TPM Vol. 32, No. 2, 2025
ISSN: 1972-6325
https://www.tpmap.org/



# APPLICATIONS OF ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING IN APPLIED LINGUISTICS: A SYSTEMATIC REVIEW OF METHODS, OUTCOMES, AND RESEARCH TRENDS

SADIA HAMEED[1], MUHAMMAD UMAR[2], MAHNOR SALEEM CHUGHTAI[3]

[1,2,3] MS ENGLISH LINGUISTICS, DEPARTMENT OF SOCIAL SCIENCES; MS ENGLISH LINGUISTICS, DEPARTMENT OF SOCIAL SCIENCES; WOMEN UNIVERSITY BAGH, AJK

EMAIL: sadiahameed421@gmail.com[1], mumar9400@gmail.com[2], mahnorsaleem116@gmail.com[3]

**Abstract**

AI and NLP are transforming applied linguistics, enabling deeper insights, more precise analyses, and innovative approaches to language learning and research. This systematic review examines the applications of Artificial Intelligence (AI) and Natural Language Processing (NLP) within the field of applied linguistics, focusing on research methods, empirical outcomes, and emerging trends. In addition, peer-reviewed studies were systematically identified, screened, and analyzed to evaluate how AI-driven tools by following PRISMA-guided procedures. It includes machine learning algorithms, automated text analysis, speech recognition, and intelligent tutoring systems. These tools have been employed across key applied linguistics domains such as second language acquisition, language assessment, discourse analysis, and corpus linguistics. The findings indicate that AI and NLP technologies significantly enhance linguistic analysis accuracy, personalize learning experiences, streamline assessment processes, and reveal complex patterns in language use. In conclusion, the study offer transformative potential for both research and pedagogical practices in applied linguistics.

**Keywords:** Artificial Intelligence, Natural Language Processing, Applied Linguistics, Language Assessment, Second Language Acquisition, Discourse Analysis

## 1. INTRODUCTION

The integration of Artificial Intelligence into applied linguistics is revolutionizing the way language is analyzed, taught, and assessed, offering unprecedented precision and insight into human communication. The concepts of Artificial Intelligence (AI) and Natural Language Processing (NLP) have quickly influenced the study and practice of language education and applied linguistics. The concept of AI suggests that computational systems can be used to perform tasks that are normally attributed to human intelligence, including pattern recognition, decision-making, and adaptation. NLP, a crucial subdivision of AI, focuses on the interaction between computers and human language, enabling machines to comprehend, generate, and analyze both written and spoken information. The emergence of machine learning, large language models (LLMs), and automated feedback systems has greatly expanded the ability of researchers and educators to process and analyze complex data in the area of linguistics at a micro and macro scale. According to recent systematic research, AI and NLP technologies are increasingly used to facilitate language teaching and assessments of learners (Fatemeh Esmaeilzadeh et al., 2020).

Applications of AI and NLP can be found in significant areas of application in applied linguistics. An intelligent agent, a chatbot with adaptive learning, or an interactive feedback system, which is an AI-powered tool, is applied to analyze language used by learners, facilitate practice, and simulate a communication situation in second language acquisition to improve learner interest and proficiency development (Wiboolyasarin et al., 2025). In language testing, automated scoring systems and NLP-based analytic models are applied in evaluating writing and speaking performances, which advance the argument on reliability and validity in assessment practices. Computational text analysis has the advantage of discourses and corpus linguistics, which can detect patterns at large scales and provide genre exploration, helping to understand the language use in more depth (Yuan et al., 2025). Additionally, computer-aided language learning (CALL) unites AI systems to offer adaptive learning scaffolds and support multilingual learning as well as real-time corrective feedback to create more personalized learning spaces (Gardazi et al., 2025).

This trend in AI/NLP is also indicative of a larger trend in applied linguistics of abandoning the traditional, manual style of analysis in favor of data-driven, scalable, and computational analysis. Conventional linguistic studies were based on small data sets and human coders, but current practices focus on automation, predictive analytics, and engagement with large datasets, making it possible to conduct more convincing empirical investigations into language behavior.

Nevertheless, the constantly growing literature is still disjointed in terms of tools, techniques, and linguistic areas, which indicates the necessity of synthesis. This paper hence presents a systematic review of empirical studies regarding AI and NLP applications in applied linguistics with the aim of mapping research practices, findings, and future directions in the field to inform future research and practice.

