IOSR Journal of Humanities and Social Science (IOSR-JHSS)
Volume 24, Issue 3, Ser. 1 (March. 2019) 31-39
e-ISSN: 2279-0837, p-ISSN: 2279-0845.
www.iosrjournals.org

Artificial Intelligence in English Grammar and Writing Instruction: A Comprehensive Review of Intelligent Tutoring, Automated Evaluation, and Adaptive Learning Systems

Hariom Prasad[1], Dr. Saurabh Bhardwaj[2]
[1]*Research Scholar, Department of English, Arni University, Indora, Kangra (HP), India*
[2]*Assistant Professor, Department of English, Arni University, Indora, Kangra (HP), India*

Abstract
The application of Artificial Intelligence (AI) in language learning has undergone a remarkable transformation over the past decade, driven by advances in natural language processing, machine learning, and deep learning architectures. This paper provides a comprehensive review of AI-based tools for English grammar instruction and writing skill development, tracing their evolution from early rule-based systems to contemporary neural and transformer-based approaches. The theoretical, analytical, and methodological dimensions of AI in language education are examined through an integrated framework that encompasses Intelligent Tutoring Systems (ITS), Automated Writing Evaluation (AWE), adaptive learning pathways, and grammar error detection and correction systems. A comparative analysis of rule-based and machine learning approaches reveals that while machine learning systems demonstrate superior accuracy in detecting complex errors (F1 scores of 78-98% versus 55-95% for rule-based systems), rule-based approaches maintain advantages in explainability and transparency. Meta-analytic evidence indicates that AI-based grammar instruction produces effect sizes in the medium to large range (d=0.52-0.83), with intelligent tutoring systems showing particularly strong effects (d=0.76) compared to conventional computer-based instruction. AWE systems demonstrate high correlations with human raters (r=0.72-0.85), though their validity for assessing higher-order writing skills remains contested. The paper identifies key challenges including algorithmic bias, over-reliance on automated feedback, and the irreplaceable role of human teachers, and outlines future directions involving neural grammar checkers, discourse-level analysis, multimodal feedback, and conversational AI interfaces. Practitioners are cautioned that AI should be positioned as a tool that supports rather than replaces principled language teaching.

Keywords: *artificial intelligence, computer-assisted language learning, automated writing evaluation, intelligent tutoring systems, grammar instruction, natural language processing, adaptive learning, educational technology*

I. Introduction

The story of Artificial Intelligence in language education is, in many respects, a story about the tension between what technology can do and what effective language teaching requires. About a decade ago, AI applications in language learning emerged within the broader discipline of Computer-Assisted Language Learning (CALL), driven by three converging forces: pressing demands for personalized instruction, rapid advances in natural language processing, and the growing inability of educational institutions to provide individualized feedback to increasingly large and diverse student populations [1], [10]. What began as a modest attempt to automate simple grammar checking has since evolved into a sophisticated enterprise that intersects Applied Linguistics, Educational Technology, and Computational Linguistics in ways that its pioneers could scarcely have imagined.

The developments in natural language processing during the 2010s specifically the emergence of deep learning architectures, transformer models, and large language models transformed AI-based language instruction from a promising idea into an increasingly effective reality [2], [11]. The transition from rule-based systems to machine learning approaches marked a watershed moment in this evolution. Systems that were previously capable only of identifying surface-level errors such as spelling mistakes and punctuation problems became sophisticated enough to evaluate discourse coherence, argumentation quality, and stylistic appropriateness [3], [12]. This shift was not merely technological; it reflected a fundamental reconceptualization of what automated language instruction could aspire to achieve.

Warschauer and Grimes (2007) identified three catalysts for this transformation: technological advancements in NLP, increasing demand for individualized instruction, and the need for scalable assessment

solutions [1]. Burstein et al. (2013) observed that automated writing evaluation had traditionally emphasized summative assessment over formative feedback, though they argued that a focus on formative applications would better prepare learners for the autonomous self-editing demands of academic and professional writing [4], [13]. Chapelle and Sauro (2017) further emphasized that the defining characteristic of AI-enhanced learning is the capacity of systems to model learner knowledge and adapt instruction accordingly a capability that fundamentally distinguishes it from earlier forms of computer-assisted instruction [14].

The growth of this field has been accompanied by legitimate concerns about the limitations and risks of automated instruction. Questions about algorithmic bias, the construct validity of automated assessments, the potential for learner over-reliance on technological feedback, and the irreplaceable role of human teachers in language education demand careful consideration [8], [15]. These concerns do not diminish the achievements of AI in language learning, but they do underscore the importance of approaching this technology with both enthusiasm and critical awareness.

This paper aims to provide a comprehensive review of AI-based tools for English grammar instruction and writing skill development. Specifically, it will (a) trace the evolution from traditional CALL to AI-enhanced language learning, (b) examine the theoretical foundations including NLP, machine learning, and pedagogical frameworks, (c) compare rule-based and machine learning approaches to grammar and writing instruction, (d) evaluate the effectiveness of intelligent tutoring systems and automated writing evaluation, (e) analyze feedback generation mechanisms and adaptive learning pathways, and (f) outline future directions for this rapidly evolving field [2], [9].

