is considered correct only if both the entity boundary (i.e., start and end positions) and the entity type precisely match the ground truth [20]. In contrast, relaxed evaluation allows partial matches or disregards entity type distinctions. We adopt a strict evaluation protocol in this study to rigorously assess the entity recognition performance. All reported F1-scores reflect strict boundary and type matching requirements, ensuring a high-precision evaluation of the proposed model's capability in accurately detecting and classifying entity mentions.

IV. EXPERIMENTAL RESULTS

4.1. Experimental Environment

All experiments in this study were conducted on a high-performance computing environment configured with Ubuntu 24.04.2 LTS as the operating system. The hardware setup included dual NVIDIA GeForce RTX 4090 GPUs, each equipped with 24 GB of VRAM, an Intel Core i9-10900X processor with 10 cores and 20 threads, and 125 GB system memory. The software environment consisted of NVIDIA driver version 550.144.03, CUDA 12.4, Python 3.8.20, PyTorch 2.4.1 (cu121 build), and the Transformers library version 4.46.3.

Model training was performed with the following configurations: batch size of 32, maximum sequence length of 128 tokens, and a learning rate of 2×10^{-5}. To mitigate overfitting, a dropout rate of 0.1 was applied, and the AdamW optimizer was used for parameter updates. An early stopping strategy was incorporated based on validation loss, with a patience threshold of three epochs. A summary of the hyperparameters is provided in Table 2.

Table 2. Hyperparameters used for model training.

Hyperparameter	Value
Batch size	32
Max sequence length	128
Epochs	100
Learning rate	2×10^{-5}
Dropout	0.1
Optimizer	AdamW
Early stopping patience	3 epochs

To analyze the impact of dataset size and entity label distribution on model performance, three dataset configurations were designed:

- Original Distribution: Maintains the natural distribution of filtered entities in the dataset.
- Balanced: Reduces overrepresented entity categories to alleviate class imbalance while preserving overall dataset size.
- Uniform Small: Enforces an equal number of samples for each entity category to facilitate fair comparative evaluation.

All datasets were partitioned into training, validation, and test sets using stratified sampling with an 80:10:10 split ratio to preserve entity distribution across subsets.

For text preprocessing, tokenization was performed using the KoELECTRA WordPiece tokenizer, which segments input sentences into subword units to effectively handle out-of-vocabulary expressions, including neologisms. Entity annotations followed the standard BIO tagging scheme, where B- denotes the beginning of an entity, I- represents tokens inside an entity span, and O indicates non-entity tokens. The detailed dataset statistics for each configuration are summarized in Table 3.

4.2. Results

Table 4 presents the F1-scores obtained by the proposed NER model across three dataset configurations: Original Distribution, Balanced, and Uniform Small. Both macro-

Table 3. Overview of dataset composition.

Entity	Balanced	Original	Uniform
OG	100,000	3,137,400	8,331
AF	100,000	522,900	8,331
CV	100,000	3,801,400	8,331
FD	100,000	132,800	8,331
TR	100,000	83,000	8,331
TM	100,000	630,800	8,331
NW	8,331	8,331	8,331
Total	608,331	8,316,631	58,317

Table 4. F1-scores by entity type across dataset configurations.

Entity	Balanced F1	Original F1	Uniform F1
OG	0.88	0.91	0.77
AF	0.84	0.81	0.70
CV	0.90	0.92	0.82
FD	0.92	0.87	0.86
TR	0.85	0.76	0.81
TM	0.85	0.83	0.74
NW	0.91	0.89	0.92
Micro average	0.88	0.90	0.79
Macro average	0.88	0.86	0.80

averaged and micro-averaged F1-scores are reported to provide a comprehensive evaluation of overall performance. For ease of comparison, Fig. 5 visualizes the results as a bar graph, illustrating performance variations across entity types and dataset conditions.

The experimental results provide several key insights into the influence of dataset composition and distribution on the performance of deep learning-based NER models, particularly for low-resource entity types such as neologisms.

First, the model demonstrated strong performance in the Original Distribution configuration, especially for dominant categories such as Civilization (CV) and Organization (OG), where the availability of extensive training data contributed to high F1-scores of 0.92 and 0.91, respectively. This observation aligns with well-established findings in NER research that model performance correlates positively with sample abundance and linguistic diversity within the training set. However, this performance advantage for high-frequency entities comes at the cost of recognition fairness, as reflected in the relatively lower F1-score of 0.76 for the Theory/Rule (TR) category, which suffers from data scarcity.

