Lecture Notes on Language and Literature (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/langl.2026.090120
ISSN 2523-5869 Vol. 9 Num. 1

Characterization and Mechanisms of Lexical Complexity in AI-Generated Texts: A Comparative Corpus-Based Study

Lulu Chen

School of International Education, Tianjin Foreign Studies University, Tianjin, 300270, China

Keywords: AI-generated text, Corpus, L2 writing, Lexical complexity

Abstract: Based on a corpus-based methodology, this study analyzes the intrinsic reasons for the high level of lexical complexity observed in Artificial Intelligence Generated Content (AIGC). The research compares 24 English argumentative essays written by AI with 24 second-language (L2) learner essays reaching the IELTS Writing Task 2 Band 7 level. Under controlled conditions of identical genre and topic, the study performs quantitative statistics across three dimensions: lexical sophistication, semantic abstraction, and information density. Statistical results indicate that the frequency of advanced vocabulary in AI texts is significantly higher, approximately 2.3 times that of human texts. The proportion of abstract nouns reached 9.14%, far exceeding the 3.32% found in human texts, suggesting that AI expressions tend toward nominalization and conceptualization. Regarding overall information organization, the lexical density of AI texts was 69.9%, also surpassing the 60.3% of human texts, reflecting a stronger tendency for information condensation and phrasal structures. The analysis points out that the complexity of AI text primarily stems from its mechanism of selecting vocabulary based on probability distributions. This mechanism favors longer words, abstract nouns, and words with high semantic content, thereby forming a highly compact linguistic surface. Such complexity is essentially a formal feature at the statistical level and is not entirely equivalent to the proficiency levels corresponding to human L2 acquisition. These findings provide empirical references for AI text identification, the refinement of writing evaluation standards, and L2 writing pedagogy.

1. Introduction

1.1 Research Background

With the rapid development of generative artificial intelligence, large language models (LLMs), exemplified by ChatGPT, have demonstrated a high degree of automation in text generation. Relying on large-scale textual pretraining and the probabilistic prediction mechanism of the Transformer architecture, these models are capable of producing structurally complete and stylistically appropriate argumentative essays within a short time. They have already been widely applied in contexts such as academic writing, business communication, and standardized test

preparation [1]. At the level of surface linguistic features, AI-generated discourse often appears highly conventionalized and complex. Existing studies indicate that, in standardized writing tasks, AI-generated corpora tend to exhibit a higher degree of complexity than that of intermediate-level second language learners across several linguistic indices, including lexical richness and syntactic sophistication [2,3]. Specifically, this is reflected in a higher proportion of long words, more frequent use of nominalization, and more compact information structuring. However, such complexity does not arise from the logical construction and contextualized activation of conceptual networks within the human cognitive system; rather, it is an external manifestation of probabilistic distribution patterns acquired through large-scale data pretraining, representing a form of statistical sophistication distinct from natural language acquisition.

At the same time, the relatively low barrier to access and high efficiency of generative AI technologies are exerting tangible impacts on academic integrity and language education [1,4]. Students can generate complete texts that meet scoring criteria with minimal prompts. While this convenience improves writing efficiency, it may also foster cognitive dependence and reduce learners' engagement in independent idea generation and linguistic organization. More importantly, AI-assisted writing has introduced new forms of academic misconduct, blurring the boundaries of traditional plagiarism detection, as machine-generated content and human-revised texts become increasingly difficult to distinguish [5,6]. Although various detection tools, such as GPTZero and DetectGPT, have been developed to address this issue, their accuracy and stability across topics and genres remain limited. Due to substantial overlap between AI-generated texts and high-level human writing in statistical features, existing tools continue to suffer from relatively high rates of false positives and false negatives [7,8].