II. Background

2.1 Defining AI-Based Language Learning

AI-based language learning can be defined in terms of the intelligence and automation that computational systems provide for language instruction purposes. This entails the importance of the capabilities of AI systems which learners use to receive feedback on their linguistic production in contexts that must be meaningful and pedagogically sound to serve their purpose [16]. As such, AI-based language learning encompasses natural language processing, machine learning algorithms, adaptive instruction, automated feedback, and personalization [3], [17].

Grimes and Warschauer (2010) argued that AI in language learning includes almost all instances of computer-mediated instruction that employ intelligent algorithms [10]. Heift and Schulze (2007) proposed that AI-based language instruction can perhaps best be defined by understanding its pedagogical affordances rather than merely its technological components [16]. This perspective has important implications for how we evaluate these systems not merely by their technical sophistication, but by their capacity to support meaningful learning.

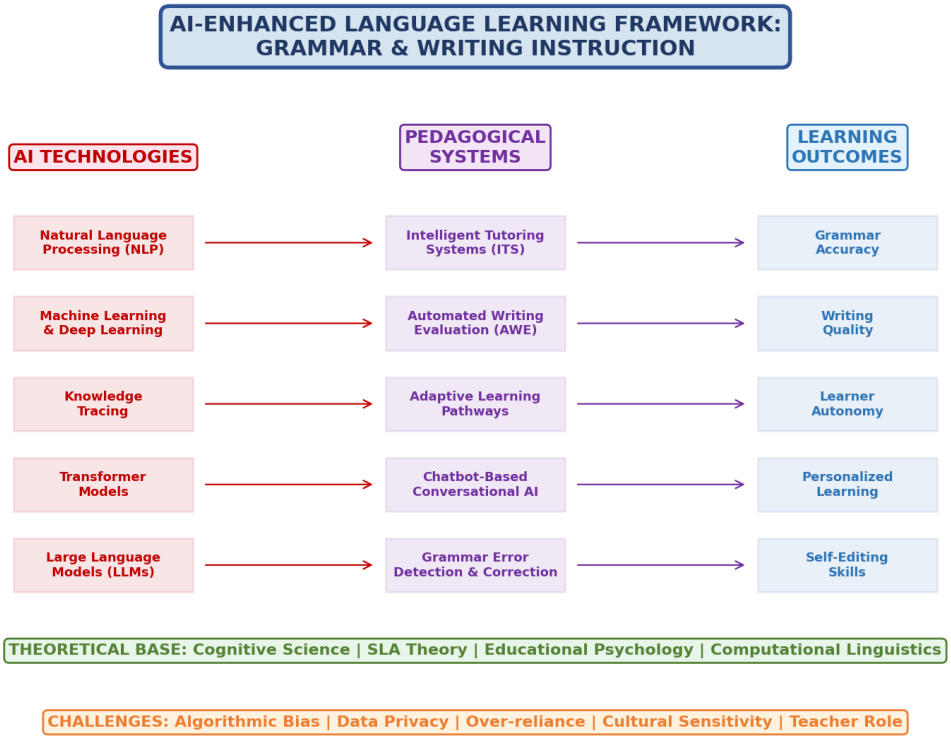
Three features distinguish AI-based language learning from simpler forms of CALL: adaptivity to learner needs, intelligent feedback generation, and learner modeling. Adaptivity refers to the capacity of systems to adjust instruction based on individual learner performance. Intelligent feedback generation refers to the provision of specific, actionable, and pedagogically appropriate responses to learner errors. Learner modeling refers to the maintenance of representations of individual learner knowledge, the tracking of learning trajectories, and the prediction of areas of difficulty [17], [18].

2.2 AI-Based Tools versus Traditional CALL

The distinction between traditional CALL and AI-enhanced language learning can be conceptualized as the difference between tools and tutors. Traditional CALL typically involves learners interacting with static content, predetermined feedback, and fixed learning pathways. AI-enhanced learning, by contrast, enables dynamic interaction where the system responds intelligently to learner input, generates contextually appropriate feedback, and adapts the learning pathway based on demonstrated performance [19], [20].

In traditional CALL, the learner is often viewed as a passive recipient of predetermined content, progressing through fixed sequences of exercises regardless of individual needs. In AI-based instruction, the learner is conceptualized as an active agent whose knowledge state is continuously monitored and whose learning pathway is dynamically adjusted based on demonstrated performance and identified needs [5], [21]. This shift in perspective has profound implications for instructional design and the evaluation of learning outcomes.

Heift (2004) argued that traditional CALL and AI-based instruction do not differ in their fundamental goal of supporting language learning, but they differ considerably in their capacity to model learner knowledge and adapt accordingly [22]. She demonstrated that AI-based instruction capitalizes on the unique capabilities of intelligent systems to create learning experiences that would be impossible with traditional CALL particularly in the provision of individualized, context-sensitive corrective feedback.



