**International Journal of Linguistics, Literature and Translation**
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

**| RESEARCH ARTICLE**

# A Corpus-Based Multidimensional Analysis of Linguistic Features between Human-Authored and ChatGPT-Generated Compositions

**Jinliang Wu**
*Undergraduate Student, School of Foreign Languages, Guizhou Medical University, Guiyang, China*
**Corresponding Author**: Jinliang Wu, **E-mail**: 2093313576@qq.com

**| ABSTRACT**

This study presents a corpus-based multidimensional comparative analysis of linguistic features in human-authored and ChatGPT-generated English compositions, with a focus on four core dimensions: lexical difficulty, syntactic complexity, textual cohesion, and error patterns. A total of 120 compositions were analyzed—60 produced by ChatGPT-4 and 60 authored by Chinese L2 English learners from the Ten-thousand English Compositions of Chinese Learners corpus—equally distributed across three educational proficiency levels: primary, secondary, and tertiary. Quantitative analyses indicate that human-authored compositions exhibit a progressive increase in lexical complexity aligned with educational advancement, while ChatGPT-generated texts demonstrate limited differentiation between primary and secondary levels, followed by a sharp lexical elevation at the tertiary level. This pattern suggests an algorithmic reliance on generalized discourse rather than sensitivity to developmental variation. In terms of syntactic complexity, ChatGPT consistently produces structurally uniform texts with high usage of subordinate clauses and logical subordination, whereas human writing displays greater contextual flexibility, albeit with occasional simplification. Regarding textual cohesion, ChatGPT-generated compositions—particularly at the tertiary level—rely heavily on overt logical connectors and referential markers, resulting in structurally coherent but stylistically formulaic discourse. In contrast, human-authored texts, while sometimes lacking explicit cohesion markers, employ more nuanced devices such as collocations and implicit semantic links. Error analysis reveals a near absence of grammatical, lexical, and orthographic errors in ChatGPT outputs, contrasting with the relatively high error frequency in human compositions, especially at lower proficiency levels. These findings highlight ChatGPT's strengths in producing grammatically accurate and syntactically complex texts, yet also underscore its limitations in mimicking authentic learner development and stylistic variability. The study concludes that while generative AI can serve as an effective auxiliary tool in L2 writing instruction, its pedagogical integration should be carefully calibrated to avoid undermining learners' development of rhetorical sensitivity, authorial voice, and context-appropriate expression.

**| KEYWORDS**

Second Language Writing; ChatGPT-Generated Compositions; Human-Authored Compositions; Corpus-Based Analysis; Educational Proficiency Levels

**| ARTICLE INFORMATION**

**ACCEPTED:** 01 April 2025 **PUBLISHED:** 12 May 2025 **DOI:** 10.32996/ijllt.2025.8.5.10

## 1. Introduction

Recent advancements in artificial intelligence (AI) have spurred the development of significantly enhanced–and, in some cases, entirely novel–digital writing tools (Godwin-Jones, 2022). Generative AI systems like ChatGPT have emerged as pivotal resources in second-language (L2) writing contexts, prompting widespread interest in their capacity to mimic human linguistic performance.

Proponents highlight AI's ability to produce contextually appropriate, grammatically coherent outputs that may be indistinguishable to readers. Critics, however, contend that such texts often exhibit a "neutral" or "style-flattened" tone,

characterized by uniform syntactic structures, conservative lexical choices, and an absence of a distinct authorial voice (Amirjalili et al., 2024). This stylistic uniformity, while enhancing clarity, contrasts sharply with the personalized nature of human composition–especially in L2 writing, where authors integrate cultural references, personal experiences, and stylistic nuances to convey identity and communicative intent. In turn, AI's "correct but characterless" outputs raise personalized concerns: over-reliance on such tools may hinder learners from developing unique writing voices or engaging in deep reflective processes. This gap underscores the need to evaluate whether AI can adapt to register-specific norms that govern human writing. Guided by this need, the study aims to assess the human-like quality of ChatGPT-generated compositions through a quantitative, multidimensional framework (Sardinha, 2024). This goal was operationalized by comparing AI outputs with human-written counterparts across dimensions of error analysis, syntactic complexity, lexical difficulty, and cohesion devices–key indicators of the differences between human and AI writing. These choices reflect our focus on how AI navigates situational demands (e.g., formality, audience awareness), which are central to effective communication but understudied in existing literature.