It is evident that while artificial intelligence is reshaping the landscape of text production, it also poses new challenges to academic integrity and language pedagogy. In response, research in this field is gradually shifting from external detection techniques toward the analysis of intrinsic linguistic features, aiming to systematically capture the micro-level differences between AI-generated and human-written texts. Such efforts are essential for developing more effective identification methods, refining writing assessment systems, and informing pedagogical adjustments [2]. A fine-grained analysis of lexical features constitutes a central component in understanding text properties and conducting discourse classification. Patterns of lexical choice, distribution, and collocation directly reflect the generative logic and cognitive basis of a text. Therefore, a systematic comparison of lexical features between AI-generated discourse and human writing can help explain the underlying causes of the elevated complexity observed in AI texts, reveal the statistical properties of algorithmic generation mechanisms, and identify discriminative features between human and machine-produced texts [3].

In the field of AI text detection, early approaches primarily relied on statistical indicators such as perplexity and burstiness. However, as large language models continue to evolve, the statistical distributions of AI-generated texts increasingly overlap with those of human writing, reducing the effectiveness of detection methods based solely on such features. Gao et al. (2023), by comparing summaries generated by ChatGPT with those written by humans, found that the accuracy of current detection tools declines significantly when applied to AI-generated content that has been post-edited by humans [7]. Shah et al. (2023), employing explainable AI techniques and stylistic features for detection, achieved some success on specific datasets, but their models still exhibited limited generalizability across domains [8]. Collectively, these studies suggest that surface-level statistical features alone are insufficient for reliably distinguishing between human and AI-generated texts, underscoring the need to explore more fine-grained and discriminative features at the intersection of lexical and syntactic levels.

1.2 Theoretical Framework

Current research on the linguistic features of AI-generated texts is primarily grounded in the theoretical perspectives of lexicometry and Systemic Functional Linguistics (SFL). Lexicometry provides methodological support for the extraction and quantitative comparison of textual features, while SFL offers a theoretical basis for understanding how texts encode information, realize metafunctions, and employ grammatical metaphor. Within this interdisciplinary framework, a growing body of empirical studies has demonstrated that large language models significantly outperform typical human writing samples in measures such as lexical density, lexical diversity, and lexical sophistication. In particular, they exhibit notable statistical advantages in deploying low-frequency vocabulary associated with academic registers and in constructing complex noun phrases [15].

In the field of second language (L2) writing research, lexical complexity is widely regarded as a key indicator of writing quality and language development [10,11]. Lu (2010) found a significant positive correlation between lexical richness and the quality of L2 learners' oral narratives, although this relationship varies depending on genre and task type. McCarthy and Jarvis (2010) further showed that different measures of lexical diversity, such as MTLD and HD-D, differ in their stability when applied to text evaluation, and that the choice of metric can directly influence research outcomes [12]. These findings suggest that, when comparing lexical features between human and AI-generated texts, a multidimensional analytical framework incorporating multiple indices is preferable to reliance on a single measurement. To operationalize the combined insights of lexicometry and SFL and to enable a comparable quantitative analysis of lexical features across AI-generated and L2 learner texts, this study adopts three core indicators: the proportion of advanced vocabulary, the proportion of abstract nouns, and lexical density. These indices not only capture key dimensions of textual complexity but also provide a micro-level lens for examining algorithmic generation mechanisms.

Within corpus linguistics, word length is often used as an indirect indicator of lexical complexity [13]. Words with a greater number of characters tend to have lower frequency and higher semantic load, and their appropriate use requires a broader vocabulary repertoire and more precise contextual judgment. According to Biber's multidimensional analysis framework, words consisting of six or more characters show a clear positive correlation with the degree of informational integration in discourse [14]. Following this operational definition, the present study classifies words with a length of no fewer than six characters as advanced vocabulary, using this measure to reflect lexical sophistication and formal complexity [14]. It should be noted that word length is used here as an operational proxy for lexical sophistication rather than a comprehensive measure, primarily for the purpose of ensuring methodological consistency and computational feasibility. Therefore, the results should be interpreted as indicative of formal complexity rather than a direct measure of lexical proficiency.Abstract nouns serve as an important indicator of nominalization and semantic abstraction in discourse, functioning as a key resource for information compression and conceptual expression in academic English. According to the theory of grammatical metaphor in SFL, nominalization transforms dynamic processes into static entities, condensing clause-level structures into single noun phrases and thereby increasing both information density and conceptual layering [9]. In this study, abstract nouns are identified through morphological features, and their proportion is used as a core metric of semantic abstraction. While this morphological approach does not capture all instances of semantic abstraction, it provides a consistent and replicable operationalization suitable for large-scale corpus comparison. Lexical density, in turn, is a macro-level measure of informational load and corresponds directly to the information-packaging mechanism discussed in SFL. Content words-nouns, verbs, adjectives, and adverbs-carry core