Second, the Balanced Dataset results highlight the trade-off between absolute performance and class-wise consistency. By limiting the maximum sample size per entity type to 100,000, the model exhibited stable performance across categories, achieving comparable F1-scores for both high-resource and low-resource entities. The Neologism (NW) category, in particular, achieved a notable F1-score of 0.91, suggesting that the proposed model effectively generalizes to lexically diverse and context-sensitive expressions even under balanced data conditions. These results indicate that mitigating class imbalance enhances overall model robustness, addressing limitations commonly observed in real-world NER applications where entity frequency distributions are highly skewed.

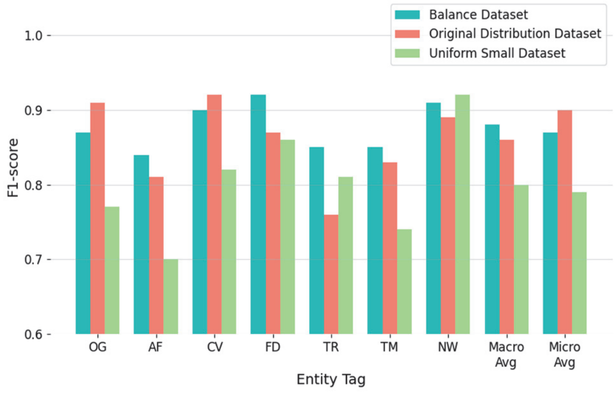

Fig. 5. Comparison of F1-scores across three dataset configurations.

In contrast, the Uniform Small Dataset configuration, which imposed strict sample count uniformity across all entity types, revealed limitations in the model's ability to maintain high recognition performance, particularly for previously dominant categories. The decline in F1-scores for OG, AF, and TM to 0.77, 0.70, and 0.74, respectively, underscores the critical role of data volume in enabling accurate entity recognition, especially for linguistically complex or syntactically variable expressions. Interestingly, the NW category achieved its highest F1-score of 0.92 under this setting, which, upon closer examination, is likely attributable to reduced data diversity rather than true generalization improvement. This result suggests that performance gains in low-resource conditions may not directly translate to real-world robustness, highlighting the need for careful dataset design that balances diversity and fairness.

A closer examination of the evaluation metrics reveals distinctive patterns in neologism recognition. While most low-frequency categories such as TR and TM exhibited performance instability across configurations, the NW category maintained consistently high F1-scores of 0.89, 0.91, and 0.92 across the Original, Balanced, and Uniform Small datasets, respectively. Although the slight increase observed in the Uniform Small setting warrants careful interpretation given the constraints of reduced data diversity, the overall stability suggests that the model effectively leverages contextual features for neologism recognition. Furthermore, the convergence of macro- and micro-averaged F1-scores at 0.88 in the Balanced Dataset demonstrates that equitable treatment across entity types can be achieved without sacrificing overall accuracy.

Overall, these findings emphasize the importance of dataset composition in optimizing NER performance, particularly for specialized categories such as neologisms, which are characterized by high variability, short life cycles, and low-frequency occurrence. The results also validate the proposed model's ability to adapt to both imbalanced and balanced data environments, demonstrating practical applicability in real-world scenarios where linguistic resources are often unevenly distributed.

In future work, incorporating data augmentation techniques, semi-supervised learning, or continual learning frameworks may further enhance recognition performance for low-resource entity types while preserving fairness across diverse linguistic categories.

V. CONCLUSION

In this paper, we proposed a deep learning-based NER model capable of effectively identifying neologisms, which have traditionally posed significant challenges for dictionary- or rule-based detection approaches. By leveraging context-aware learning mechanisms, the model demonstrated robust recognition performance for neologisms characterized by lexical diversity, short life cycles, and context dependency—properties that hinder traditional linguistic resources from timely adaptation. Furthermore, the model exhibited stable prediction performance across various entity types, confirming its applicability beyond formal domains such as news and media, and extending to dynamic environments where neologisms frequently emerge, including everyday language use and online platforms.

The experimental results validate the potential of our proposed model as a technical foundation for enhancing neologism detection systems, thereby contributing to the timely recognition of evolving linguistic expressions. Such advancements are crucial for narrowing generational language gaps and improving equitable access to information in a rapidly evolving linguistic landscape.

Future research directions include extending the model to handle informal, conversational, and colloquial language, where neologisms tend to emerge most actively and unpredictably. Adapting the model for application in online communities, social media interactions, and daily interpersonal communication will be essential to ensure practical, real-world effectiveness. Moreover, incorporating techniques such as domain adaptation, continual learning, or semi-supervised approaches may further enhance the model's ability to generalize to diverse linguistic settings.

REFERENCES

[1] Korea Research, 2025 Generation Perception and Intergenerational Conflict survey: Weekly Report no. 322-1, 2025. https://hrcopinion.co.kr/en/archives/32460.

