

A comprehensive review of the application of NLP technology in language learning

Jiale Peng

Wuhan University, School of Computer Science, Wuhan, 430072, China

canny626626@163.com

Abstract. The rapid development of Natural Language Processing (NLP) technology has provided new perspectives and tools for the acquisition of second languages. As our world becomes increasingly interconnected, the role of NLP in facilitating language learning has become more prominent. This paper reviews the multifaceted applications of NLP technology in language learning, including auxiliary teaching functions such as reading assistance, writing feedback, oral interaction, and personalized learning. These applications have significantly enhanced language learners' abilities in vocabulary acquisition, grammatical application, pronunciation accuracy, and reading comprehension through real-time feedback, enhanced interactivity, and promotion of cultural understanding. Despite the immense potential of NLP technology in second language learning, challenges such as technical accuracy, cultural adaptability, and data privacy exist. This paper proposes strategies, such as ensuring technical accuracy, curating diverse datasets and safeguarding data privacy to address these challenges. Looking forward, the future development of NLP technology in second language education holds great promise. As NLP continues to evolve, it is expected to contribute to more personalized, effective, and accessible language learning solutions. This paper aims to provide valuable references and insights for educators, technology developers, and policymakers, enabling them to harness the transformative potential of NLP in language education and to navigate the future of educational technology with confidence and foresight.

Keywords: Natural language processing, second language acquisition, technological application, educational innovation.

1. Introduction

Language learning is a multifaceted cognitive process that encompasses not only the mastery of vocabulary, grammar, and pronunciation but also a deep understanding of cultural contexts. In today's globalized world, proficiency in one or more foreign languages is crucial for personal career development and international communication. Traditional language learning methods, however, are often limited by the distribution of educational resources and the uniformity of teaching approaches, making it difficult to meet the personalized needs of learners. With the rapid development of artificial intelligence, Natural Language Processing (NLP) has brought revolutionary changes to language learning. NLP has many practical uses, such as categorizing text, identifying the tone of language, recognizing names in text, translating languages, and answering questions [1]. This paper explores the

application of NLP in language learning and analyzes its impact on educational practice, providing valuable insights for educators, technology developers, and policymakers.

2. Research background and current status

In Bax's [2] and Zheng et al.'s [3] review of three decades of the Computer Assisted Language Learning (CALL) development, the field has evolved through CALL (1990-1999), Mobile Assisted Language Learning (MALL) (2000-2009), and Intelligent Computer Assisted Language Learning (I-CALL) (2010-2020), each characterized by its socio-technical context [4]. The early stage was a time of conceptual exploration, with a focus on AI integration into CALL, despite the absence of empirical data in system evaluations. The mid-term phase saw the introduction of the hypermedia-internet-based English learning system by Lo et al. [5], which signified the essence of MALL and the shift towards Internet-based educational environments. System evaluations began using corpora, although quasi-experimental designs were not widely adopted until later.

The I-CALL era is distinguished by the transformative integration of intelligent AI functionalities in language teaching, with Natural Language Processing (NLP) occupying a central role. As a significant branch of artificial intelligence, NLP is dedicated to endowing computers with the capabilities to understand, interpret, and generate human language. NLP's interdisciplinary approach, which intersects Linguistics, Computer Science, and robotics [6], ensures that technological advancements are both pedagogically sound and attuned to the intricacies of language learning [7]. It facilitates adaptive and personalized learning experiences across various domains of language learning, such as reading, writing, and oral communication, providing tailored support that enhances comprehension and offers real-time pronunciation feedback. In essence, the I-CALL era encapsulates the convergence of empirical research and practical application, positioning NLP at the forefront of developing intelligent systems that address the diverse requirements of language learners.

3. Analysis of the application of NLP technology in language learning

3.1. Assisted reading

Reading is an essential pathway for acquiring information and improving language skills in language learning. The application of NLP in assisted reading provides robust support for learners.