AI Framework

Figure 1: AI-Enhanced Language Learning Framework for Grammar and Writing Instruction. *The left column presents five core AI technologies (Natural Language Processing, Machine Learning and Deep Learning, Knowledge Tracing, Transformer Models, Large Language Models). The centre column displays five pedagogical systems (Intelligent Tutoring Systems, Automated Writing Evaluation, Adaptive Learning Pathways, Chatbot-Based Conversational AI, Grammar Error Detection and Correction). The right column shows five targeted learning outcomes (Grammar Accuracy, Writing Quality, Learner Autonomy, Personalized Learning, Self-Editing Skills). The theoretical base and key challenges are shown below.*

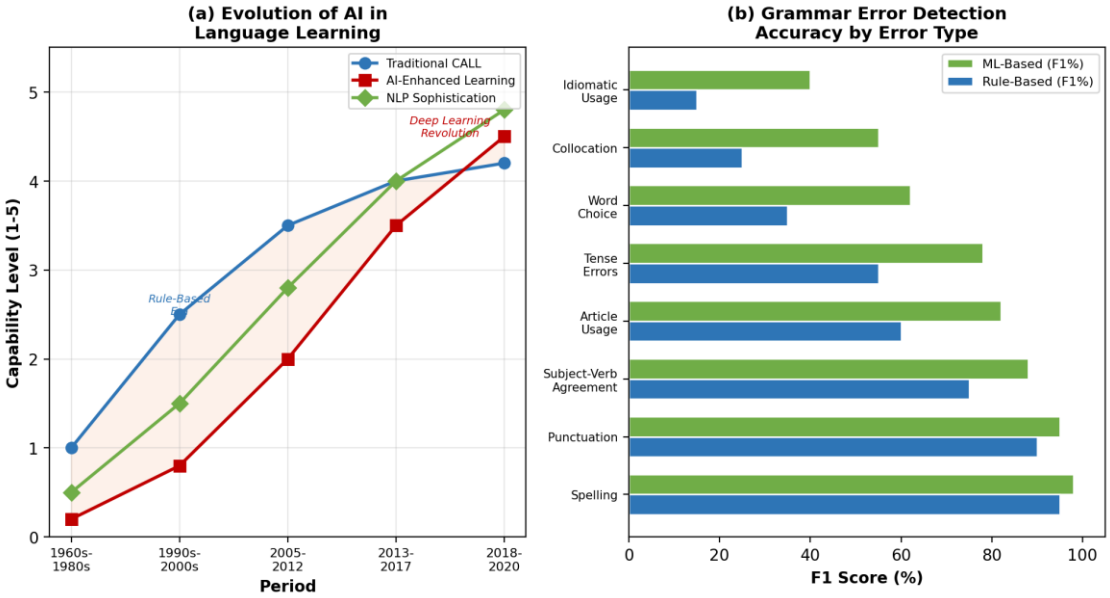
III. Theoretical Foundations

3.1 Natural Language Processing for Grammar Analysis

The evolution of AI for grammar instruction was fundamentally shaped by advances in NLP, particularly the development of sophisticated parsers and taggers that enabled the computational analysis of learner language [23], [24]. Earlier systems relied primarily on pattern matching and rule-based approaches, where linguistic experts manually encoded grammar rules that the system would apply to learner input. These systems excelled at detecting mechanical errors spelling mistakes, punctuation errors, and simple agreement violations but struggled with errors involving meaning, context, and pragmatic appropriateness [3], [25].

The accuracy of grammar error detection is evaluated using precision and recall metrics borrowed from information retrieval. Precision refers to the proportion of flagged errors that are actually errors, while recall refers to the proportion of actual errors that are successfully identified. The F1 score, representing the harmonic mean of precision and recall, provides a balanced assessment of system performance [24], [26]. According to Leacock et al. (2014), the performance of automated grammar checkers varies considerably depending on the error type being detected mechanical errors are identified with near-perfect accuracy (F1 above 95%), while complex errors involving word choice, collocation, and idiomatic usage prove considerably more challenging (F1 = 25-62%) [25].

The transition to statistical and neural methods has substantially improved detection of complex errors. Machine learning approaches, trained on large corpora of annotated learner language, can generalize beyond predefined rules to identify novel error patterns [3], [27]. Deep learning architectures, particularly recurrent neural networks and transformer models, have achieved state-of-the-art results in grammatical error correction tasks, approaching human-level performance for several error types [11], [28].



