

Application of Natural Language Processing Technology in Foreign Language Teaching

Jiuzhang Chen^{1*}

¹School of European Languages and Culture, Beijing Foreign Studies University, Beijing, 100089, China

Abstract. This paper will summarize the development and current situation of the application of natural language processing technology in foreign language teaching. According to the comprehensive literature, the application of NLP in foreign language teaching has gradually upgraded from the initial application in writing assessment and grammar error correction to the key technology supporting personalized learning and resource generation. It is estimated that it will develop in the direction of human-machine collaboration in the future. This paper will first clarify the use of NLP in foreign language teaching scenarios such as automatic marking and personalized learning, and then analyze the problems that affect the actual effect of technology, such as feedback engagement and learner differences. Finally, it will discuss the technical challenges, ethical considerations and future directions in the context of the large language model, so as to provide reference for the realization of more efficient and responsible technologies in the future, and provide literature support for the development of foreign language teaching in the future.

1 Introduction

In today's digital era, as the core branch of artificial intelligence, natural language processing technology is profoundly reshaping the face of foreign language teaching. Since the Page team at Duke University in the United States developed the first intelligent essay evaluation system, PEG, in the 1960s, the application of NLP technology in foreign language teaching has undergone nearly six decades of development. From the initial rule-based grammar checking, to the error correction model based on statistical machine translation, to the current intelligent application based on deep learning and large language model, each iteration of NLP technology brings new possibilities for language teaching. The concept of technology enabled education has gradually moved from ideal to reality. Applications such as intelligent composition correction, personalized learning recommendation, and conversational language practice have entered the daily teaching scene.

However, the traditional foreign language teaching system has obvious bottlenecks in personalized training. On the one hand, under the large class teaching mode, it is difficult for teachers to provide timely and detailed feedback for each student's writing, and the feedback lag problem has plagued front-line teaching for a long time; On the other hand, there are significant differences in learners' language level, mother tongue background and learning style, and unified textbooks and teaching methods are difficult to meet personalized needs. Research in second language acquisition posits that mother tongue transfer constitutes a significant factor

contributing to learners' grammatical errors. Learners from varied mother tongue backgrounds tend to commit different types of mistakes, thereby further underscoring the imperative for personalized instruction. In addition, the limitation of teaching resources also restricts the improvement of learning effect, and high-quality teaching materials are difficult to achieve dynamic adaptation and instant generation. The existence of these problems calls for more intelligent and accurate technical solutions.

In recent years, a substantial amount of research has been undertaken by scholars both domestically and internationally concerning the application of NLP in foreign language instruction. In the field of writing feedback, Yang Yingli and Gao Zihan sorted out four turns in the field from external factors to internal psychology, from feedback effect to feedback process, from teacher-student feedback to digital intelligence empowerment feedback through the bibliometric analysis of foreign literature from 1996 to 2024 [1]. Tang Jinlan and Wu Yian reviewed the application research of online English writing automatic evaluation system, which laid a theoretical foundation for subsequent research [2]. In terms of the effectiveness of intelligent feedback, Liu Shujun and others verified the positive impact of intelligent composition feedback on middle school students' writing revision practice, composition quality and writing motivation through experiments [3]. In terms of personalized learning support, Zhang Shengsheng and others proposed personalized grammar error correction for Chinese learners, which significantly improved the error correction performance by adapting to the level of learners and the domain of their mother tongue [4].

* Corresponding author: Chenjz202@outlook.com

Nadejde and Tetreault's research on English learners shows that the system performance is significantly improved after the grammar error correction model is adapted to the learners' native language and level [5]. In the field of international Chinese education, Diao Jiangzhi analyzed the enabling role of AI technology in the generation of teaching resources, classroom interaction and teacher professional development [6]. In terms of the study of learners' cognitive mechanism, Zheng and Yu discussed the cognitive, behavioral and emotional dimensions of feedback engagement in second language writing through multiple case studies of low-level learners, providing a new perspective for understanding the effectiveness of feedback [7]. Boudouaia and others focused on the promoting effect of generative AI on English writing learning, and compared the effect difference between AI feedback and teacher feedback [8]. These studies reveal the application value and potential challenges of NLP technology in foreign language teaching from different angles.

To sum up, the application of NLP technology in foreign language teaching has made remarkable progress. However, most of the existing studies focus on a single technology or specific scenario, and lack the integration of the evolution path from "automatic correction" to "intelligent collaboration". The comprehensive analysis of key factors affecting the effectiveness of technology, such as feedback engagement, personalized adaptation, human-machine collaboration, is still insufficient. The new opportunities and challenges in the era of large language model also need to be sorted out. Based on these twelve documents, this paper aims to comprehensively review the application evolution, core influencing factors and future trends of NLP in foreign language teaching, in order to provide theoretical reference and practical enlightenment for the further development of technology enabled foreign language teaching.