One important application of NLP in assisted reading is text enhancement. Text enhancing tools within NLP can identify and interpret complex structures within texts, aiding learners in better understanding the content. In earlier years, CALL systems like SmartReader [8], WERTi [9], and FLAIR [10] leveraged NLP to facilitate reading and language learning through various text enhancements. To be specific, SmartReader offers a reading assistant that processes texts to highlight key information, providing definitions, grammatical insights, and encyclopedic data. While the WERTi system facilitates text enhancements and exercises for English, Spanish, and German, allowing users to interact with text by clicking, filling gaps, or writing words. It employs parsing and expressions to highlight seven linguistic structures. And FLAIR is an online information retrieval system that efficiently retrieves and ranks web documents based on grammatical constructions. It parses documents and recognizes a wide range of grammatical structures, offering users an annotated and highlighted text experience. Based on those former systems, Zilio et al. developed Smart and Intelligent Language Learning Environment (SMILLE) system [11], an innovative system that applies the Noticing Hypothesis and input enhancement techniques to address the subdued prominence of grammatical details in user-selected online content. Utilizing NLP, SMILLE highlights grammatical elements, potentially increasing the user's intake of metalinguistic information. It processes texts to bring in-context grammatical structures to the user's attention, provides glosses and definitions for vocabulary, and offers explanations for various grammatical forms.

Another important application is that NLP can provide authentic native language reading material for language learners [12]. The most popular Large Language Models (LLM) at present, the BERT model developed by Google and the ChatGPT developed by OpenAI, have powerful text generation

capabilities. Their ability to generate realistic dialogues, news, articles, or reading passages, can provide learners with authentic language use examples [13]. This can boost learners' reading and comprehension skills and provide exposure to authentic language material for improving language proficiency [14].

3.2. *Writing feedback*

Writing is a complex skill in language learning that involves not only a firm grasp of grammar and vocabulary but also the capacity to articulate thoughts clearly and coherently. Natural Language Processing (NLP) technologies have significantly enhanced the writing process by offering real-time grammatical checks, stylistic advice, and personalized feedback through automated scoring systems. These advancements are complemented by AI systems like Automated Writing Evaluation (AWE) and Intelligent Tutoring Systems (ITS), which provide comprehensive diagnostic insights into language use, facilitating learners' understanding and improvement. In the realm of AWE and Automated Essay Scoring (AES), NLP has been instrumental in analyzing language features and generating predictive models for essay scoring. Studies such as those by McNamara et al. [15] have applied hierarchical classification approaches to evaluate essays based on length and quality, demonstrating higher accuracy in score predictions through the use of specific thresholds. Alexopoulou et al. [16] utilized NLP to reveal that professional task writing, characterized by its structured format, tends to have lower error rates compared to narrative tasks.

Further research by Kyle and Crossley [17] has shown that fine-grained indices of phrasal complexity, extracted through NLP, are potent predictors of writing quality in TOEFL essays. Similarly, Vajjala [18] identified document length and discourse features as significant predictors in AES and AWE systems, respectively. Specific AI applications, such as the Genie Tutor system [19], have been designed to offer real-time grammar detection and expression suggestions, directly aiding language development. The EJP-Write ITS [20] supports academic writing by providing functionalities like reference citation and template searches, which are particularly beneficial for enhancing students' adherence to academic writing conventions.