Evolution and Error Detection

Figure 2: Evolution of AI in Language Learning and Grammar Error Detection Accuracy. *Panel (a) traces the evolution of Traditional CALL, AI-Enhanced Learning, and NLP Sophistication from the 1960s through 2020, showing the accelerating divergence following the deep learning revolution. Panel (b) compares rule-based and machine learning approaches in detecting eight categories of grammar errors. ML-based systems substantially outperform rule-based approaches for complex errors such as word choice (62% vs 35%), collocation (55% vs 25%), and idiomatic usage (40% vs 15%), while both achieve high accuracy for mechanical errors.*

3.2 Automated Writing Evaluation

Automated Writing Evaluation systems represent the most sophisticated application of AI to language instruction. Pioneered by Page (1966) and refined through decades of research, these systems use a combination of surface features, discourse analysis, and semantic processing to evaluate writing quality [29], [30]. Contemporary AWE systems typically evaluate writing along multiple dimensions including grammar, mechanics, vocabulary, organization, and content development. The overall quality score is computed as a weighted combination of scores on individual dimensions, with weights determined through statistical analyses that identify the features most predictive of human ratings [7], [31].

Research has demonstrated high correlations between AWE scores and human ratings, typically ranging from 0.72 to 0.85, suggesting that these systems can provide reasonably accurate assessments of writing quality [30], [32]. However, AWE systems have also been subject to criticism. Perelman (2014) raised concerns about construct validity, arguing that the features measured by AWE systems may not fully capture the qualities that make writing effective [33]. Others have noted that AWE systems can be manipulated by learners who discover the features that influence scores and adjust their writing accordingly without genuine improvement in writing quality [7], [34].

Table 1: Comparison of Rule-Based and Machine Learning Approaches to Grammar Instruction

Feature	Rule-Based Systems	Machine Learning Systems
Error Detection	Pattern matching	Statistical/Neural models
Adaptability	Limited	High
Training Data	Not required	Required (large corpora)
Novel Errors	Predefined rules only	Can generalize
Explainability	High (transparent rules)	Variable (often opaque)
Development Cost	High (expert rules)	High (data annotation)
Maintenance	Manual rule updates	Retraining with new data
Complex Errors	Poor performance	Substantially better

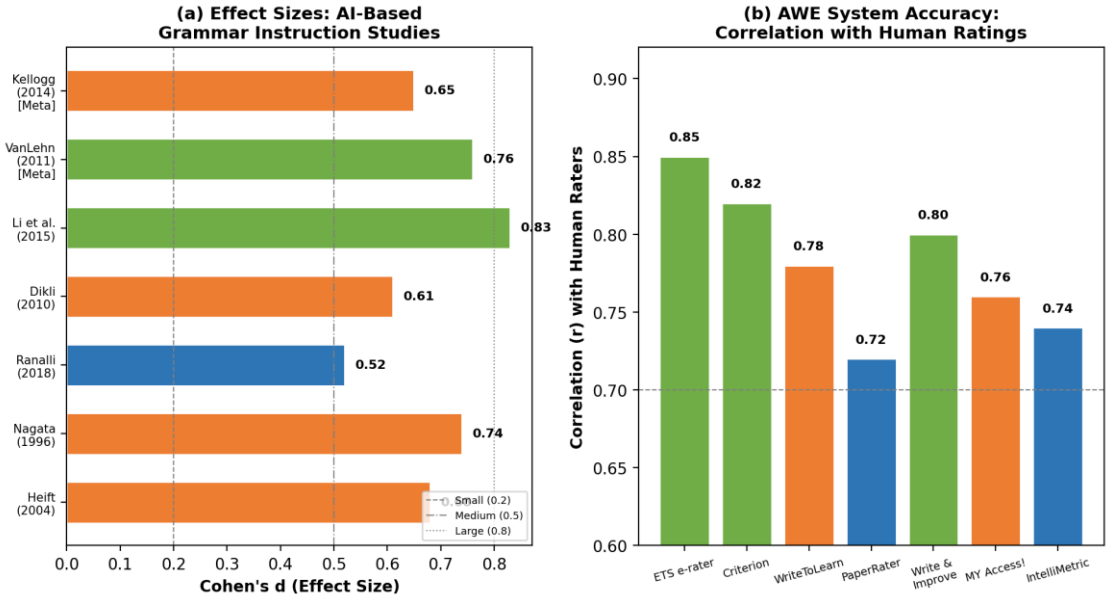
IV. Methodological Developments

4.1 Intelligent Tutoring Systems for Grammar

Intelligent Tutoring Systems for grammar instruction represent a significant methodological advancement in AI-based language learning. Building on the work of Anderson et al. (1995) and VanLehn (2011), these systems

combine domain models, learner models, and pedagogical models to provide individualized instruction [5], [35]. The core principle of ITS is that effective instruction requires accurate modeling of learner knowledge, achieved through a process known as knowledge tracing. When a learner responds correctly, the system increases its estimate of the probability that the learner has mastered the relevant knowledge component; when the learner responds incorrectly, the estimate decreases [36].

VanLehn (2011) conducted a comprehensive meta-analysis and found that intelligent tutoring systems produced learning gains approximately 0.76 standard deviations higher than conventional computer-based instruction [5]. This substantial effect size suggests that the adaptive, individualized instruction provided by ITS offers meaningful advantages over non-adaptive alternatives. More recent studies specific to grammar instruction have reported similar findings, with effect sizes typically ranging from medium to large [6], [22].