Crucially, linguistic features such as error analysis, syntactic complexity, lexical difficulty, and textual cohesion are pivotal for register variation and form an essential part of this study's scope, as they directly reflect how language adapts to communicative contexts. While these features are actively examined, however, the investigation is constrained by resource and scope limitations that preclude us from exploring two additional dimensions: the influence of register-specific contexts on language use and the natural fluency of textual expression, both of which remain unaddressed in this exploratory analysis. Instead, we focus on dimensions most directly tied to the perceived "human-likeness" of AI outputs in real-world interactions, deferring the unstudied contextual and fluency-related aspects to future research. Our framework aligns with text-linguistic theories emphasizing context-driven language use, hypothesizing systematic variations in AI performance across communicative situations–an assumption that requires empirical validation beyond the current study's constraints.

## 2. Literature Review

The emergence of generative artificial intelligence (AI) technologies has generated significant scholarly interest concerning their prospective applications in language education, with particular emphasis on second language (L2) writing instruction. Generative AI models, such as ChatGPT, have been recognized for their capacity to generate grammatically precise and contextually relevant text that closely mirrors human-authored writing in terms of structure and coherence (Rasul et al., 2023). These advanced language models have been integrated into various writing assistance tools, automated feedback systems, and serve as interactive learning companions to support second language (L2) learners in their writing development (Yan, 2023). A growing corpus of research has identified a salient stylistic limitation inherent in ChatGPT-generated compositions: their marked tendency toward stylistic neutrality. Such outputs are frequently characterized as "accurate yet devoid of distinctive voice," often lacking emotional subtlety, explicit authorial stance, and individualized expression (Amirjalili et al., 2024). This depersonalization is particularly evident in the limited deployment of modal verbs, hedging devices, and epistemic markers–linguistic features that human authors routinely employ to convey attitude, degree of certainty, and reader engagement (Mo & Crosthwaite, 2025). Amidst writing, these linguistic elements fulfill not only stylistic but also critical rhetorical functions, enabling scholars to negotiate meaning, articulate nuanced positions, and establish epistemic credibility. Human-authored L2 writing, by contrast, tends to be rich in personal voice, with language shaped by sociocultural identity, emotional expression, and communicative purpose (Zhao, 2019) L2 learners often draw upon their backgrounds and experiences to develop a unique writing identity, which can be seen in their lexical variation, rhetorical strategies, and stance-taking language The growing concern is that frequent reliance on AI tools may diminish opportunities for L2 learners to experiment with voice, build rhetorical awareness, and engage in reflective thinking. Several scholars have attempted to assess the effectiveness of AI in writing through performance comparison. Sardinha & Pinto (2020) have proposed multidimensional frameworks for examining linguistic variation in compositions, which have been adapted to explore ChatGPT-generated language.

Notwithstanding the growing body of research in this area, a notable void exists in systematic, register-specific comparative investigations between human-composed and ChatGPT-generated compositions. The majority of existing studies have centered on assessments via readability metrics or grammatical evaluations, rather than engaging in meticulous text-linguistic analyses (Xia et al., 2019; Warschauer, 2023). Furthermore, scholarly work has largely neglected the potential impacts of stylistic uniformity in AI-produced writing on reader engagement, communicative purposes, and second language acquisition outcomes.

This study aims to systematically compare human and ChatGPT-generated writing across linguistic dimensions–lexical difficulty, syntactic complexity, textual cohesion, and error patterns–to evaluate AI's suitability as a pedagogical tool for fostering second language (L2) writing development. The research is anchored in an integrated theoretical framework that merges discourse cohesion theory, interlanguage theory, error analysis, and quantitative approaches to linguistic complexity, providing a multi-dimensional lens to identify stylistic and pragmatic constraints in AI outputs within educational contexts. Lexical complexity, measured via type-token ratio and lexical density (Flesch, 1948), further reveals AI's reliance on generic terms and statistical