3.3. *Oral interaction*

Oral communication stands as the most direct and natural form of interaction in language learning, an area where NLP has made significant inroads. The integration of speech recognition technology allows computers to accurately identify and understand learners' pronunciation, offering immediate and constructive feedback. This advancement has led to the development of dialogue systems that facilitate real-time conversation practice, thereby enhancing learners' fluency and practical language application. Furthermore, oral assessment tools powered by NLP analyze the pronunciation accuracy and naturalness of language use, providing learners with targeted suggestions for improvement. Building upon these capabilities, researchers like Ayedoun et al. have crafted conversational agents that emphasize communication strategies and affective backchannels [21]. These agents enable interactive practice, allowing learners to refine their conversational skills through dynamic questioning and by receiving responsive feedback from the AI system. Expanding on this interactive approach, Ruan et al. developed EnglishBot, an AI chatbot designed for language learning [22]. EnglishBot engages students in college-themed discussions, offering tailored feedback that is aimed at refining their linguistic proficiency. This bot represents the convergence of NLP advancements with educational technology, providing a platform for learners to practice oral communication in a simulated college environment. Li et al. introduce the Latent Dirichlet Integrated Deep Learning (LDiDL) framework for assessing student English proficiency [23]. It collects a diverse dataset of spoken English and extracts features like acoustics and linguistics. Using Latent Dirichlet Allocation (LDA), the system identifies underlying topics and integrates these with a deep learning model to analyze student performance. The model generates feedback on vocabulary, grammar, fluency, and pronunciation.

3.4. *Personalized learning*

Personalized learning has emerged as a pivotal trend in contemporary education, advocating for tailored educational experiences that cater to the distinct requirements and proclivities of learners. The incorporation of Natural Language Processing (NLP) has been instrumental in realizing this vision, enabling the creation of granular learner profiles that inform the recommendation of learning materials and the construction of bespoke learning trajectories. This approach has been shown to augment both the efficacy and the motivational engagement of learners. Chen et al. introduced the Personalized Intelligent Multimedia System (PIMS), an early application that curated English news articles to match learners' language proficiency, thereby mitigating cognitive overload and bolstering reading comprehension [24]. Advancing the integration of affective computing, Chao et al. developed an Affective Tutoring System sensitive to learners' emotional states, modulating the learning content to align with affective cues and support a more conducive learning atmosphere [25]. Pandarova et al. [26] furthered the precision of personalized learning with a system that dynamically adjusts the complexity of grammar lessons to accommodate varying language capabilities, fostering a self-paced learning environment. R. Vasile et al. [27] contributed to the field with Deep Tutor, an intelligent tutoring system that facilitates science learning through dialogue, providing immediate feedback and hints that promote independent problem-solving and significant learning gains. In a more recent development, FX. Risang Baskara et al. [28] have underscored the transformative potential of ChatGPT in language learning. This technology is positioned to deliver highly personalized instruction by aligning lesson plans with the unique educational needs and interests of individual students, thereby amplifying the effectiveness and engagement in language learning environments.

4. **Challenges and future prospects**

Despite the extensive potential of NLP in revolutionizing language learning, several challenges have surfaced that could impede its full integration and efficacy. Technical accuracy stands out as an urgent concern. NLP systems are tasked with the complex job of interpreting and generating human language, which requires a high degree of precision. The nuances of language, including idiomatic expressions, slang, and context-dependent meanings, pose significant hurdles for NLP algorithms. Cultural adaptability adds another layer of complexity. Language is deeply intertwined with culture, and NLP systems must navigate this relationship adeptly. The same words or phrases may carry different connotations and implications across cultures, and an NLP system's inability to recognize these differences could lead to misunderstandings or offenses. Data privacy is another critical challenge of the digital age, and NLP systems are no exception. As these systems collect and analyze vast amounts of data to improve their performance, they must do so while safeguarding user privacy.

To solve these problems, future work should focus more on ensuring technical accuracy, curating diverse datasets, and safeguarding data privacy. Ensuring technical accuracy involves developing hybrid models that integrate rule-based systems with machine learning to adeptly handle linguistic diversity and evolve with language use. Cultural adaptability is addressed by curating diverse datasets and consulting cultural experts to ensure NLP systems are sensitive and appropriate across various cultural contexts. Furthermore, safeguarding data privacy is crucial and is achieved by employing stringent cyber security practices and complying with global data protection regulations, such as GDPR, while maintaining transparent communication with learners about data usage policies. These concerted efforts are essential for NLP to fulfill its promise of enhancing personalized and effective language learning experiences.