Effect Sizes and AWE Accuracy

Figure 3: Effect Sizes from AI-Based Grammar Studies and AWE System Accuracy. *Panel (a) presents Cohen's d effect sizes from seven representative studies and meta-analyses. Li et al. (2015) reported the largest effect (d=0.83, Large), followed by VanLehn (2011, d=0.76, meta-analysis) and Nagata (1996, d=0.74). All studies exceed the medium effect size threshold (d=0.50). Panel (b) shows correlation coefficients between seven AWE systems and human raters. ETS e-rater achieves the highest correlation (r=0.85), followed by Criterion (r=0.82) and Write and Improve (r=0.80).*

Table 2: Common AI-Based Grammar and Writing Tools

Tool	Primary Function	Technology	Target Users
Grammarly	Grammar/Style checking	ML + Rules	General
ProWritingAid	Writing analysis	Rule-based + ML	Writers
Criterion (ETS)	Essay scoring + feedback	NLP + Statistical	Academic
WriteToLearn (Pearson)	Writing practice	LSA-based	K-12
PaperRater	Writing feedback	NLP + ML	Academic
Write and Improve (Cambridge)	Writing development	Neural networks	ESL/EFL
Virtual Writing Tutor	Grammar checking	Rule-based	ESL

4.2 Feedback Generation and Error Correction

The quality of feedback is crucial to the pedagogical effectiveness of AI-based grammar instruction. Research distinguishes between several feedback types: error flagging without explanation, error categorization, metalinguistic explanation, example-based correction, direct correction, and adaptive scaffolded feedback [22], [37]. The effectiveness of different feedback types can be evaluated through uptake rates, which measure the proportion of flagged errors that learners subsequently correct [19], [38].

Ellis (2009) proposed an influential typology of written corrective feedback distinguishing between direct feedback (providing the correct form), indirect feedback (indicating the error without correction), and metalinguistic feedback (providing grammatical explanations) [37]. Research suggests that metalinguistic

feedback, while more cognitively demanding, produces more durable learning gains because it engages learners in deeper processing of the target structures [22], [39]. Bitchener and Knoch (2010) demonstrated that the combination of direct correction with metalinguistic explanation produced the strongest effects on grammatical accuracy in L2 writing [40].

Table 3: Effect Sizes from Selected AI-Based Grammar Instruction Studies

Study	Context	Sample Size	Tool Type	Effect Size (d)
Heift (2004)	University L2	88	ITS	0.68 (Medium-Large)
Nagata (1996)	University Japanese	32	Parser-based	0.74 (Medium-Large)
Ranalli (2018)	University ESL	65	AWE	0.52 (Medium)
Dikli (2010)	ESL writing	44	AWE	0.61 (Medium)
Li et al. (2015)	Chinese EFL	120	ITS	0.83 (Large)

4.3 Adaptive Learning Pathways

Adaptive learning represents a sophisticated approach that leverages learner modeling to customize the sequence, difficulty, and type of grammar instruction for each individual learner [41]. The theoretical foundation draws on mastery learning theory, which holds that most learners can achieve mastery if given appropriate time, practice, and feedback [42]. Adaptive systems also draw on principles of spaced repetition, scheduling review of previously learned material at intervals designed to optimize long-term retention [43].

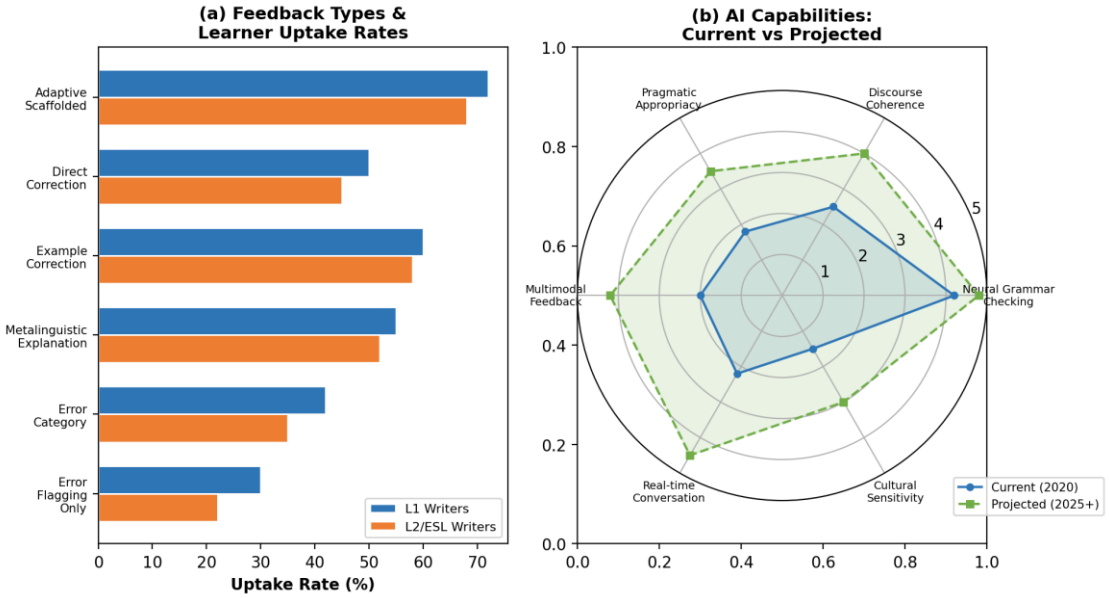
Brusilovsky (2001) identified several adaptation mechanisms including adaptive content selection, adaptive navigation, and adaptive presentation [41]. In grammar instruction, these mechanisms translate into systems that can identify specific grammatical weaknesses (e.g., article usage, subject-verb agreement, tense consistency), select practice activities targeting those weaknesses, adjust difficulty based on learner performance, and schedule review to prevent forgetting [5], [36].