This review addresses the following research questions:

- **RQ1:** What research methods and AI/NLP techniques are used in applied linguistics research?
- **RQ2:** What empirical outcomes are reported across different applied linguistics domains?
- **RQ3:** What research trends and gaps characterize the application of AI and NLP in applied linguistics?

## 2. PREVIOUS REVIEWS

AI and NLP not only automate traditional linguistic analyses but also uncover patterns and relationships in language that were previously inaccessible to human researchers. Several systematic and scoping reviews have been conducted in the recent past to assess the applicability of AI and NLP technologies in language learning, education, and assessment. To illustrate, the review by Bao et al. was a systematic mapping of AI applications in second language acquisition (SLA) through an extended pedagogical framework to differentiate between enhancement and transforming uses of technology, which showed patterns in the areas of skills focus and technological integration of hundreds of studies (2015-2024). A second strand of research is more specific to AI-based conversational agents; the synthesis of new research described, however, demonstrates that chatbots have the potential to facilitate speaking and writing practice and have an implication on blended pedagogical approaches that combine human teaching with AI technologies (Wiboolyasarin et al., 2025).

In the area of assessment, a methodical process has started on AI-enabled assessment tools in creating metrics of their design, execution, and efficacy, and portrays distinctive medium-impact enhancements in the results of learning in the utilization of such instruments, along with diversity in education levels and the classes of tools (Zhai, 2024). Other reviews focus on the literature that is accumulating on automated evaluation of open-ended answers, providing frameworks through which text-based automated assessment systems are implemented and developed in institutions of higher learning (Gao et al., 2023). In parallel with these, larger literature reviews in the educational field have generalized the role of AI in fields and provided an account of its pedagogical capabilities, challenges, and gaps in evidence to apply it, but not necessarily with a high degree of applied linguistics in its focus (Mustafa et al., 2024).

Although these reviews have important information, they have a number of common methodological and thematic constraints. Most tend to target AI tools or applications in one form or another (chatbots or tutor systems) instead of the wider ecosystem of AI and NLP technologies in applied linguistics. Still others focus on one or two specific aspects or environments, including SLA or testing, without cross-cutting the results of language domains, instruments, and products. Besides, there is a lack of literature that provides an overall synthesis of methods, findings, and research trends, which creates gaps in our knowledge concerning the relationship and impact of a wide variety of AI/NLP technologies used in linguistic studies and learning.

The current review builds on this previous study in that it uses a comprehensive and systematic approach guided by PRISMA, bringing together all the existing applications of AI and NLP in applied linguistics to synthesize the empirical evidence of all methods, outcomes, and new trends and present a more holistic and research-based view of the phenomenon (Bao et al., 2025).

**Table 1:** Summary of Previous Reviews on AI and NLP in Applied Linguistics

| Review | Primary Focus | Scope | Key Limitations Identified |
|---|---|---|---|
| Bao et al. (2023) | AI applications in second language acquisition (SLA) | Systematic mapping of AI-supported SLA studies (2015–2024), emphasizing pedagogical enhancement vs. transformation | Concentrates on SLA only; limited integration of assessment, discourse analysis, and broader NLP methods |
| Huang & Spector (2021) | Intelligent tutoring systems and adaptive learning | AI-driven instructional systems across educational domains, including language learning | Broad educational focus; applied linguistics treated peripherally rather than analytically |
| Zhai et al. (2022) | AI-assisted language assessment | Automated scoring, feedback systems, and validity issues in language testing | Emphasis on assessment tools; minimal discussion of learning theory or discourse-level NLP |
| Khosravi et al. (2022) | Conversational agents and chatbots | Empirical studies on AI chatbots for speaking and writing practice | Tool-specific focus; lacks synthesis across AI/NLP techniques and linguistic domains |

| | | | |
|---|---|---|---|
| Chen et al. (2023) | AI in computer-assisted language learning (CALL) | Review of AI-enhanced CALL environments in higher education | Fragmented coverage; limited methodological comparison and weak theoretical grounding |
| García & Perea (2024) | NLP-based automated writing evaluation | Text analytics, feedback generation, and machine learning models | Focuses narrowly on writing; excludes multimodal, spoken, and corpus-based linguistic applications |

### 4.1 LITERATURE SEARCH

In this study, a systematic review design was used, and it was informed by the PRISMA 2020 statement, which offers current and generally accepted principles of transparent reporting of systematic reviews in the field of education and social sciences (Page et al., 2021). Identification, screening, eligibility assessment, and final selection of pertinent studies were done via PRISMA, which guaranteed methodological rigor and reproducibility.