semantic meaning, whereas function words-such as prepositions, articles, conjunctions, and auxiliaries-primarily serve grammatical and cohesive functions. A higher lexical density indicates a greater amount of new information per unit of text and a higher degree of semantic saturation [11,13]. This measure captures the overall distribution of content versus function words and provides a direct indication of how AI systems emulate the information-dense characteristics of advanced academic writing through probabilistic prediction.

Based on the above theoretical and empirical background, this study adopts a corpus-based comparative design and employs quantitative methods to systematically identify differences between AI-generated texts and L2 learner texts across three dimensions: lexical sophistication, semantic abstraction, and informational density. It further explores differences in lexical decision-making between AI systems and human L2 learners. The study addresses three main research questions: (1) whether significant differences exist between AI-generated and L2 learner texts in terms of the proportion of advanced vocabulary, abstract nouns, and lexical density; (2) how these differences reflect the fundamental distinction between probabilistic algorithmic mechanisms and human cognitive processes; and (3) what implications these findings hold for AI text detection, the refinement of writing assessment systems, and L2 writing pedagogy.

2. Methodology

2.1 Corpus Compilation

In constructing the human-written corpus, this study strictly adheres to the comparability requirements of contrastive corpus research by selecting texts produced by authentic English learners in standardized writing tasks. The human data are primarily drawn from publicly available IELTS writing samples. After careful screening, all selected texts are Task 2 argumentative essays with scores approximating Band 7. A Band 7 level represents a good command of English, where writers are capable of handling relatively complex linguistic structures and developing detailed arguments. This level is chosen because it excludes overly simplified expressions resulting from low proficiency while also avoiding near-native performance that may obscure characteristics of second language acquisition. As such, it serves as a representative benchmark for upper-intermediate L2 writing ability.

In terms of topic control, all texts are strictly limited to the domains of education and society. These topics are among the most frequently tested in IELTS Task 2 and constitute core areas in academic English writing, with relatively well-defined academic registers. During the selection process, all texts were manually reviewed to exclude essays exhibiting clear template dependence, repetitive content, or insufficient argumentative development. This ensures that the final corpus consists of texts with complete structure, generally coherent argumentation, and natural language use. All essays were independently produced by learners under timed conditions, with each text ranging from 200 to 300 words. A total of 24 essays were included, comprising 6,814 words.

For the construction of the AI-generated corpus, this study employed standardized prompts to elicit responses from ChatGPT (GPT-4 architecture), ensuring consistency with the human corpus in terms of task type, genre conventions, and register requirements. The prompts were framed within the IELTS Writing Task 2 argumentative format, with topics restricted to education and society, a target length of approximately 250 words, and a formal academic style. The specific instruction used was: “Write an IELTS Task 2 essay on an education-related/society-related topic. The essay should be approximately 250 words in length and written in a formal academic style. Present a clear position, develop coherent arguments, and support them with relevant examples. Avoid repetition and use varied vocabulary and sentence structures.” To minimize potential systematic effects arising from model context retention, all 24 essays were generated in independent

sessions, with each interaction limited to a single writing task. From the second to the twenty-fourth essay, an additional constraint-"Use different arguments and examples from typical responses."-was included to enhance diversity in thematic content and argumentation and to reduce formulaic repetition caused by prompt similarity. Following generation, all texts were manually screened to remove samples with repetitive content, highly similar structures, or clear deviations from the argumentative genre. Ultimately, 24 AI-generated essays were retained for analysis, with a total word count of 5,986. The basic word count and text quantity of the two self-built corpora are summarized as shown in Table 1.