approximations, contrasting with human writers' context-sensitive vocabulary diversification. Syntactic complexity is analyzed using metrics such as Lu's (2010) Mean Length of Clause (MLC) and T-unit complexity, revealing that proficient human writers adapt structures contextually (e.g., transforming Chinese paratactic sentences into English hypotactic complexes), while AI exhibits polarized tendencies—either oversimplifying with repetitive simple sentences or overcomplicating with mechanical clause nesting, resulting in logical disconnections. Drawing on Halliday and Hasan's (1976) discourse cohesion theory, it examines how human writers enhance coherence through strategic use of grammatical devices (e.g., specifying referents to clarify "it"/"this" and deploying causal connectives like "therefore") and lexical resources (e.g., synonymy and collocation), in contrast to AI's tendency toward formulaic or redundant cohesive patterns that often lead to under-explicitation or contextual vagueness. Corder's (1981) interlanguage theory frames AI output as an "algorithmic interlanguage," highlighting its formal grammatical correctness but pragmatic limitations, analogous to learners' systematic errors from overgeneralization or L1 transfer (e.g., misusing "imply" for "suggest" or failing to distinguish register-specific collocations like "data analysis" vs. "information analysis"). By integrating these theories with corpus-based quantification and factor analysis, the study situates AI outputs within the functional profiles of L2 academic writing (e.g., balancing "involved" personal engagement and "informational" objectivity), critically evaluating whether AI can effectively scaffold learners' syntactic sophistication, lexical richness, and error awareness while addressing its limitations in capturing contextual and cultural subtleties.

## 3. Methodology

This study employs a corpus-based and multidimensional comparative design to evaluate linguistic divergences between ChatGPT-generated and human-authored compositions in second-language (L2) English writing. Anchored in Flesch's (1948) readability scores, Lu's (2010) syntactic complexity metrics, Halliday and Hasan's (1976) cohesion theory, and Corder's (1981) interlanguage framework, the analysis targets four dimensions: lexical difficulty, syntactic complexity, textual cohesion, and error patterns. To enable fine-grained comparison, six parallel sub-corpora were constructed based on educational level (primary, secondary, tertiary) and authorship type (AI vs. human), with 20 compositions per sub-corpus.

### *3.1 Corpora*

The AI corpus was generated using ChatGPT-4, prompted with standardized L2 writing tasks tailored to each educational level. The human corpus consisted of learner-authored compositions drawn from the Ten‐thousand English Compositions of Chinese Learners (the TECCL Corpus). All compositions were matched for genre, topic, and length to ensure comparability. This design supports both intra-level (AI vs. human) and inter-level (primary to tertiary) comparisons, allowing investigation into how ChatGPT-generated writing aligns with human developmental patterns in L2 writing across proficiency stages.

**Table 1. Corpora Size of Six Groups (words)**

| File Number | Human-authored compositions | | | ChatGPT-generated compositions | | |
|---|---|---|---|---|---|---|
| | primary | secondary | tertiary | primary | secondary | tertiary |
| 1 | 115 | 170 | 240 | 108 | 175 | 242 |
| 2 | 82 | 105 | 348 | 85 | 88 | 304 |
| 3 | 144 | 100 | 157 | 142 | 84 | 174 |
| 4 | 79 | 139 | 251 | 81 | 140 | 244 |
| 5 | 162 | 241 | 320 | 140 | 203 | 314 |
| 6 | 108 | 113 | 344 | 85 | 91 | 332 |
| 7 | 125 | 142 | 318 | 109 | 118 | 308 |
| 8 | 110 | 116 | 325 | 87 | 104 | 320 |
| 9 | 112 | 114 | 349 | 92 | 95 | 343 |
| 10 | 75 | 130 | 356 | 53 | 112 | 345 |
| 11 | 156 | 163 | 422 | 129 | 159 | 423 |
| 12 | 102 | 129 | 351 | 87 | 109 | 361 |
| 13 | 99 | 125 | 477 | 84 | 131 | 472 |
| 14 | 116 | 116 | 456 | 122 | 115 | 469 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 15 | 122 | 81 | 347 | 101 | 85 | 330 |
| 16 | 129 | 104 | 314 | 124 | 106 | 281 |
| 17 | 99 | 160 | 319 | 81 | 142 | 315 |
| 18 | 142 | 323 | 489 | 130 | 290 | 514 |
| 19 | 117 | 121 | 515 | 98 | 123 | 571 |
| 20 | 146 | 167 | 301 | 150 | 177 | 284 |
| Average | 117 | 143 | 350 | 104 | 132 | 347 |

### *3.2 Data Processing*

Lexical difficulty was measured using the BFSU Readability Analyzer 3, which calculates type-token ratio, lexical density, and the distribution of high- and low-frequency words. Human-authored writing often relied heavily on repetitive and familiar terms (e.g., freedom, network, real name), whereas AI outputs demonstrated broader synonym use (e.g., cyberbullying, inappropriate content, trustworthy community) and more abstract vocabulary.