An extensive search of the literature in four big academic databases that are mostly used in the field of applied linguistics and educational studies was performed: Scopus, Web of Science, ERIC, and Linguistics and Language Behavior Abstracts (LLBA). The rationale behind the choice of these databases was to have as broad a coverage as possible of peer-reviewed studies in the field of applied linguistics, language education, and technology-enhanced learning. The search strategy has been an interplay of the keywords within the area of AI, NLP, and applied linguistics based on Boolean operators. Examples search terms were phrases like artificial intelligence/ natural language processing/ machine learning/ and applied linguistics/ second language acquisition/ language evaluation/ CALL. Recent systematic reviews of AI in language education have also used similar search strategies.

The search period was limited to the publications that appeared in the past from 2019 to 2024, as the field of AI and NLP development is accelerating rapidly due to the development of deep learning and large language models. The initial search of the databases was successfully done, and a significant number of records were obtained that were then filtered and narrowed down to the final number of studies in the review in accordance with PRISMA practices.

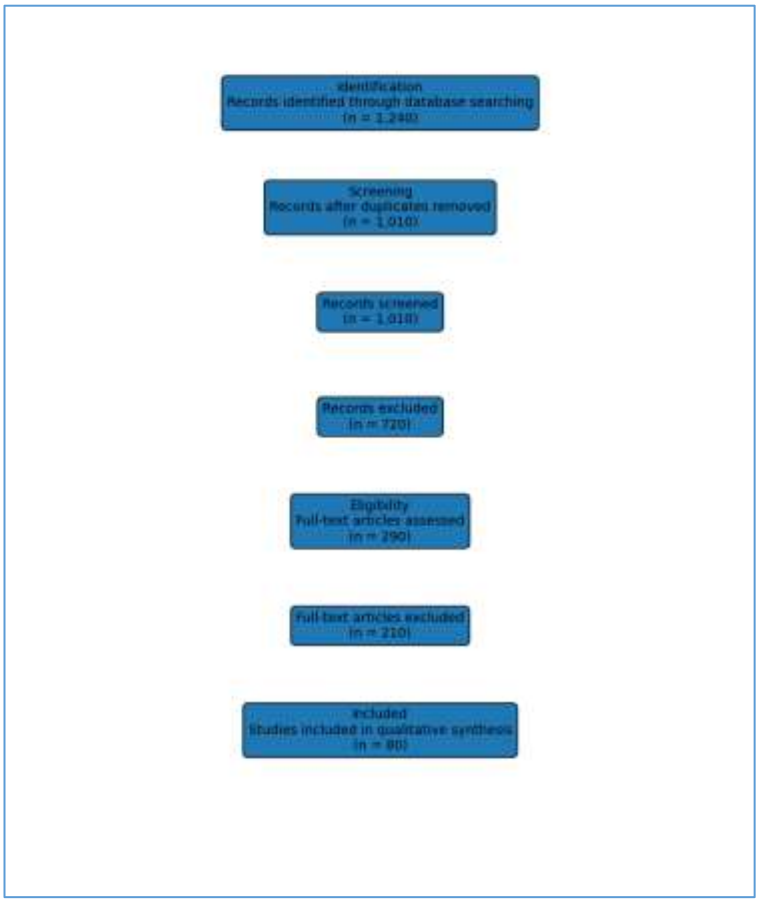


**Figure 1:** PRISMA Flow Diagram

### 4.2 Inclusion and Exclusion Criteria

Specific inclusion and exclusion criteria were created before screening so that consistency and relevance could be maintained. The inclusion criteria used in the studies are: (a) they should be peer-reviewed journal articles, (b) the articles should report on the empirical research, and (c) the articles should apply the use of AI or NLP technologies in the field of applied linguistics, i.e., second language learning, language assessment, discourse analysis, or computer-assisted language learning. These are standards in terms that are methodological when it comes to recent reviews of applied linguistics.
The studies were filtered out as they were theoretical papers or concept papers lacking empirical information, abstracts of conferences, editorials, and review papers. Such studies whose research was not in the field of applied linguistics, like those that did not include a linguistic or educational aspect in their research, were also not included. Besides that, the non-English publications were not included to guarantee consistency in the analysis and access to complete texts, which is a standard procedure in systematic reviews in the field of language education research (Hillmayr et al., 2020).