With the continuous advancement of technology, the application of NLP in the field of language learning will become more extensive and in-depth. It is expected that future NLP technology will be more intelligent and personalized, providing more precise and efficient language learning support. At the same time, NLP technology will also drive innovation in educational models, achieving more flexible and open learning environments. Educators and technology developers need to work closely together to continuously explore new applications of NLP in language education to meet the growing learning needs.

5. Conclusion

This paper summarizes the extensive application of Natural Language Processing (NLP) technology in the field of second language learning and its profound impact on teaching and learning processes. NLP technology provides comprehensive language learning support for learners, significantly improving the efficiency of learning in various aspects such as vocabulary, grammar, pronunciation, and cultural understanding. Although the integration of NLP technology has brought many conveniences to second language learning, it also comes with challenges such as technical accuracy, cultural adaptability, and data privacy. To overcome these challenges, strategies including algorithm optimization, cross-cultural design, and data protection must be adopted to ensure the sustainability and effectiveness of the technology. Looking forward, as technology continues to evolve, it is expected that NLP will play an even more critical role in second language learning, promoting the development of personalized learning experiences and further enhancing learners' language abilities. Educators should actively integrate NLP technology into teaching practice, pay attention to the latest technological advancements, and improve their professional capabilities in this field to more effectively guide learners. Overall, the application of NLP technology in second language learning not only broadens the boundaries of education but also poses new educational challenges. By continuously optimizing technological applications and resolving existing problems, NLP technology has the potential to play a more important role in the future field of language education, providing learners with richer and more efficient learning pathways.

References

- [1] Rohit Kumar Yadav, Aanchal Madaan & Janu. (2024). Comprehensive analysis of natural language processing. *Global Journal of Engineering and Technology Advances*, 19(1), pp. 083-090.
- [2] Bax, S. (2011). Normalisation revisited: The effective use of technology in language education. *International Journal of Computer-Assisted Language Learning and Teaching*, 1(2), 1-15.
- [3] Zheng, P., Liang, X., Huang, G., & Liu, X. (2016). Mapping the field of communication technology research in Asia: Content analysis and text mining of SSCI journal articles 1995-2014. *Asian Journal of Communication*, 26(6), 511-531.
- [4] Jiacing Liang, Gwojen Hwang, Meirong Alice Chen, Darmawansah (2021): Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach, *Interactive Learning Environments*, DOI: 10.1080/10494820.2021.1958348.
- [5] Lo, J. J., Wang, H. M., & Yeh, S. W. (2004). Effects of confidence scores and remedial instruction on prepositions learning in adaptive hypermedia. *Computers & Education*, 42(1), 45-63.
- [6] Toprak-Yıldız, T. E. (2024). Natural Language Processing Applications in Language Assessment. *Advances in Educational Technologies and Instructional Design*, pp.216-234 .
- [7] Detmar Meurers. (2012). Natural language processing and language learning. *The Encyclopedia of Applied Linguistics*.
- [8] Mahmoud Azab, Ahmed Salama, Kemal Oflazer, Hideki Shima, Jun Araki, and Teruko Mitamura. (2013a). An English reading tool as a nlp showcase. In *The Companion Volume of the Proceedings of IJCNLP 2013: System Demonstrations*. Asian Federation of Natural Language Processing, Nagoya, Japan, pp.5-8.
- [9] Detmar Meurers, Ramon Ziai, et al. (2010). Enhancing authentic web pages for language learners. *Proceedings of the NAACL HLT 2010 Fifth Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics, pp.10-18.
- [10] Maria Chinkina and Detmar Meurers. (2016) Linguistically aware information retrieval: providing input enrichment for second language learners. In *Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications*, San Diego, CA.
- [11] Leonardo Zilio, Rodrigo Wilkens, C'edrick Fairon. (2017) *Proceedings of Recent Advances in Natural Language Processing*, Varna, Bulgaria, Sep 4-6, pp.839-846.