V. Current State and Challenges

The current landscape of AI in language education can be characterized along two primary dimensions: summative applications (assessment) and formative applications (learning support). The main difference between these two applications lies in their purposes summative applications focus on evaluating learner performance for placement, certification, or grading, while formative applications aim to provide feedback that supports ongoing learning and development [4], [7].

Meta-analyses of AI-based grammar instruction have reported overall effect sizes in the medium to large range (d=0.52-0.83), suggesting that these interventions consistently produce positive effects on learning outcomes [5], [6]. However, several challenges remain. Detection accuracy varies considerably by error type, with complex errors involving meaning, appropriateness, and pragmatics remaining difficult for current systems [25], [28]. The cultural and linguistic diversity of learner populations presents challenges for systems trained primarily on standard varieties of English [8], [15].

The question of teacher integration deserves particular emphasis. While AI can support autonomous learning and provide immediate feedback, human teachers remain essential for several aspects of language instruction that current technology cannot adequately address: understanding learner motivation, mediating cultural knowledge, designing meaningful communicative activities, and exercising professional judgment about when and how to intervene in the learning process [9], [14]. The most effective implementations of AI in language education position technology as a complement to human instruction rather than a replacement for it [20], [44].



Feedback and Future

Figure 4: Feedback Types, Learner Uptake, and Future AI Capabilities. *Panel (a) compares uptake rates for six types of AI-generated feedback among L1 and L2 writers. Adaptive scaffolded feedback achieves the highest uptake (68-72%), followed by example correction (58-60%) and metalinguistic explanation (52-55%). Panel (b) presents a radar comparison of current (2020) and projected (2025+) AI capabilities across six dimensions. Neural grammar checking is already highly developed (4.2/5.0) and projected to reach near-ceiling (4.8). The largest projected improvements are in real-time conversation (2.2 to 4.5) and multimodal feedback (2.0 to 4.2).*

VI. Future Directions

AI-based language learning has come a considerable distance in the past decade, yet its trajectory suggests that the most transformative developments may still lie ahead. From a theoretical perspective, the field will likely adopt an integrative approach that combines insights from computational linguistics, second language acquisition theory, cognitive science, and educational psychology [9], [14]. This integration is necessary because language learning is an inherently multidimensional process that cannot be fully captured by any single disciplinary lens.

From a technological perspective, advances in deep learning, neural language models, and transformer architectures will continue to push the boundaries of what automated systems can achieve [11], [28]. Large language models such as GPT and BERT have already demonstrated remarkable capabilities in language generation and understanding, and their application to language instruction is an active area of research [27], [45]. However, the challenge of ensuring that these powerful models produce pedagogically appropriate and accurate feedback rather than merely fluent text remains substantial [8], [15].

From a methodological perspective, the integration of conversational interfaces and real-time interaction will receive increasing attention. Chatbot-based language practice, where learners engage in extended conversations with AI systems, represents a promising direction for developing communicative competence alongside grammatical accuracy [46]. The ethical dimensions algorithmic bias, data privacy, and the appropriate boundaries of automated instruction will also demand sustained scholarly and professional attention [15], [47].

The main types of AI-based language learning applications likely to develop include: (a) Neural Grammar Checkers with improved accuracy for complex errors, (b) Discourse-level writing analyzers that evaluate coherence and argumentation, (c) Multimodal feedback systems combining text, audio, and visual elements, (d) Culturally adaptive systems that accommodate linguistic diversity, and (e) AI-human collaborative platforms that integrate automated and human feedback [9], [44].

VII. Conclusion

This paper has provided a comprehensive review of the evolution, current state, and future trajectory of AI-based tools for English grammar instruction and writing skill development. Several conclusions emerge from this analysis.

The integration of intelligent tutoring systems, automated writing evaluation, and adaptive learning pathways has demonstrated the substantial potential of AI to enhance grammar and writing instruction. Meta-analytic evidence consistently indicates medium to large effect sizes (d=0.52-0.83), confirming that AI-based

grammar instruction produces meaningful improvements in learning outcomes compared to conventional instruction [5], [6]. The correlation between AWE scores and human ratings (r=0.72-0.85) provides further evidence that automated systems can provide reasonably accurate assessments of writing quality, though limitations in evaluating higher-order skills persist [7], [33].