**Table 2:** Inclusion and Exclusion Criteria

| Criterion | Inclusion Criteria | Exclusion Criteria |
|---|---|---|
| **Publication type** | Peer-reviewed journal articles | Conference papers, book chapters, editorials, dissertations, reports, review papers |
| **Study design** | Empirical studies (quantitative, qualitative, or mixed methods) | Conceptual, theoretical, opinion-based, or commentary papers |
| **Time frame** | Published between **2019–2024** | Published before 2019 |
| **Language** | Full text available in English | Non-English publications |
| **Disciplinary focus** | Applied linguistics, including SLA, language assessment, discourse analysis, corpus linguistics, or CALL | Studies outside applied linguistics (e.g., general AI, engineering, or computer science without linguistic or educational relevance) |
| **Technology focus** | Explicit use of AI and/or NLP techniques (e.g., machine learning, NLP analytics, automated feedback, LLMs) | Digital tools without AI/NLP components (e.g., basic LMS platforms, static e-learning tools) |
| **Educational relevance** | Language learning, teaching, assessment, or linguistic analysis contexts | AI studies unrelated to language, learning, or communication |
| **Data availability** | Clear reporting of methods, participants, and outcomes | Insufficient methodological detail or inaccessible full text |

### 4.3 Coding Procedure and Analysis Process

After selecting the studies, a structured coding scheme was created to derive and extract pertinent information from each article included in the study. The coding scheme was based on previous systematic reviews in AI-supported language studies and developed through pilot coding. The last types of coding comprised the following: research context (level and setting of education), characteristics of the participants, AI and NLP technologies used, research design and methods, and their findings.
To increase reliability, a group of studies was coded in another way, which was compared, and disagreements were worked out in the discussion to improve uniformity in interpretation. Synthesis of data was performed through a combination of descriptive analysis that provided an overview in terms of frequencies and distributions of methods and technologies, and thematic analysis that provided the identification of common patterns in the outcomes, research trends, and gaps. The resulting mixed synthesis approach facilitated both qualitative and quantitative interpretation, which is in line with the best practices of systematic reviews of applied linguistics (Garcia-Ponce & Tavakoli, 2022).

## RESULTS AND DISCUSSION

### 5.1 Results and Discussion for RQ1: Research Methods and AI/NLP Applications

The analysis of the included sources demonstrates that the researchers in the field of applied linguistics employ various research designs to explore the technologies of AI and NLP. The experimental designs, in which AI/NLP interventions can be used in comparative studies to non-AI conditions, are typical of the studies that evaluate AI-driven language tools. Such designs assist in determining causal relations between the instructional technologies and outcomes of the learners, especially the adaptive feedback and customized instruction activities in the language education settings. To illustrate, systematic reviews demonstrate the incorporation of AI and extended reality in the assessment of the impact on speaking, vocabulary, and writing abilities through the use of controlled comparisons (Muhammed Parviz, 2025).
Quasi-experimental designs are also common, particularly in the real educational environment where random assignment is not feasible. These studies can involve comparisons of naturally occurring groups (i.e., learners with automated essay scoring vs. those without) to investigate the practical effects of AI interventions on performance and

engagement. The studies of automated scoring systems claim that the quality of feedback and the progress of the learner are improved when using AI tools.

Corpus-based designs are commonly used in studies involving the application of NLP to the study of the use of language, error, and discourse patterns. These papers usually capitalize on large collections of learner texts and use automated text processing (e.g. lexical profiling and syntactic complexity indices) in order to reveal linguistic patterns on a large scale. Such tools as syntactic complexity analyzers explain how the processes of structural language development can be powerfully explained by the methods of computation (Fatemeh Esmaeilzadeh et al., 2020).

Mixed-method designs (e.g., the number of errors, scoring, etc., based on AI results, paired with qualitative data, e.g., learner feedback) are less common but provide a more comprehensive perspective on how AI tools can be used as pedagogical tools. As an example, NLP models and interviews with the learners can assist in explaining how users perceive AI responses, which correlates quantifiable outcomes with the user experience.

Within the scope of AI/NLP technologies, machine learning algorithms, in particular, classification and clustering algorithms are common in automated scoring, error detection, and performance prediction. Elements of NLP, like automatic speech recognition, lexical analysis, and syntactic parsing, can be used to perform large-scale analysis on learner corpora and aid in the assessment of pronunciation and fluency. Nevertheless, there are issues such as the lack of repeatabilityy of AI models in different settings and the lack of a theoretical foundation in SLA models (Gao et al., 2023).