Table 1: Basic Statistics of the Corpora

Corpus Name	Text Source	Number of Texts	Number of Words
L2-Human Corpus	IELTS Writing Task 2 sample essays (Band 7 level)	24	6814
AI-Generated Corpus	Generated by ChatGPT in independent sessions	24	5986

2.2 Data Processing and Annotation

Before formal analysis, the raw corpus was first organized and cleaned. Specifically, the 24 AI-generated texts were grouped into one dataset and the 24 human-written texts into another, and each group was merged into a single plain-text file (named AI_all.txt and Human_all.txt respectively), thereby forming two corpora of comparable size. During the merging process, all texts were encoded in UTF-8 and saved in plain-text format (.txt) to ensure compatibility with subsequent processing procedures. Basic formatting standardization was also carried out, including the removal of redundant blank lines, unification of paragraph spacing, and the use of line breaks to separate individual texts, so as to avoid potential interference caused by unclear text boundaries in subsequent statistical analyses. No linguistic modifications were made to the content at this stage, thereby preserving the original form and natural characteristics of the corpus as far as possible.

After corpus integration, TagAnt was used to conduct automatic part-of-speech (POS) tagging for both corpora. Specifically, *AI_all.txt* and *Human_all.txt* were imported into TagAnt, and its built-in English POS tagging model was applied. During tagging, each word in the corpus was assigned a corresponding POS tag, and the output was formatted as "word_POS" (e.g., *education_NN*, *is_VBZ*, *important_JJ*). Upon completion, the tagged results were saved as *AI_all_tagged.txt* and *Human_all_tagged.txt* respectively for subsequent lexical feature extraction and statistical analysis. Through this process, the raw corpus was transformed into a structured and annotated format, providing standardized input data for quantitative analysis using AntConc.

2.3 Quantitative Operation of Indicators

2.3.1 Sophisticated vocabulary

This study uses the Word List tool in AntConc to conduct statistical analysis. A minimum word length threshold of six characters is set to automatically extract words that meet this criterion and record their total frequency of occurrence. The proportion of long words is then calculated by dividing the frequency of these words by the total number of words in the corpus. This procedure relies on character-length-based automatic filtering, which helps reduce potential subjectivity associated with manual judgment.

2.3.2 Abstract Nouns

The measurement of abstract noun proportion in this study is based on a morphological criterion. Words containing typical nominalization suffixes (e.g., -tion, -ment, -ity, and -ness) and simultaneously tagged as nouns are operationally defined as abstract nouns. In practice, AntConc is used to retrieve words that satisfy both morphological and part-of-speech conditions. The frequencies of these items are then calculated and aggregated to obtain the total number of abstract nouns. Subsequently, the proportion of abstract nouns is computed by dividing the total number of abstract nouns by the total number of words in both the AI corpus and the human corpus, respectively. This percentage is used as the index of abstract noun usage in each corpus. Although this operational definition cannot capture all forms of abstract semantic expression, it ensures methodological consistency and enhances the comparability between the two corpora.

2.3.3 Lexical Density

The tagged corpora (AI_all_tagged.txt and Human_all_tagged.txt) were imported into AntConc, and content words were retrieved in the Concordance interface by searching for specific part-of-speech tags. Specifically, noun frequencies were obtained by searching "_NN", "_NNP", and "_NNS"; verb frequencies by searching "_VB", "_VBD", "_VBG", "_VBN", "_VBP", and "_VBZ"; adjective frequencies by searching "_JJ", "_JJR", and "_JJS"; and adverb frequencies by searching "_RB", "_RBR", and "_RBS". The frequencies of each category were recorded separately and then summed to obtain the total number of content words. Finally, lexical density was calculated by dividing the total number of content words by the total number of words in each corpus, yielding the lexical density index. The operational standards of the three lexical analytical dimensions are clearly defined as shown in Table 2.

Table 2: Operational Definitions of Lexical Measures

Analytical Dimensions	Key Indicators	Operational Definitions
Lexical Sophistication	Sophisticated vocabulary	≥6-character words
Semantic Abstraction	Abstract Nouns	Nominalized nouns ending in -tion, -ment, -ity, and -ness
Information Density	Lexical Density	Proportion of nouns, verbs, adjectives, and adverbs in total words

Through quantitative analysis and comparison across the above three dimensions, this study aims to reveal the lexical distributional characteristics of AI-generated texts, specifically examining the lexical strategies through which such texts construct linguistic features that appear to exceed the conventional expressive level of second language learners.