[2] Z. R. Sultanova, “Neologisms in the modern world: Language that changes reality,” *Pedagogik Tadqiqotlar Jurnali*, vol. 3, no. 1, pp. 188-193, 2025.

[3] I. Aziza and Z. Aktamova, “Modern neologisms and their impact on society,” *International Journal of Multidisciplinary Research and Growth Evaluation*, vol. 6, no. 2, pp. 491-494, 2025.

[4] S. T. Vacalares, A. F. R. Salas, B. J. S. Babac, A. L. Cagalawan, and C. D. Calimpong, “The intelligibility of internet slangs between millennials and Gen Zers: A comparative study,” *International Journal of Science and Research Archive*, vol. 9, no. 1, pp. 400-409, 2023.

[5] K. I. Nam, New Words of 2019 Investigation, National Institute of Korean Language, Seoul, Dec. 2019. https://www.korean.go.kr/front/reportData/reportDataView.do?mn_id=207&report_seq=1104.

[6] I. H. Kim, “Neologisms in the era of big data,” *Saegugeosaenghwal*, vol. 29, no. 3, pp. 55-67, Sep. 2019.

[7] National Information Society Agency, AI Hub: Large Web Data-based Korean Corpus Dataset, AI Hub, 2022. https://aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=data&dataSetSn=624.

[8] J. Park, KoELECTRA: Pretrained ELECTRA Model for Korean, GitHub Repository, 2020. https://github.com/monologg/KoELECTRA.

[9] K. Clark, M. T. Luong, Q. V. Le, and C. D. Manning, “ELECTRA: Pre-training text encoders as discriminators rather than generators,” *arXiv Preprint arXiv:2003.10555*, 2020.

[10] J. H. Kim, H. C. Kim, and Y. S. Choi, “Feature generation of dictionary for named-entity recognition based on machine learning,” *Journal of Information Management*, vol. 41, no. 2, pp. 31-46, 2010.

[11] V. Yadav and S. Bethard, “A survey on recent advances in named entity recognition from deep learning models,” *arXiv Preprint arXiv:1910.11470*, 2019.

[12] J. Li, A. Sun, J. Han, and C. Li, “A survey on deep learning for named entity recognition,” in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp. 50-70, Jan. 2022.

[13] Z. Huang, W. Xu, and K. Yu, “Bidirectional LSTM-CRF models for sequence tagging,” *arXiv Preprint arXiv:1508.01991*, 2015.

[14] X. Ma and E. Hovy, “End-to-end sequence labeling via bi-directional LSTM-CNNs-CRF,” in *Proceedings of the 54th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Berlin, Germany, Aug. 2016, pp. 1064-1074.

[15] A. Sherstinsky, “Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network,” *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020.

[16] M. Munnangi, “A brief history of named entity recognition,” *arXiv Preprint arXiv:2411.05057*, 2024.

[17] Y. Li, “Sources of English neologisms,” *Linguistics and Education*, vol. 42, no. 1, pp. 210-215, 2024.

[18] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *arXiv Preprint arXiv:*

1810.04805, 2018.

[19] A. Mansouri, L. S. Affendey, and A. Mamat, “Named entity recognition approaches,” *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp. 339-344, Feb. 2008.

[20] E. F. Tjong Kim Sang, and F. De Meulder, “Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition,” *arXiv Preprint arXiv:cs/0306050*, 2003.

AUTHORS

Junhaeng Lee is currently pursuing the B.S. degree in Computer Engineering at Sungkyul University, Korea. His research interests include natural language processing, AI agents, and deep learning.

Hyunyoung Oh is currently pursuing the B.S. degree in Computer Engineering at Sungkyul University, Korea. His research interests include natural language processing, large language models, AI agents, and deep learning.

Young-myoung Kang is an Assistant Professor in the Department of Computer Engineering at Sungkyul University, Korea, and also serves as the Head of the Sungkyul Artificial Intelligence Convergence Institute. Prior to joining Sungkyul University in 2021, he worked for nine years as a Principal Engineer at Samsung Electronics. He received his Ph.D. and M.S. degrees from Seoul National University in 2013 and 2003, respectively, and his B.S. degree from Gyeongsang National University in 2000. From 2003 to 2006, he worked as a Software Engineer at LG Electronics. His research interests include wireless and wired networks, Wi-Fi, Bluetooth, communication systems, IoT, artificial intelligence, and deep learning. He also serves as a professional reviewer for numerous international journals and conferences, including IEEE Transactions, Human-centric Computing and Information Sciences (HCIS), Journal of Sensors, and The Journal of Supercomputing.