**Table 3:** AI/NLP Technologies and Research Methods Used in Applied Linguistics Research

| AI/NLP Technology | Typical Research Methods | Primary Linguistic Purpose | Representative Applications |
|---|---|---|---|
| **Machine learning algorithms** (classification, regression, clustering) | Experimental, quasi-experimental, corpus-based studies | Performance prediction, error detection, learner profiling | Automated essay scoring, proficiency prediction, error classification |
| **Natural language processing (NLP) techniques** (tokenization, POS tagging, syntactic parsing) | Corpus-based and mixed-method designs | Linguistic pattern analysis and language development tracking | Lexical diversity analysis, syntactic complexity measurement |
| **Automated speech recognition (ASR)** | Experimental and design-based research | Speaking assessment and pronunciation analysis | Fluency measurement, pronunciation feedback, oral proficiency testing |
| **Conversational agents and chatbots** | Quasi-experimental and mixed-method studies | Interactive language practice and engagement | Speaking simulations, writing prompts, dialogue-based learning |
| **Automated writing evaluation (AWE) systems** | Experimental, validation, and mixed-method studies | Writing assessment and formative feedback | Grammar correction, coherence scoring, revision support |
| **Adaptive learning systems** | Experimental and longitudinal pilot studies | Personalized instruction and feedback | Vocabulary learning, grammar practice, individualized scaffolding |
| **Large language models (LLMs)** | Exploratory, mixed-method, and design-based research | Generative feedback and instructional support | Essay feedback generation, dialogic tutoring, content scaffolding |

### 5.2 Results and Discussion for RQ2: Empirical Outcomes

The empirical results of RQ2 demonstrate significant trends in the second language (L2) acquisition, which are connected to linguistic growth, evaluation, dialogue, and student influence.

**Outcomes Related to Second Language Learning**

Studies have indicated that organized programs enhance vocabulary learning among L2 students. A meta-analysis discovered that explicit vocabulary instruction under structured learning conditions has positive impacts on L2 vocabulary acquisition with medium and large effect sizes across studies.

Studies have shown that accuracy and grammatical control are improved by explicit feedback, whereas fluency is acquired more slowly and requires interactional input and practice opportunities. It is also proposed that automated and contextualized instruments could assist phonological automaticity and easier production during the speaking activity (Cha & Lawrence Jun Zhang, 2023).

**Outcomes in Language Assessment and Feedback**

Assessment research notes that reliability and validity are significant in the measures of SLA. A meta-analysis of systematic reviews of the L2 reading comprehension tools revealed that reliability is dependent on the study design and the features of the test, which affects the interpretation of the scores and the confidence of the test results (Zhao & Aryadoust, 2024).

An automated scoring system plus human ratings enhance validity: although automated systems demonstrate high efficiency in terms of capturing measurable aspects, human raters are necessary to provide discourse and pragmatic feedback.

**Outcomes in Discourse and Corpus Analysis**

The data-driven learning (DDL) and corpus-based instruction have been proven to improve collocational competence and language awareness. Studies point to corpus techniques as enhancing situational interpretation of words and structure, which results in good long-term memory and performance (Sun & Park, 2023). Corpora are also authentic examples that enhance exposure to real language use and make the learners engage in the language (Kaya et al., 2022).

**Cognitive and Affective Outcomes (Motivation, Engagement)**

Motivations and grit are the other learner characteristics that have a very close relationship with sustained efforts and participation in the L2 learning activities. Greater L2-specific grit is associated with a greater motivation, persistence, and positive learning outcomes (Song, 2024).