Syntactic complexity was assessed using BFSU Stanford Parser and BFSU Syntactic Complexity Analyzer, focusing on metrics such as mean clause length, the use of subordinate clauses, and the variety of syntactic constructions. While human compositions displayed erratic structures and frequent simplifications, ChatGPT-generated compositions featured more consistent use of complex sentence patterns and logical subordination (e.g., concessive and causal constructions).

Textual cohesion were evaluated through manual annotation and ChatGPT-assisted analysis. AI writing made extensive use of logical connectors (e.g., on the one hand, therefore, in conclusion), ensuring structural coherence. In contrast, human writing frequently exhibited underuse of cohesive markers and illogical progression (e.g., abrupt transitions like Suppose of all of us do use⋯).

Error analysis was conducted to examine the frequency and types of grammatical inaccuracies in human-written compositions, such as subject-verb disagreement (e.g., a fair proportion of people⋯ has not⋯), incomplete syntactic structures, and pronoun misuse (e.g., you real name). In contrast, ChatGPT-generated–produced via ChatGPT-4–exhibited minimal surface-level errors but tended to lack individualized linguistic features. ChatGPT itself was employed to qualitatively assess authorial voice and stylistic distinctiveness in both corpora.

## 4. Results

### *4.1 Comparison of Lexical Difficulty between Human-authored and ChatGPT-generated Compositions*

Table 2 presents a comparative analysis of lexical difficulty between human-authored and ChatGPT-generated compositions, evaluated across two key dimensions: Reading Ease and Grade Level, categorized by educational levels (primary, secondary, and tertiary). The Reading Ease metric, typically derived from formulas such as the Flesch Reading Ease score, quantifies the readability of a text on a scale from 0 to 100, where higher scores indicate greater ease of comprehension. This measure accounts for factors such as sentence length and word complexity, providing insight into the accessibility of the text for readers at different educational stages. The Grade Level metric, often based on the Flesch-Kincaid Grade Level formula, estimates the U.S. educational grade level required to understand the text, with lower values indicating simpler content suitable for younger readers and higher values reflecting greater complexity appropriate for advanced readers. Together, these dimensions offer a robust framework for assessing the readability and linguistic complexity of the compositions, highlighting differences in lexical difficulty between human-authored and ChatGPT-generated compositions across varying educational contexts.

**Table 2. Lexical Difficulty between Human-authored and ChatGPT-generated Compositions**

| Lexical difficulty | Human-authored compositions | | | ChatGPT-generated compositions | | |
|---|---|---|---|---|---|---|
| | primary | secondary | tertiary | primary | secondary | tertiary |
| Reading Ease | 90.681 | 71.319 | 52.510 | 85.87 | 74.298 | 40.524 |
| Grade Level | 2.835 | 7.368 | 11.761 | 3.477 | 5.783 | 12.038 |

From Table 2, it can be concluded that human-authored compositions show a clear decline in reading ease from primary (90.681) to tertiary (52.510), accompanied by a steady increase in grade level. In contrast, ChatGPT-generated compositions present a

smaller drop in reading ease from primary (85.87) to secondary (74.298), but a sharp decline at tertiary level (40.524), suggesting the model distinctly elevates lexical complexity for higher education while showing limited differentiation between lower levels.