3. Results and Discussion

3.1 Results of Lexical Analysis

The results of the data analysis are as follows. In terms of advanced lexical items, the AI corpus contains 1,095 words with a character length of no fewer than six, accounting for approximately 18.3% of the total word count, whereas the human corpus contains 537 such words, representing approximately 7.9%. Regarding the proportion of abstract nouns, the AI corpus includes a total of 547 nouns ending in -tion, -ment, -ity, and -ness, accounting for 9.14% of the total words. In

comparison, the human corpus contains 226 such items, accounting for 3.32% of the total word count. The frequency of abstract noun usage in AI-generated texts is nearly three times that of human-written texts. In terms of lexical density, the AI corpus contains 1,944 nouns, 989 verbs, 911 adjectives, and 341 adverbs. These four categories of content words sum to 4,184 occurrences, accounting for approximately 69.9% of the total word count. In the human corpus, nouns occur 1,731 times, verbs 1,182 times, adjectives 746 times, and adverbs 476 times, yielding a total of 4,110 content-word tokens, accounting for approximately 60.3% of the total word count. The quantitative gaps between AI-generated and L2 human writing on three core lexical indicators are presented as shown in Table 3.

Table 3: Comparative Quantitative Results of Key Indicators in AI and Human Texts

Indicators	AI-Generated Corpus	L2-Human Corpus
Sophisticated vocabulary (Frequency / Percentage)	1,095 / 18.3%	537 / 7.9%
Abstract Nouns (Frequency / Percentage)	547 / 9.14%	226 / 3.32%
Lexical Density (Frequency / Percentage)	4184 / 69.9%	4110 / 60.3%

3.2 Discussion of Lexical Patterns

In sophisticated vocabulary, the results indicate that the proportion of advanced vocabulary in the AI corpus is 18.3%, significantly higher than the 7.9% observed in the human corpus. This finding is consistent with Jiang and Li (2026), who reported that generative artificial intelligence demonstrates strong statistical capability in retrieving low-frequency academic vocabulary [2,9]. This pattern can be attributed to the fact that large language models are pretrained on extensive domain-relevant corpora, within which a large number of low-frequency formal lexical items are encoded in the parameter space, enabling the generation of academic expressions that exceed the lexical range typically available to language learners [1]. Further analysis of the distribution of advanced lexical items shows that long words in AI-generated texts appear not only in content words (e.g., development, globalization, significantly), but also frequently in functional and transitional expressions (e.g., furthermore, nevertheless, consequently). This distributional pattern suggests that the model has learned discourse cohesion strategies at a probabilistic level, preferentially selecting longer and more formally marked cohesive devices rather than the shorter connectives commonly used by L2 learners (e.g., but, so, also). The systematic differences in the distribution of advanced lexical items confirm that AI-generated texts exhibit a form-based advantage in lexical sophistication. As a result, AI texts appear closer to high-level academic writing at the surface level and may even exceed the typical performance range of second language learners in certain respects [16, 17].

Regarding the distribution of abstract nouns, the proportion of words ending in typical nominalization suffixes such as -tion, -ment, -ity, and -ness in the AI corpus is 9.14%, which is substantially higher than the 3.32% observed in the human corpus. This difference is not only statistically evident but also produces noticeable stylistic effects at the discourse level. The high frequency of abstract nouns makes AI-generated texts more conceptual and generalized in nature, with argumentative development often shifting away from concrete situational grounding toward abstract-level reasoning. For example, lexical items such as development, globalization, improvement, and accessibility occur with high frequency [2]. From the perspective of Systemic Functional Linguistics, this strong tendency toward nominalization reflects a tendency observed in

AI-generated discourse [16]. As a primary form of grammatical metaphor, nominalization transforms dynamic processes and complex propositions into static entities, thereby enhancing the technicality and objectivity of discourse. During pretraining, large language models acquire this feature from large-scale academic corpora and subsequently reproduce it through probabilistic generation mechanisms. This finding is consistent with Zhang et al. (2026), who note that AI-generated texts tend to employ more complex noun phrases and favor conceptually dense lexical choices in constructing academic discourse, thereby compressing substantial informational content into nominal structures and enhancing formal register features [14].