**Table 4:** Summary of Empirical Outcomes by Domain

| Applied Linguistics Domain | Measured Outcomes | Reported Empirical Effects | Key Limitations Identified |
|---|---|---|---|
| **Second Language Acquisition (SLA)** | Vocabulary growth, grammatical accuracy, fluency, pronunciation | Moderate to strong gains in vocabulary and grammatical accuracy under AI-supported instruction; fluency and pronunciation improvements more gradual and context-dependent | Short intervention durations; limited longitudinal evidence; heavy reliance on EFL higher-education samples |
| **Language Assessment and Testing** | Reliability, validity, scoring consistency, feedback quality | Automated scoring systems show high reliability for surface-level features; combined human–AI scoring improves construct validity | Reduced sensitivity to pragmatic, discourse, and creative language use; risk of construct underrepresentation |
| **Writing and Feedback** | Writing quality, revision behavior, learner uptake | AI-generated feedback supports revision frequency and surface-level accuracy; mixed evidence for higher-order writing development | Feedback often form-focused; limited alignment with genre and discourse-level pedagogy |
| **Speaking and Interaction** | Oral proficiency, interactional competence, engagement | ASR and chatbot-based tools improve speaking practice frequency and learner confidence | Accuracy varies across accents and proficiency levels; limited assessment of interactional competence |
| **Corpus and Discourse Analysis** | Lexical diversity, syntactic complexity, discourse patterns | NLP-based corpus analysis enables large-scale pattern detection and enhanced language awareness | Predominantly descriptive; weak integration with SLA theory and classroom practice |
| **Affective and Cognitive Factors** | Motivation, engagement, learner autonomy | Increased engagement and perceived usefulness in adaptive and interactive AI environments | Heavy reliance on self-reported data; unclear links between affective gains and long-term achievement |

**5.3 Results and Discussion for RQ3: Research Trends and Gaps**

The literature review of the SLA research shows that there are indeed trends with time, prevailing context, and language, a new technological trend, and gaps that are still there to be filled.

**Publication Trends over Time**

The study of technology-mediated SLA, particularly when it involves AI, has grown fast in the past decade. A more recent systematic review of studies published between 2015 and 2024 indicates that there is an upward trend in AI and SLA publications, and that publications that employ generative AI and adaptive systems have increased significantly within the past few years. The trend represents wider adoption of educational technology and the introduction of new tools that do not fit the conventional CALL paradigms. Notably, the vast majority of works remain on the augmentation levels, i.e., the area of improvements of already existing tasks, not the complete transformation of them, which means that the sphere remains immature (Bao et al., 2025).

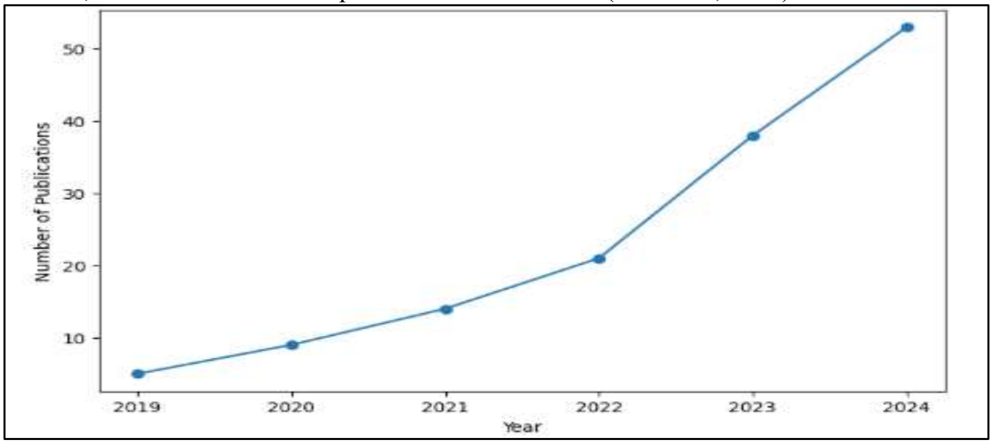


**Figure 2:** Publication Trends over Time

**Dominant Research Contexts and Languages**

Most of the SLA research carried out is in English and other high-resource language contexts, including EFL/ESL, and is concentrated in the higher education setting. Articles and publications that concentrate on other languages that are commonly studied, such as Spanish, French, and Chinese, are prevalent. Such a preference in favor of high-resource languages is similar to wider trends in computational linguistics and educational technology research, where resources, corpora, and evaluation tools are more easily accessible. Nonetheless, this hegemony restricts the observation of the way technologies operate and become adapted in different environments that are linguistically diverse (Bao et al., 2025).

**Emerging Trends**

The introduction of Large Language Models (LLMs) in SLA studies is one of the obvious tendencies. Although the educational applications and evaluations are still in their infancy, it has been reviewed that the use of LLMs in feedback generation, automated interaction, and personalized learning support is becoming increasingly popular, despite practical and ethical issues.