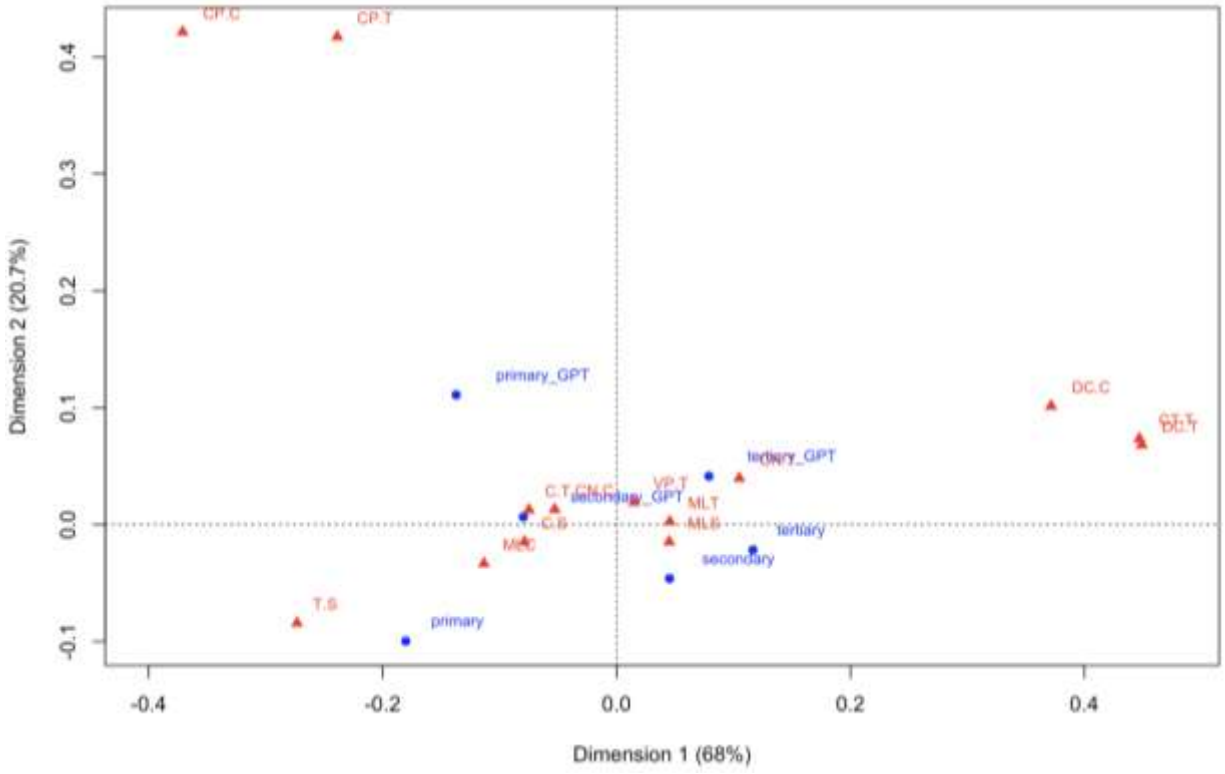
Notably, in ChatGPT-generated compositions, the negligible variation in lexical difficulty between primary and secondary school levels implies that the model lacks refined sensitivity to the developmental disparities in L2 writing abilities across these stages. This might be caused by the limited depiction of age-specific language patterns in its training data, especially for younger learners, leading to generalized outputs with inadequate distinctions. In contrast, the significant rise in lexical sophistication at the tertiary level shows that the model associates "university writing" with advanced vocabulary, abstract ideas, and formal academic structures.

This change probably originates from ChatGPT's exposure to extensive corpora that highlight scholarly language, thus strengthening its internalized schema, which links higher education to greater linguistic complexity. Furthermore, when given university-level tasks, the model may demonstrate a form of compensatory overgeneration, systematically using elevated vocabulary and syntactic structures to meet its perceived standards of academic excellence. These inclinations indicate that ChatGPT inherently presumes a certain baseline of academic literacy for the university level, affecting its output regardless of the actual variations among learners.

### *4.2 Comparison of Syntactic complexity between Human-authored and ChatGPT-generated Compositions*

Figure 1 compares syntactic complexity across six subgroups by authorship and educational level, using fourteen indices from Lu (2010), such as MLC, MLS, C/T, and CN/T. Blue points represent data groups; red points indicate syntactic metrics. Closer proximity implies stronger associations and more consistent structural use, while greater dispersion reflects syntactic variability. The analysis captures clause length, subordination, coordination, and phrasal complexity as key indicators of developmental and stylistic divergence.

**Figure 1. Corresponding Analysis of Syntactic complexity between Human-authored Compositions and ChatGPT-generated Compositions**




Human-authored compositions at the primary level of language proficiency, here referred to as Blue Primary Points, consistently exhibit linguistic profiles characterized by low syntactic complexity. Quantitative analysis reveals that these texts cluster tightly around short Mean Length of Sentence (MLS) and Mean Length of T-unit (MLT) values. Furthermore, these compositions demonstrate minimal use of subordination, as indicated by low Dependent Clauses per Clause (DC/C), alongside limited coordination, reflected in low Coordinate Phrases per Clause (CP/C). Such patterns are emblematic of early second language (L2) learners, who predominantly utilize simple independent clauses (e.g., “She went to school. She studied math.”) and basic T-units.