- [12] Detmar Meurers. (2012) Natural language processing and language learning. *The Encyclopedia of Applied Linguistics*.
- [13] George, A. S., & George, A. H. (2023). A Review of ChatGPT AI's Impact on Several Business Sectors. *Partners Universal International Innovation Journal*, 1(1), 9-23.
- [14] FX. Risang Baskara, FX. Mukarto. (2023) Exploring the Implications of ChatGPT for Language Learning in Higher Education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, Vol. 7(2).
- [15] McNamara, D. S., Crossley, S. A., Roscoe, R. D., Allen, L. K., & Dai, J. (2015). A Hierarchical classification approach to automated essay scoring. *Assessing Writing*, 23, 35-59.
- [16] Alexopoulou, T., Michel, M., Murakami, A., & Meurers, D. (2017). Task effects on linguistic complexity and accuracy: A Large - scale learner corpus analysis employing natural language processing techniques. *Language Learning*, 67(S1), 180-208.
- [17] Kyle, K., & Crossley, S. A. (2018). Measuring syntactic complexity in L2 writing using fine - grained clausal and phrasal indices. *The Modern Language Journal*, 102(2), 333-349.
- [18] Vajjala, S. (2018). Automated assessment of non-native learner essays: Investigating the role of linguistic features. *International Journal of Artificial Intelligence in Education*, 28(1), 79-105.
- [19] Lee, K., Kwon, O. W., Kim, Y. K., & Lee, Y. (2015). A Hybrid approach for correcting grammatical errors. *Critical CALL- Proceedings of the 2015 EUROCALL Conference*, Padova, Italy. Research-publishing.net, pp. 362-367.
- [20] Lin, C. C., Liu, G. Z., & Wang, T. I. (2017). Development and usability test of an e-learning tool for engineering graduates to develop academic writing in English: A Case study. *Educational Technology & Society*, 20(4), 148-161.
- [21] Ayedoun, E., Hayashi, Y., & Seta, K. (2019). Adding communicative and affective strategies to an embodied conversational agent to enhance second language learners' willingness to communicate. *International Journal of Artificial Intelligence in Education*, 29(1), 29-57.
- [22] Sherry Ruan, Liwei Jiang, Qian Yao Xu et al. (2021) EnglishBot: An AI-Powered Conversational System for Second Language Learning. *IUI'21*, April 14-17, College Station, TX, USA.
- [23] Li, W. & Mohamad, M.(2023). An Efficient Probabilistic Deep Learning Model for the Oral Proficiency Assessment of Student Speech Recognition and Classification. *International Journal on Recent and Innovation Trends in Computing and Communication*, pp. 411-424 .
- [24] Chen, C. M., Hsu, S. H., Li, Y. L., & Peng, C. J. (2006). Personalized intelligent m-learning system for supporting effective English learning. In *2006 IEEE International Conference on Systems, Man and Cybernetics*, Vol. 6, IEEE, pp. 4898-4903.
- [25] Chao, C. J., Lin, H. K., Huang, T. C., Hsu, K. C. & Hsieh, C. Y. (2012). The Application of affective tutoring systems (ATS) in enhancing learners' motivation. *Workshop Proceedings of the 20th International Conference on Computers in Education (ICCE)*, Asia-Pacific Society for Computers in Education, pp. 58-66.
- [26] Pandarova, I., Schmidt, T., Hartig, J., Boubekki, A., Jones, R. D., & Brefeld, U. (2019). Predicting the difficulty of exercise items for dynamic difficulty adaptation in adaptive language tutoring. *International Journal of Artificial Intelligence in Education*, 29(3), 342-367.
- [27] R. Vasile, N. Niraula, and R. Banjade. (2015) DeepTutor: an effective, online intelligent tutoring system that promotes deep learning, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, vol. 29, no. 1.
- [28] FX. Risang Baskara, FX. Mukarto. (2023) Exploring the Implications of ChatGPT for Language Learning in Higher Education, *Indonesian Journal of English Language Teaching and Applied Linguistics*, Vol. 7(2).