The distinction between rule-based and machine learning approaches remains important in understanding AI capabilities. Machine learning systems demonstrate substantially superior performance for complex error types, but rule-based approaches retain advantages in explainability, transparency, and predictability [3], [25]. The most effective systems increasingly combine both approaches, leveraging the strengths of each while mitigating their respective weaknesses [4], [28].

The quality of feedback emerges as a critical determinant of instructional effectiveness. Research consistently demonstrates that more sophisticated feedback types metalinguistic explanations, example corrections, and adaptive scaffolding produce higher uptake rates and more durable learning gains than simple error flagging or direct correction [22], [37], [38]. This finding has important implications for system design, suggesting that investment in feedback quality is at least as important as improvements in error detection accuracy.

The role of the teacher in AI-enhanced language instruction deserves particular emphasis. While AI can support autonomous learning and provide immediate, individualized feedback, human teachers remain essential for understanding learner motivation, mediating cultural knowledge, designing meaningful communicative activities, and exercising professional judgment [9], [14]. The most promising implementations of AI in language education position technology as a tool that complements and enhances human instruction rather than replacing it [20], [44].

Looking ahead, advances in neural networks, transformer architectures, and large language models will continue to expand the capabilities of AI-based language instruction [11], [45]. However, the enduring principles of effective language teaching meaningful practice, appropriate feedback, attention to individual differences, and integration with human instruction will continue to provide the pedagogical foundation upon which technological innovations must be built [1], [9]. Practitioners are encouraged to remain open to technological innovations while maintaining critical awareness of their limitations, ensuring that the deployment of AI in language education serves the fundamental goal of supporting learners in their journey toward communicative competence [8], [47].

References

[1] M. Warschauer and D. Grimes, "Audience, authorship, and artifact: The emergent semiotics of Web 2.0," Annual Review of Applied Linguistics, vol. 27, pp. 1-23, 2007.
[2] J. Burstein, "The e-rater scoring engine: Automated essay scoring with natural language processing," in Automated Essay Scoring, M. D. Shermis and J. Burstein, Eds. Mahwah, NJ: Lawrence Erlbaum, pp. 113-121, 2003.
[3] M. Warschauer and D. Grimes, "Automated writing assessment in the classroom," Pedagogies: An International Journal, vol. 3, no. 1, pp. 22-36, 2008.
[4] J. Burstein, J. Tetreault, and N. Madnani, "The e-rater automated essay scoring system," in Handbook of Automated Essay Evaluation, M. D. Shermis and J. Burstein, Eds. New York: Routledge, pp. 55-67, 2013.
[5] K. VanLehn, "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems," Educational Psychologist, vol. 46, no. 4, pp. 197-221, 2011.
[6] J. Li, S. Link, and V. Hegelheimer, "Rethinking the role of automated writing evaluation in ESL writing instruction," Journal of Second Language Writing, vol. 27, pp. 1-18, 2015.
[7] A. J. Kellogg, "A meta-analysis on the effectiveness of automated essay scoring," Dissertation Abstracts International, vol. 75, no. 4-A, 2014.
[8] A. Zwaan and M. Verhoef, "Ethics of artificial intelligence in education," in Encyclopedia of Educational Philosophy and Theory, M. Peters, Ed. Singapore: Springer, pp. 1-6, 2018.
[9] R. Godwin-Jones, "Emerging technologies: Data-informed language learning," Language Learning and Technology, vol. 21, no. 3, pp. 9-27, 2017.
[10] D. Grimes and M. Warschauer, "Learning with laptops: A multi-method case study," Journal of Educational Computing Research, vol. 38, no. 3, pp. 305-332, 2008.
[11] S. Crossley, K. Kyle, and D. McNamara, "Sentiment analysis and social cognition engine (SEANCE)," Behavior Research Methods, vol. 49, no. 3, pp. 803-821, 2017.
[12] D. Ferris, Treatment of Error in Second Language Student Writing, 2nd ed. Ann Arbor: University of Michigan Press, 2011.
[13] J. Ranalli, "Automated written corrective feedback: How well can students make use of it?" Computer Assisted Language Learning, vol. 31, no. 7, pp. 653-674, 2018.
[14] C. A. Chapelle and S. Sauro, Eds., The Handbook of Technology and Second Language Teaching and Learning. Hoboken, NJ: Wiley-Blackwell, 2017.
[15] M. Warschauer and R. Kern, Eds., Network-Based Language Teaching: Concepts and Practice. Cambridge: Cambridge University Press, 2000.
[16] T. Heift and M. Schulze, Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues. New York: Routledge, 2007.
[17] T. Heift and M. Schulze, "Tutorial CALL and feedback," in The Handbook of Technology and Second Language Teaching, C. A. Chapelle and S. Sauro, Eds. Hoboken: Wiley-Blackwell, pp. 94-110, 2017.
[18] P. M. McCarthy and S. Jarvis, "MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment," Behavior Research Methods, vol. 42, no. 2, pp. 381-392, 2010.
[19] S. Dikli, "An overview of automated scoring of essays," Journal of Technology, Learning, and Assessment, vol. 5, no. 1, pp. 1-35, 2006.
[20] M. Levy and G. Stockwell, CALL Dimensions: Options and Issues in Computer-Assisted Language Learning. Mahwah, NJ: Lawrence Erlbaum, 2006.