This syntactic simplicity serves to prioritize semantic clarity over structural complexity, consistent with findings in the literature (Ortega, 2003).

In parallel, ChatGPT-generated compositions at the Blue Primary Points level are designed to replicate these linguistic characteristics. The generated texts maintain short MLS and MLT measures, alongside minimal subordination and coordination, thereby mirroring the structural simplicity observed in human-authored compositions by early L2 learners. This alignment suggests that ChatGPT effectively models the syntactic features typical of elementary proficiency writing, emphasizing clarity and straightforward expression.

The model positions slightly above its human counterparts in terms of Clauses per T-unit (C/T) and Mean Length of Clause (MLC), reflecting occasional deployment of simple subordinate constructions (e.g., "She played outside when the sun shone"). Nonetheless, its close alignment with human primary-level data indicates a tendency to conform to simplified syntactic patterns, eschewing more complex clause structures in favor of accessibility. This approach corresponds with fundamental task demands but consequently constrains the range of structural variation exhibited in the compositions.

At the secondary level, representing intermediate proficiency, human-authored compositions (indicated by the blue secondary points) demonstrate a moderate shift toward greater syntactic complexity. This is evidenced by increases in Mean Length of T-unit (MLT), reflecting longer and more elaborated T-units, as well as higher Clauses per T-unit (C/T), indicating more frequent clause combination within single T-units. Such patterns correspond to the emerging use of compound sentence constructions—for example, “He studied hard, and he passed the exam”—alongside the incorporation of simple relative clauses, as illustrated by “The book that I read was interesting.” These developments collectively suggest a growing ability among intermediate learners to integrate multiple clauses and enhance sentence complexity through coordination and subordination.

A significant increase in Coordinate Phrases per T-unit (CP/T) reflects learners' efforts to improve textual cohesion through phrasal coordination (e.g., "quickly and efficiently"), consistent with Lu (2010)’s observations of intermediate learners expanding their syntactic repertoire. In contrast, ChatGPT-generated compositions (denoted as Blue Secondary Points) exhibit minimal divergence from primary-level groups, demonstrating only slight rises in T-units per Sentence (T/S) and Dependent Clauses per T-unit (DC/T), indicating limited advancement in clausal coordination and subordination complexity.

The model prioritizes fluency via standardized structures (e.g., predictable "because" or "although" clauses) rather than stage-specific complexity, suggesting a one-size-fits-all strategy that may not align with intermediate learners’ need to practice varied syntactic forms (e.g., conditional clauses, participial phrases).

At the tertiary level (advanced proficiency), human-authored compositions predominantly align with metrics indicative of high linguistic complexity, characterized by the employment of sophisticated syntactic structures. Key evaluative indicators encompass elevated measures of mean length of clause (MLC) and mean length of T-unit (MLT).

### *4.3 Comparison of Textual Cohesion between Human-authored and ChatGPT-generated Compositions*

Table 3 provides a quantitative comparison of textual cohesion strategies in human- and ChatGPT-generated compositions across primary, secondary, and tertiary levels. Cohesive devices are categorized into grammatical (reference, substitution, ellipsis, conjunction) and lexical (repetition, synonymy, hyponymy, meronymy, collocation) types. The analysis interprets these patterns through the lens of established cohesion theories, with attention to their implications for AI-generated discourse structure and stylistic tendencies

**Table 3. Textual Cohesion between Human-authored and ChatGPT-generated Compositions**

| Cohesion Devices | Human-authored compositions | | | ChatGPT-generated compositions | | |
|---|---|---|---|---|---|---|
| | primary | secondary | tertiary | primary | secondary | tertiary |
| Reference | 416 | 371 | 466 | 402 | 26 | 898 |
| Substitutions and Ellipsis | 0 | 0 | 0 | 0 | 0 | 0 |
| Conjunction | 0 | 28 | 66 | 1 | 16 | 248 |
| Repetition | 101 | 89 | 124 | 118 | 121 | 248 |
| Synonymy | 0 | 0 | 0 | 0 | 0 | 0 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Hyponymy | 0 | 0 | 0 | 0 | 4 | 67 |
| Meronymy | 0 | 0 | 0 | 0 | 2 | 0 |
| Collocation | 74 | 64 | 64 | 86 | 86 | 251 |