2 The progression of core applications in NLP-empowered foreign language instruction

2.1 Automated assessment and correction

Represented by the automatic writing evaluation system and grammar error correction tool, NLP technology is the first to realize the real-time and large-scale feedback of language forms. Foreign countries have developed PEG and other systems since the 1960s, while China has taken the introduction of foreign scoring systems by Liang Maocheng as the starting point, and the related research has been continuously deepened in the past two decades [9]. In English teaching, after the integration of intelligent composition feedback, students' final writing scores were significantly improved compared with the first draft, that is, the three paired t-test P values were <0.01; The number of words in the composition increased from an average of 515 words to 708 words, that is, $t=7.47$, $p<0.01$. The frequency of the use of stylistic elements such as argument and interpretation

has increased significantly [3]. In the teaching of Chinese as a second language, the grammar error correction model based on transformer has also achieved good results on the Lang-8 and HSK data sets [4].

NLP technology, as an auxiliary tool, mainly replaces manual completion of repetitive and low-level correction tasks, and realizes the immediacy and large-scale coverage of feedback. Its advantage is to significantly reduce the burden on teachers, provide objective and consistent scores, and stimulate students' writing revision motivation through instant feedback. The system's ability to assess high-level language abilities such as content, structure and logic is limited, and the feedback often stays at the lexical and grammatical levels, which is difficult to provide in-depth writing guidance; In addition, the early system pays insufficient attention to the individual differences of learners, and the feedback is lack of pertinence, so students are prone to aesthetic fatigue or dependence.

2.2 Personalized learning support and resource generation

The focus of technology has gradually shifted to the learner model. The study found that learners at different levels responded differently to automatic feedback. Based on the ZPD theory, it is found that students in the low-level group improve their scores more significantly with the help of automatic feedback, while there is no significant change in the high-level group [9]. Zhang Shengsheng and others proposed personalized grammar error correction for Chinese learners, which significantly improved the performance of the model by adapting to the learners' level and native language. It exceeded the random baseline by 3.59 percentage points on level A and 4.98 percentage points on native English speakers in Singapore [4]. In the field of international Chinese education, AI technology is also used to generate multimodal teaching resources and create personalized learning paths, such as generating picture books and pictures through Doubao and generating test papers using ChatGPT [6].

Technology is no longer satisfied with general feedback, but attempts to provide customized support according to learners' language level, mother tongue background, learning style and other characteristics, and meet individual needs by generating diversified learning resources. Its advantage is that it can more accurately diagnose learners' problems and provide learning materials that are in line with their recent development zone, so as to improve learning efficiency and motivation. But the challenge is also outstanding. The construction of personalized model needs large-scale annotation data and complex algorithm support, and the problem of domain adaptation has not been completely solved. At the same time, excessive personalization may lead learners to fall into the information cocoon and lack of contact with the multilingual environment. In addition, the automation and quality assurance of resource generation still need to be further optimized.

2.3 Towards intelligent collaboration and process support

The current research is no longer limited to the outcome evaluation, but pays more attention to the writing revision process and the learner interaction. According to the analysis of 253 modification points by Liu Shujun et al., the most common modification behaviors used by students were increase, accounting for 55.73%, and replacement, accounting for 33.60%. The proportion of low-order modifications was 60.38%, which was higher than that of high-order modifications. The modification success rate of students responding to intelligent feedback is 93%, which is higher than 81% of autonomous modification [3]. Yang Yingli and Gao Zihan found through bibliometrics that the research on feedback in L2 writing abroad has experienced four turns: from external factors to internal psychology, from feedback effect to feedback process, from teacher-student feedback to digital intelligence and empowerment feedback. Zheng & Yu, Zhang & Hyland and other key literatures all focus on learners' multidimensional engagement in feedback [1,7,10].

Technology has changed from a one-way feedback provider to a participant in the teaching process, emphasizing the support for learners' revision process, and exploring the formation of a multi collaborative feedback ecology with teachers and peers. Its advantage is that it can track the learning process, provide process support, and stimulate deep cognitive engagement through interaction with learners. However, the support of current technology for high-level thinking ability, such as critical thinking and creative expression, is still insufficient. The collaborative mechanism between intelligent system and teachers and peers is not yet mature. How to effectively integrate multiple feedback and avoid information overload is an urgent problem to be solved. In addition, the mechanism of learners' affective and motivational factors in the collaborative process needs further study.