[21] B. P. Woolf, Building Intelligent Interactive Tutors. Burlington, MA: Morgan Kaufmann, 2009.
[22] T. Heift, "Corrective feedback and learner uptake in CALL," ReCALL, vol. 16, no. 2, pp. 416-431, 2004.
[23] D. Meurers, "Natural language processing and language learning," in Encyclopedia of Applied Linguistics, C. A. Chapelle, Ed. Oxford: Wiley-Blackwell, pp. 4193-4205, 2012.
[24] N. Madnani and J. Burstein, "NLP approaches to the analysis of student writing," in Handbook of Natural Language Processing and Machine Translation, J. Olive et al., Eds. New York: Springer, pp. 439-445, 2011.
[25] C. Leacock, M. Chodorow, M. Gamon, and J. Tetreault, Automated Grammatical Error Detection for Language Learners, 2nd ed. San Rafael, CA: Morgan and Claypool, 2014.
[26] D. Dahlmeier and H. T. Ng, "Better evaluation for grammatical error correction," in Proc. NAACL-HLT, pp. 568-572, 2012.
[27] C. Napoles, K. Sakaguchi, and J. Tetreault, "JFLEG: A fluency corpus and benchmark for grammatical error correction," in Proc. EACL, pp. 229-234, 2017.
[28] Z. Yuan and T. Briscoe, "Grammatical error correction using neural machine translation," in Proc. NAACL-HLT, pp. 380-386, 2016.
[29] E. B. Page, "The imminence of grading essays by computer," Phi Delta Kappan, vol. 47, pp. 238-243, 1966.
[30] M. D. Shermis and J. Burstein, Eds., Handbook of Automated Essay Evaluation: Current Applications and New Directions. New York: Routledge, 2013.
[31] L. M. Rudner and T. Liang, "Automated essay scoring using Bayes' theorem," Journal of Technology, Learning, and Assessment, vol. 1, no. 2, pp. 3-21, 2002.
[32] S. Koltovskaia, "Student engagement with automated written corrective feedback provided by Grammarly: A multiple case study," Assessing Writing, vol. 44, art. 100450, 2020.
[33] L. Perelman, "When the state of the art is counting words," Assessing Writing, vol. 21, pp. 104-111, 2014.
[34] E. Choi, "The impact of automated essay scoring on learner writing behavior," Computer Assisted Language Learning, vol. 30, no. 5, pp. 386-405, 2017.
[35] J. R. Anderson, A. T. Corbett, K. R. Koedinger, and R. Pelletier, "Cognitive tutors: Lessons learned," Journal of the Learning Sciences, vol. 4, no. 2, pp. 167-207, 1995.
[36] A. T. Corbett and J. R. Anderson, "Knowledge tracing: Modeling the acquisition of procedural knowledge," User Modeling and User-Adapted Interaction, vol. 4, pp. 253-278, 1994.
[37] R. Ellis, "A typology of written corrective feedback types," ELT Journal, vol. 63, no. 2, pp. 97-107, 2009.
[38] J. Bitchener and U. Knoch, "The contribution of written corrective feedback to language development," Journal of Second Language Writing, vol. 19, no. 4, pp. 207-217, 2010.
[39] D. Ferris, "Second language writing research and written corrective feedback in SLA," Studies in Second Language Acquisition, vol. 32, no. 2, pp. 181-201, 2010.
[40] J. Bitchener and U. Knoch, "Raising the linguistic accuracy level of advanced L2 writers with written corrective feedback," Journal of Second Language Writing, vol. 19, pp. 207-217, 2010.
[41] P. Brusilovsky, "Adaptive hypermedia," User Modeling and User-Adapted Interaction, vol. 11, no. 1-2, pp. 87-110, 2001.
[42] B. S. Bloom, "Learning for mastery," Evaluation Comment, vol. 1, no. 2, pp. 1-12, 1968.
[43] N. Kornell, "Optimising learning using flashcards: Spacing is more effective than cramming," Applied Cognitive Psychology, vol. 23, no. 9, pp. 1297-1317, 2009.
[44] L. Fryer and R. Carpenter, "Bots as language learning tools," Language Learning and Technology, vol. 10, no. 3, pp. 8-14, 2006.
[45] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT, pp. 4171-4186, 2019.
[46] R. Godwin-Jones, "Second language writing online: An update," Language Learning and Technology, vol. 22, no. 1, pp. 1-15, 2018.
[47] N. Selwyn, "Should robots replace teachers? AI and the future of education," Cambridge: Polity Press, 2019.