UDC 004.89:811.111

S. Y. Pidopryhora
V. V. Chopliak

NATURAL LANGUAGE PROCESSING ARCHITECTURES FOR CONTEXTUAL ERROR DETECTION IN ENGLISH ACADEMIC WRITING

Vinnytsia National Technical University

Abstract

The paper investigates the implementation of Natural Language Processing (NLP) technologies for automated error detection in English academic discourse. While traditional rule-based algorithms are limited to basic syntactic and lexical corrections, modern transformer-based architectures allow for deep semantic analysis and contextual disambiguation. This study explores the theoretical transition from heuristic spell-checkers to advanced neural networks capable of understanding English linguistic nuances. The research highlights how these computational tools enhance the quality of academic writing for non-native speakers by identifying complex contextual anomalies.

Keywords: natural language processing, English academic writing, contextual error detection, transformer architectures, computational linguistics.

Анотація

У роботі досліджено впровадження технологій обробки природної мови (NLP) для автоматизованого виявлення помилок в англомовному академічному дискурсі. Тоді як традиційні алгоритми, засновані на правилах, обмежені базовими синтаксичними та лексичними виправленнями, сучасні архітектури на базі трансформерів дозволяють проводити глибокий семантичний аналіз та контекстуальну дизамбігуацію. У дослідженні розглядається теоретичний перехід від евристичних систем перевірки орфографії до передових нейронних мереж, здатних розуміти лінгвістичні нюанси англійської мови. Робота підкреслює, як ці обчислювальні інструменти підвищують якість академічного письма для неносіїв мови шляхом виявлення складних контекстуальних аномалій.

Ключові слова: обробка природної мови, англомовне академічне письмо, контекстуальне виявлення помилок, архітектури трансформерів, комп'ютерна лінгвістика.

Introduction

In the era of globalized scientific communication, English serves as the primary medium for disseminating technical and academic knowledge. For non-native speakers, producing high-quality English academic texts presents significant challenges, particularly regarding lexical precision, stylistic consistency, and contextual coherence. Historically, software engineers addressed this issue by developing text-processing software that relied on hard-coded grammatical rules and static dictionaries. However, the intricacies of the English language – such as polysemy, idiomatic expressions, and domain-specific terminology – often bypass traditional algorithmic detection. The advent of Natural Language Processing (NLP) and machine learning has fundamentally transformed computational linguistics, enabling software systems to analyze English text at a semantic level rather than merely a syntactic one [1].

Limitations of rule-based correction systems

First-generation proofreading software utilized Boolean logic and heuristic algorithms to identify spelling and basic grammatical errors. These systems compare inputted text strings against a predefined database of English vocabulary and grammatical templates. While effective for identifying typographical errors or obvious subject-verb agreement issues, rule-based systems lack contextual awareness. For instance, a traditional algorithm cannot differentiate between "affect" and "effect" or "principal" and "principle" if both words are spelled correctly but used in the wrong semantic context. This limitation results in a high rate of false positives and overlooked contextual anomalies, rendering such tools insufficient for rigorous academic writing [2].

The shift to transformer-based NLP architectures

The paradigm shift in error detection occurred with the introduction of deep learning models, specifically transformer architectures such as BERT (Bidirectional Encoder Representations from Transformers) [3].

Unlike sequential processing algorithms, transformers analyze the entire English sentence simultaneously, evaluating the relationship and weight of each word relative to its surrounding context. This bidirectional analysis allows the system to perform contextual disambiguation.

In practice, when analyzing English academic discourse, neural networks generate high-dimensional word embeddings. If a user writes a sentence with an inappropriate lexical choice that technically fits the grammatical structure, the NLP model detects a statistical anomaly in the contextual embedding and suggests a more scientifically appropriate synonym. This level of computational analysis mimics human editorial intuition.

Research methodology

This theoretical study incorporates a systematic review of contemporary literature regarding NLP applications in ESL (English as a Second Language) education. The methodology involves analyzing the architectural differences between traditional syntax parsers and modern Large Language Models (LLMs) used in text generation and correction. Furthermore, the research evaluates the impact of these technologies on mitigating cognitive load for students and researchers drafting technical documentation in English [4].

Future perspectives

Future advancements in computational linguistics are expected to focus on domain-specific fine-tuning of NLP models. While current systems excel in general English academic writing, adapting these neural networks to understand the highly specialized jargon of software engineering, quantum physics, or biotechnology remains a critical challenge. Additionally, further research is required to address the ethical implications of AI-assisted writing, ensuring that these computational tools serve to enhance user proficiency rather than completely replacing the cognitive process of language acquisition.

Conclusion

The integration of Natural Language Processing into text analysis software represents a monumental leap in automated English language assessment. By transitioning from rigid, rule-based algorithms to context-aware neural networks, modern digital tools can detect complex semantic errors that previously required human intervention. For software engineering students and researchers, leveraging these advanced computational architectures not only improves the structural integrity of English academic publications but also facilitates a deeper understanding of linguistic nuances within technical discourse.

REFERENCES

1. Jurafsky, D., & Martin, J. H. (2023). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (3rd ed.). Pearson.
2. Leacock, C., Chodorow, M., Gamon, M., & Tetreault, J. (2014). Automated Grammatical Error Detection for Language Learners. Synthesis Lectures on Human Language Technologies, 7(1), 1-170.
3. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
4. Fitria, T. N. (2021). Grammarly as an AI-powered English Writing Assistant: Students' Alternative for Writing English. Metathesis: Journal of English Language, Literature, and Teaching, 5(1), 65-78.
5. Nykyporets, S. S., Kot, S. O., Sabadosh, J. H., Chopliak, V. V., & Piddubchak, S. Y. (2025). Leveraging digital technologies in English phraseology research. *Bulletin of Science and Education. Series: Philology, 10*(40), 83-95. https://doi.org/10.52058/2786-6165-2025-10(40)-83-95.
6. Nykyporets, S. S., Pradivlyanny, M. H., Boiko, Y. V., Chopliak, V. V., & Kukharchuk, H. V. (2024). Innovative techniques in vocabulary acquisition for foreign language learning: The impact of artificial intelligence. *Society and National Interests. Series "Education/Pedagogy",* 5(5), 113-127. https://doi.org/10.52058/3041-1572-2024-5(5)-113-127.
7. Nykyporets, S. S., Kot, S. O., Boiko, Y. V., Melnyk, M. B., & Chopliak, V. V. (2024). Advanced integration of virtual information environments (VIEs) in contemporary educational methodologies. *Society and National Interests. Series "Education/Pedagogy"*, 4(4), 139-154. https://doi.org/10.52058/3041-1572-2024-4(4)-139-154.

Svitlana Y. Pidopryhora – Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: sveta26pipodo@gmail.com.

Scientific Supervisor: ***Victoria V. Chopliak*** – teacher of English, Foreign Languages Department, Vinnytsia National Technical University, Vinnytsia, e-mail: nikavnuchkova@gmail.com.

Підопригора Світлана Юріївна – студентка групи ЗПІ-25б, факультет інформаційних технологій та комп'ютерної інженерії, Вінницький національний технічний університет, м. Вінниця, e-mail: sveta26pipodo@gmail.com.

Науковий керівник: ***Чопляк Вікторія Володимирівна*** – викладач англійської мови, кафедра іноземних мов, Вінницький Національний Технічний Університет м. Вінниця, e-mail: nikavnuchkova@gmail.com.