Grammatical and lexical cohesion patterns reveal notable divergences between human- and ChatGPT-generated texts. Human-authored compositions demonstrate relatively consistent reference use across levels, whereas ChatGPT shows minimal reference at the secondary level (26) but a sharp increase at the tertiary level (898), indicating possible overuse and referential redundancy in advanced AI outputs. This reflects the model's alignment with academic discourse norms that emphasize referential precision.

Substitution and ellipsis are entirely absent in both corpora, suggesting a general avoidance of these syntactically demanding strategies—likely due to their complexity and the tendency in both human L2 writing and AI outputs to favor syntactic explicitness.

Conjunction usage in human texts increases moderately with proficiency, aligning with developmental patterns. In contrast, ChatGPT employs disproportionately high conjunctions at the tertiary level (248), signaling reliance on formulaic linking devices drawn from formal registers.

Lexical cohesion analysis further highlights ChatGPT's tendency toward repetition and collocation at the tertiary level, possibly to maintain topical focus and surface fluency. While human writers use repetition more selectively, ChatGPT frequently recycles lexical items, potentially at the expense of variation. Additionally, synonymy and hyponymy—largely absent in human texts—appear selectively in AI outputs, reflecting algorithmic efforts at lexical diversification.

Overall, ChatGPT-generated compositions display dense but formulaic cohesion at advanced levels, shaped by training data conventions rather than developmental appropriateness. In contrast, human-authored texts exhibit a more gradual and balanced progression in cohesion strategies, indicative of authentic L2 acquisition patterns.

### *4.4 Comparison of Error Patterns between Human-authored and ChatGPT-generated Compositions*

Table 4 provided compares the frequency of different error types in human-authored and ChatGPT-generated compositions across primary, secondary, and tertiary levels. It's evident that human-authored compositions contain a significantly higher number of errors in all categories compared to ChatGPT-generated compositions which exhibit zero errors.

**Table 4. Error Patterns between Human-authored and ChatGPT-generated Compositions**

| Error Patterns | Human-authored compositions | | | ChatGPT-generated compositions | | |
|---|---|---|---|---|---|---|
| | primary | secondary | tertiary | primary | secondary | tertiary |
| grammatical errors | 152 | 142 | 142 | 0 | 0 | 0 |
| lexical errors | 7 | 9 | 6 | 0 | 0 | 0 |
| spelling errors | 43 | 24 | 0 | 0 | 0 | 0 |
| discourse errors） | 6 | 0 | 0 | 0 | 0 | 0 |
| addition | 13 | 19 | 20 | 0 | 0 | 0 |
| omission | 66 | 58 | 57 | 0 | 0 | 0 |
| misformation | 105 | 68 | 65 | 0 | 0 | 0 |
| misordering | 0 | 0 | 0 | 0 | 0 | 0 |

The analysis encompasses a range of error categories, including grammatical inaccuracies, lexical mistakes, spelling errors, discourse-related issues, additions, omissions, misformations, and misordering. Each category represents a distinct aspect of linguistic performance, capturing the challenges faced by human writers at different educational stages as well as the capabilities of ChatGPT in producing error-free text. For human-authored compositions, the data reveal a notable presence of errors across all educational levels, reflecting the inherent variability in human writing proficiency. In contrast, ChatGPT-generated compositions demonstrate a remarkable absence of errors across the same categories, underscoring the AI's precision in adhering to linguistic norms.

Through this comparative framework, Table 4 illuminates the significant differences in error profiles between the two types of compositions. The findings suggest that human-authored compositions, regardless of the educational level of the writer, are prone to a variety of errors, which may impact their overall clarity and correctness. Conversely, the zero-error performance of ChatGPT-generated compositions points to the AI's ability to consistently produce linguistically accurate content, potentially making it a more reliable option for applications requiring high levels of precision. This analysis contributes to a deeper understanding of the strengths and limitations of both human and AI-generated compositions in terms of error management.