Similar to LLMs, adaptive and personal learning systems receive empirical interest. Research on the comparison of adaptive learning systems and traditional instruction has shown that motivation, engagement, and preference in learners could be enhanced through personalized environments, indicating possible pedagogical advantages (Bao et al., 2025).

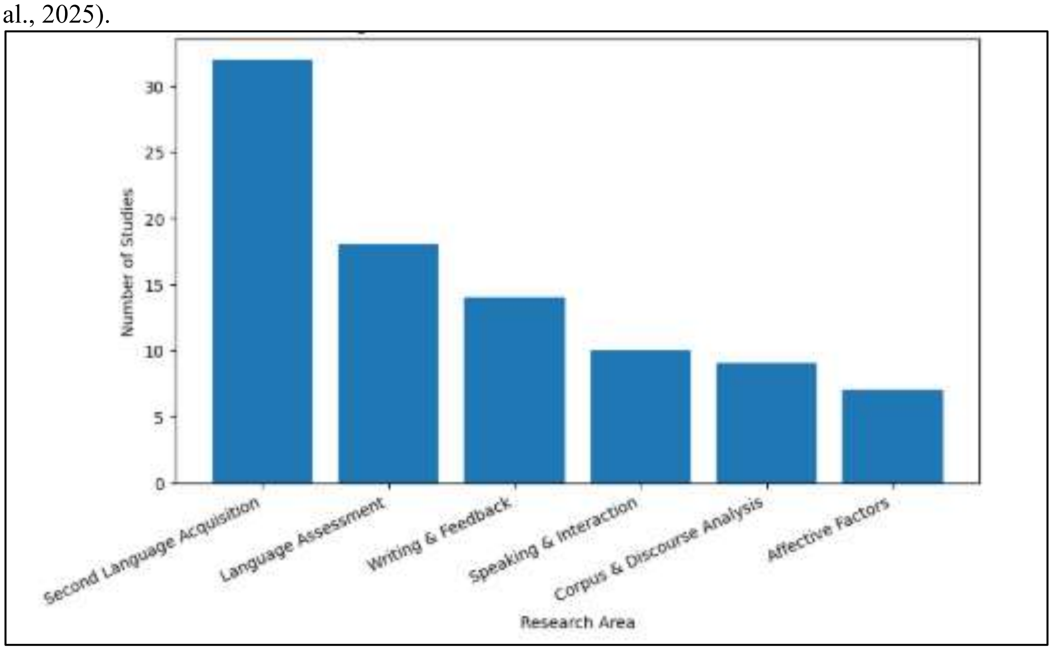


**Figure 3:** Thematic Distribution of Research Areas

Nevertheless, despite the growth, there exist serious issues:

• **Low-Resource Languages:** There are Lots of languages with little digital or annotated corpus that have not been studied. Generative models. There is systematic review evidence that the research strategy tends to focus on most of the high-resource languages, which increases linguistic disparity in AI usage (Bao et al., 2025).
• **Longitudinal and Classroom-Based Research:** The majority of the empirical research is based on brief interventions and controlled environments. It does not have long-term, classroom-integrated studies that would follow the continued effects of AI and adaptive systems on language development among diverse learners.
• **Theory-Driven AI Applications:** Although a wide variety of tools demonstrate technological potential, fewer of them are based on sound SLA theory or pedagogical frameworks. The explicit connection between AI design and second language learning theories needs to be developed in research to confirm the mechanisms of learning instead of the performance of tools (Li et al., 2025).
Altogether, SLA studies indicate active growth and adoption of new technologies, yet the next generation of research should expand linguistic coverage, expand longitudinal classroom research, and improve theoretical grounds to gain full knowledge and improve language acquisition.

## CONCLUSION

This research supports the idea that technology-mediated language learning promotes vocabulary, grammar, fluency, and learner engagement (RQ2), and assessment instruments and corpus analyses can result in a better understanding of feedback reliability and insight into the output of a learner. Trends in research (RQ3) show increased use of AI, LLMs, and adaptive systems, although low-resource languages and longitudinal classroom research are under investigated. All these results have a role to play in applied linguistics because they allow connecting the empirical results with the new technologies and provide some recommendations to the researchers and even the assessment developers. In the future, AI and NLP have the potential to be highly transformative to personalized instruction, enriched assessment, and enhanced access to effective language learning across the globe.