3 The core factors affecting the effectiveness of NLP teaching: feedback, personalized adaptation and collaboration

3.1 Depth of feedback and learner engagement

The effectiveness of technical feedback depends not only on its accuracy, but also on learners' cognition, behavior and emotional engagement. Zhang and Hyland compared teacher feedback with automatic feedback and found that the two learners' engagement levels were quite different due to individual differences, and the source of feedback and learner factors were the important reasons for the difference in engagement [1]. Learners with higher second language proficiency will question the content of automatic feedback, showing a higher level of cognitive engagement, while learners with lower second language proficiency rely excessively on feedback. Liu Shujun and others found that 80% of the students affirmed the role of intelligent scoring in

promoting writing, but the quality of feedback is the key. Specific and accurate feedback can lead to effective modification, while general suggestions are difficult to implement [3].

Feedback engagement is a mediator between technology output and learning effectiveness. Its core lies in how learners perceive, process and respond to feedback information. The advantage is that when learners can be deeply engaged, even simple error correction can trigger deep learning transfer. However, its engagement level is restricted by many factors, such as individual differences like learners' language level, feedback characteristics, learning situation and so on. The deficiency of current research lies in the insufficient attention to the dynamic changes and long-term effects of engagement, and the lack of effective intervention strategies to improve learners' engagement level.

3.2 Personalized challenges and implementation path

The real personalization needs to overcome the problem of domain adaptation. Zhang Shengsheng et al. conducted experiments by constructing data sets with different mother tongues and grades, and finally found that after the model was adapted to learners' grades, the average F0.5 score increased from 42.52 at the random baseline to 44.44; After adapting to the mother tongue, it increased from 39.66 to 41.39; After adapting to the mother tongue-grade combination, it increased from 44.70 to 46.46. This shows that the domain adaptation model integrating learners' multidimensional characteristics can significantly improve the error correction effect [4]. Song Hua's review also pointed out that the automatic evaluation system has different effects on students of different levels, and it needs to be expanded to more learning stages and groups in the future [9].

Personalization is the key path to achieve precision teaching, and its technical core lies in domain adaptation and multi-dimensional feature fusion. Its advantage is that it can significantly improve the adaptability and error correction effect of the model, and narrow the gap between learners at different levels. Personalization also faces multiple challenges. The acquisition cost of high-quality, multi-dimensional annotation data is high; It may exacerbate the digital divide; It is difficult for learners with scarce resources to enjoy the same quality of service. In short, the balance between personalized and standardized teaching needs to be carefully grasped to avoid over catering to individuals and ignoring the cultivation of common literacy.

3.3 Colour illustrations

People are free to use colour illustrations for the online version of the proceedings, but any print version will be printed in black and white unless special arrangements have been made with the conference organiser. Please check whether this is the case. If the print version will be black and white only, people should check their figure captions carefully and remove any reference to

colour in the illustration and text. In addition, some colour figures will degrade or suffer loss of information when converted to black and white, and this should be considered when preparing them.

4 Reconstruction of human-machine collaboration teaching mode

4.1 Technical challenges and ethical boundaries

The current technology still has limitations in the evaluation of high-level abilities such as discourse coherence and cultural connotation. Liu Shujun et al.'s feedback pointed out that the IN Classroom Intelligent Essay Evaluation System had the problems of exaggerated praise, general suggestions, and low recognition rate of complex grammar errors [3]. Diao Jiangzhi stressed that AI tools involve the collection and analysis of user data, which may lead to ethical disputes without proper privacy protection. He cited the "Framework for Teachers' AI Competence" and "Framework for Students' AI Competence" issued by UNESCO in 2024, calling for ensuring the safe and ethical use of AI technology by teachers and students [6]. Technological breakthroughs and ethical norms must be promoted together. Only under the premise of ensuring data security and algorithm fairness, can intelligent technology really be used continuously in the field of education.

4.2 New opportunities driven by LLM

Large language models bring new possibilities to foreign language teaching. Diao Jiangzhi specifically mentioned that ChatGPT can act as a language learning assistant, and Duolingo Max provides question-answering and role-playing functions based on GPT-4 [6]. Yang Yingli and Gao Zihan pointed out that the research hotspots in 2021-2024 include artificial intelligence and machine translation. Researchers are investigating the effect of generative AI empowering second language writing feedback, and comparing AI with teacher feedback, aiming to build a multiple collaborative model in the era of digital intelligence [1]. With the continuous iteration of large language models, human-machine collaboration will move from ideal to reality, opening up a broader space for personalized and immersive language learning.

4.3 Interdisciplinary research deepening

In the future, it is necessary to integrate the relevant theories and research methods of learning science and cognitive psychology. Yang Yingli and Gao Zihan called for the use of eye tracking, keylogging, brain imaging and other technologies to deeply explore the cognitive mechanism in the feedback process, and to carry out longitudinal tracking research on the basis of complex dynamic system theory [1]. The review by Meurers also highlights the cross-cooperation of NLP

with second language acquisition, corpus linguistics, and the support of automatic analysis for learners to measure language complexity, accuracy, and fluency [11]. Heift & Schulze's book laid the foundation for interdisciplinary research in the field of ICALL from an earlier perspective, pointing out that the design of the intelligent language learning system must consider the realizability of computing technology and the cognitive laws of second language learners [12]. This interdisciplinary communication will become an important way to deepen NLP foreign language teaching research in the future.