## 5. Conclusion

This study presents a corpus-based, multidimensional comparative analysis of linguistic features in human-authored and ChatGPT-generated English compositions, stratified across three educational proficiency levels. The results demonstrate that while ChatGPT-generated texts consistently outperform human-authored counterparts in terms of surface-level grammatical accuracy, syntactic regularity, and cohesion density, they exhibit limited responsiveness to developmental linguistic variation and register-sensitive stylistic differentiation. Specifically, the AI-generated texts fail to reflect the gradual progression typically observed in human learners' writing from primary to tertiary levels, instead displaying either stylistic uniformity or disproportionate lexical and cohesive elaboration at higher academic levels.

Moreover, the over-reliance on formulaic cohesive devices and standardized syntactic constructions in AI outputs suggests a predominance of algorithmically derived language patterns, which may inadvertently obscure individual rhetorical intent and pragmatic nuance. In contrast, human-authored compositions—despite a higher incidence of linguistic errors—manifest more contextually adaptive strategies and a wider range of authorial voice expressions, particularly in intermediate and advanced proficiency stages. These observations underscore the dual nature of AI as both a facilitator and a potential constraint within the domain of L2 writing development.

Consequently, the findings advocate for a balanced and pedagogically informed integration of AI writing tools in second language instruction. While the precision and fluency of ChatGPT-generated texts can provide valuable models for syntactic structuring and grammatical reinforcement, their stylistic limitations necessitate instructional mediation that fosters learners' engagement with authentic voice construction, contextual awareness, and genre-appropriate rhetorical strategies. Future research should further investigate how generative language models may be calibrated or fine-tuned to accommodate learner-specific linguistic trajectories and to support more dynamic and personalized writing development within diverse educational contexts.

**Conflicts of Interest**: The author declare no conflict of interest.


## References

[1] Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the Boundaries of Authorship: A Comparative Analysis of AI-generated Text and Human Academic Writing in English Literature. *Frontiers in Education, 9*, 1347421. https://doi.org/10.3389/feduc.2024.1347421

[2] Corder, S. P. (1981). *Error Analysis and Interlanguage*. Oxford University Press.

[3] Flesch, R. (1948). A New Readability Yardstick. *Journal of Applied Psychology, 32*(3), 221. https://doi.org/10.1037/h0057532

[4] Godwin-Jones, R. (2022). Partnering with AI: Intelligent Writing Assistance and Instructed Language Learning. *Language Learning & Technology, 26*(2), 5-14.

[5] Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Longman.

[6] Lu, X. (2010). Syntactic Complexity Measures and Their Relationship to L2 Writing Quality. *TESOL Quarterly, 44*(3), 492–522. https://doi.org/10.1093/applin/24.4.492

[7] Mo, Z., & Crosthwaite, P. (2025). Exploring the Affordances of Generative AI Large Language Models for Stance and Engagement in Academic Writing. *Journal of English for Academic Purposes, 75*, 101499. https://doi.org/10.1016/j.jeap.2025.101499.

[8] Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., ... & Heathcote, L. (2023). The Role of ChatGPT in Higher Education: Benefits, Challenges, and Future Research Directions. *Journal of Applied Learning and Teaching, 6*(1), 41-56.

[9] Sardinha, T. B. (2024). AI-Generated vs Human-Authored Texts: *A Multidimensional Comparison. Applied Corpus Linguistics, 4*(1), 100083.

[10] Sardinha, T. B., & Pinto, M. V. (Eds.). (2020). *Multi-Dimensional Analysis: Research Methods and Current Issues*. Bloomsbury Publishing.

[11] Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The Affordances and Contradictions of AI-Generated Text for Writers of English as a Second or Foreign Language. *Journal of Second Language Writing, 62*. https://doi.org/10.1016/j.jslw.2023.101071

[12] Xia, M., Kochmar, E., & Briscoe, T. (2019). Text Readability Assessment for Second Language Learners. arXiv Preprint arXiv:1906.07580. *Computation and Language*. https://doi.org/10.48550/arXiv.1906.07580

[13] Yan, D. (2023). Impact of ChatGPT on Learners in a L2 Writing Practicum: An Exploratory Investigation. *Education and Information Technologies, 28*(11), 13943-13967.

[14] Zhao, C. G. (2019). Writer Background and Voice Construction in L2 Writing. *Journal of English for Academic Purposes, 37*, 117-126. https://doi.org/10.1016/j.jeap.2018.11.004