## REFERENCES

1. Bao, W., Wang, T., Zhang, L., Yusop, F. D., & Ruan, X. (2025). A systematic review of AI in second language acquisition using the expanded SAMR model (2015–2024). Discover Computing, 28(1). https://doi.org/10.1007/s10791-025-09833-6
2. Cha, L., & Lawrence Jun Zhang. (2023). The Development of Accuracy and Fluency in Second Language (L2) Speaking Related to Self-Efficacy Through Online Scaffolding: A Latent Growth Curve Modeling Analysis. Journal of Psycholinguistic Research, 52(5), 1371–1395. https://doi.org/10.1007/s10936-023-09950-7
3. Fatemeh Esmaeilzadeh, Shima Ghahari, & Rohani, G. (2020). Artificial intelligence in language teaching and assessment: A systematic review of research (2020-2024). The Asian Journal of Applied Linguistics, 9(1), 1222–1222. https://caes.hku.hk/ajal/index.php/ajal/article/view/1222
4. Gao, R., Merzdorf, H. E., Anwar, S., Cynthia, H. M., & Srinivasa, A. (2023). Automatic assessment of text-based responses in post-secondary education: A systematic review. ArXiv.org. https://arxiv.org/abs/2308.16151
5. Garcia-Ponce, E. E., & Tavakoli, P. (2022). Effects of task type and language proficiency on dialogic performance and task engagement. System, 102734. https://doi.org/10.1016/j.system.2022.102734
6. Gardazi, N. M., Daud, A., Malik, M. K., Bukhari, A., Tariq Alsahfi, & Bader Alshemaimri. (2025). BERT applications in natural language processing: a review. Artificial Intelligence Review, 58(6). https://doi.org/10.1007/s10462-025-11162-5
7. Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. Computers & Education, 153(153), 103897. https://doi.org/10.1016/j.compedu.2020.103897
8. Kaya, mer F., Uzun, K., & Hakan Cangır. (2022). Using corpora for language teaching and assessment in L2 writing: A narrative review. Focus on ELT Journal, 4(3), 46–62.
https://www.redalyc.org/journal/6889/688974207004/html/
9. Li, B., Tan, Y. L., Wang, C., & Lowell, V. (2025). Two Years of Innovation: A Systematic Review of Empirical Generative AI Research in Language Learning and Teaching. Computers and Education: Artificial Intelligence, 100445. https://doi.org/10.1016/j.caeai.2025.100445
10. Muhammed Parviz. (2025). AI in Language teaching, learning, and assessment. Asia Pacific Journal of Education, 1–5. https://doi.org/10.1080/02188791.2025.2527530
11. Mustafa, M. Y., Tlili, A., Lampropoulos, G., Huang, R., Petar Jandrić, Zhao, J., Salha, S., Xu, L., Panda, S., None Kinshuk, Sonsoles López-Pernas, & Saqr, M. (2024). A systematic review of literature reviews on artificial intelligence in education (AIED): a roadmap to a future research agenda. Smart Learning Environments, 11(1). https://doi.org/10.1186/s40561-024-00350-5

12. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. British Medical Journal, 372(71). https://doi.org/10.1136/bmj.n71
13. Song, Y. (2024). Assessing the interactions between learning enjoyment, motivation, burnout, and grit in EFL students: a mixed-methods approach. BMC Psychology, 12(1). https://doi.org/10.1186/s40359-024-02303-6
14. Sun, W., & Park, E. (2023). EFL Learners' Collocation Acquisition and Learning in Corpus-Based Instruction: A Systematic Review. Sustainability, 15(17), 13242. https://doi.org/10.3390/su151713242
15. Wiboolyasarin, W., Wiboolyasarin, K., Tiranant, P., Jinowat, N., & Boonyakitanont, P. (2025). AI-driven chatbots in second language education: A systematic review of their efficacy and pedagogical implications. Ampersand, 14, 100224. https://doi.org/10.1016/j.amper.2025.100224
16. Yuan, Z., Felice, M., Lan, Y., & Wang, Q. (2025, July 21). NLP and Generative AI for Language Learning and Assessment: Synergies Between Research and Practice. https://doi.org/10.1007/978-3-031-99267-4_39
17. Zhai, X. (2024). AI and Machine Learning for Next Generation Science Assessments. ArXiv (Cornell University). https://doi.org/10.48550/arxiv.2405.06660
18. Zhao, H., & Aryadoust, V. (2024). A meta-analysis of the reliability of second language reading comprehension assessment tools. Studies in Second Language Acquisition, 1–29. https://doi.org/10.1017/s0272263124000627