5 Conclusion

Based on 12 core literatures, this paper systematically reviewed the application evolution, core influencing factors and future trends of natural language processing technology in foreign language teaching. The research first combed the development of NLP application in foreign language teaching, from early automated assessment and correction, to personalized learning support and resource generation, and then to the current intelligent collaboration and process support, clearly showing the role evolution of technology from "auxiliary tool" to "teaching collaborator". On this basis, three core factors affecting the effectiveness of technology are analyzed, namely feedback engagement, personalized adaptation and human-machine collaboration. It is concluded that the effectiveness of technology not only depends on the performance of the algorithm, but also is closely related to the learners' cognitive, behavioral and emotional engagement, the accurate adaptation degree of the system to the learners' characteristics, and the teaching mode of teacher-machine-learner multiple collaboration.

The application of NLP technology in foreign language teaching has achieved remarkable results. Automatic evaluation system can improve writing performance, personalized error correction can optimize performance through domain adaptation, and AI technology plays a substantial empowering role in multimodal resource generation and classroom interaction. However, the effectiveness of technology is restricted by multiple factors, and the lack of learner engagement, the difficulty of personalized adaptation and the immaturity of human-machine collaboration mechanism are still the main challenges. In addition, there are still limitations in the assessment of higher-order language ability by technology, and ethical issues such as data privacy and algorithmic bias need to be regulated.

In the future, large language models will bring revolutionary opportunities for foreign language teaching, and generative AI has shown great potential in feedback generation and immersive scene creation. At the same time, the introduction of interdisciplinary research methods will deepen the understanding of the cognitive mechanism of learning. The value of technology ultimately depends on its deep integration with pedagogy and learners' needs. In the future, researchers should devote ourselves to developing

intelligent tools that are more accurate, more interpretable and more ethical. In the AI era, researchers should realize a new human-machine collaborative ecology of "giving creation to AI and leaving creativity to human beings", so as to promote the comprehensive development of learners as the ultimate educational goal.

References

1. Y. Yang, Z. Gao. Research Progress in L2 Writing Feedback Abroad: A Bibliometric Analysis Based on 1996-2024. *J. Beijing Int. Stud. Univ.*, (1), 127-141 (2025)
2. J. Tang, Y. Wu. A Review of Research on the Application of Online Automated Writing Evaluation Systems in English Writing. *Foreign Lang. Teach. Res.*, 43(2), 273-282, 321 (2011)
3. S. Liu, Y. Li, Y. He, J. Wang. Can Intelligent Essay Feedback Improve the Effect of Writing Teaching in Middle School? *J. East China Norm. Univ. (Educ. Sci.)*, (9), 36-50 (2022)
4. S. Zhang, G. Pang, L. Yang, C. Wang, Y. Du, E. Yang, Y. Huang. Personalized Grammatical Error Correction for Chinese as a Second Language. *J. Chin. Inf. Process.*, 35(12), 64-73 (2021)
5. M. Nadejde, J. Tetreault. Personalizing Grammatical Error Correction: Adaptation to Proficiency Level and L1. *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, Hong Kong, 27-33 (2019)
6. J. Diao. Analysis of the Application of Artificial Intelligence Technology in International Chinese Classes. *Adv. Educ.*, 15(3), 1-6 (2025)
7. Y. Zheng, S. Yu. Student engagement with teacher written corrective feedback in EFL writing: A case study of Chinese lower-proficiency students. *Assess. Writ.*, 37, 13-24 (2018)
8. A. Boudouaia, S. Mouas, B. Kouider. A study on ChatGPT-4 as an innovative approach to enhancing English as a foreign language writing learning. *J. Educ. Comput. Res.*, (6), 1509-1537 (2024)
9. H. Song. A Review of Domestic Research on English Writing Evaluation in the Past Decade: Based on Automated Writing Evaluation Systems. *Open J. Syst.*, (2025)
10. Z. Zhang, K. Hyland. Student engagement with teacher and automated feedback on L2 writing. *Assess. Writ.*, 36, 90-102 (2018)
11. D. Meurers. Natural Language Processing and Language Learning. In C. A. Chapelle (Ed.), *Concise Encyclopedia of Applied Linguistics.*, Wiley (2019)
12. P. ten Hacken. Review of Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues by T. Heift and M. Schulze. *Comput. Linguist.*, 33(4), 583-586 (2007)