In the case of lexical density, the proportion of content words in the AI corpus is 69.9%, which is significantly higher than the 60.3% observed in the human corpus. This finding partially aligns with the trend reported by Nkhobo and Chaka (2023), who used the Coh-Metrix lexical density index to examine learner writing. Although their study found slightly higher raw lexical density scores in student essays, statistical tests did not reveal a significant difference, suggesting that the lexical density of AI-generated texts is already comparable to that of human writing [18]. Regarding internal word-class distribution, the AI corpus shows a clear advantage in nouns and adjectives, with 1,944 noun tokens and 911 adjective tokens. In contrast, the human corpus demonstrates relatively higher proportions of verbs (1,182 occurrences) and adverbs (476 occurrences). This difference reflects a broader shift in academic discourse from clause-dominated structures toward phrase-dominated structures [14]. In addition, the stability of lexical diversity measures such as MTLD and HD-D varies across different text types when evaluating lexical variation [12]. In this study, the use of lexical density further confirms that AI-generated texts already exceed second language learner writing in terms of informational load. At the same time, the relationship between lexical richness and text quality appears to be highly context-dependent [11]. Zhu et al. (2024), based on Chinese corpora, similarly found a clear difference in lexical density between human responses and ChatGPT-generated texts, with human language exhibiting stronger narrative and interactional features [3]. This cross-linguistic consistency suggests that the relatively high lexical density of AI-generated texts may represent a more generalizable characteristic across languages rather than a phenomenon specific to English argumentative writing.

A higher lexical density indicates that a greater amount of information is encoded per unit of text, while the proportion of function words (e.g., prepositions, articles, and conjunctions) decreases. As a result, the surface form of the text appears more compact and information-dense. This pattern may reflect a tendency of language models may tend to prioritize words with higher semantic load during probabilistic prediction, resulting in higher informational load. However, high informational density does not necessarily correspond to higher discourse quality. The reduction of function words may weaken inter-sentential cohesion and overall textual fluency. In such cases, logical progression is often achieved through the accumulation of content words rather than through natural transitions and cohesive devices grounded in contextual reference. This helps explain why some AI-generated texts may appear “informationally rich yet stylistically rigid during reading. In addition, the dominance of nouns and adjectives in the AI corpus reflects a broader shift from clause-based structures toward phrase-based constructions. In contrast, the human corpus shows a relatively higher proportion of verbs and adverbs, indicating a stronger narrative orientation and a more dynamic mode of expression.

4. Conclusion

This study compares ChatGPT-generated texts with argumentative essays produced by second language learners and confirms significant quantitative differences between the two corpora at the lexical level. The findings indicate that AI-generated texts exhibit a high degree of consistency and

a clear statistical advantage across key indices, including the proportion of advanced lexical items, the proportion of abstract nouns, and lexical density. These differences reflect the probabilistic generation mechanism underlying generative artificial intelligence, which tends to construct highly information-dense and abstract discourse by preferentially selecting semantically rich lexical items. By contrast, human learners' writing is more narrative-oriented and constrained by cognitive processing capacity, resulting in a stronger reliance on concrete, context-dependent lexical choices. The systematic divergence in these linguistic indicators not only provides empirical evidence for AI text detection but also offers a reference for delineating the boundary between machine-generated writing and naturally acquired language production. These findings contribute to a more precise understanding of the distinction between statistical linguistic complexity and cognitively grounded language proficiency.

From a pedagogical perspective, artificial intelligence can be utilized as an auxiliary learning resource. During the revision stage, learners may draw on AI-generated suggestions to learn how to employ complex noun phrases and cohesive devices to enhance argumentative cohesion. This process requires integrating fragmentary high-level expressions provided by AI into the learner's own coherent logical structure. At the same time, assessment systems require corresponding adjustment. Traditional quantitative measures such as lexical complexity or average sentence length are no longer sufficient to comprehensively evaluate writing quality. Greater emphasis should instead be placed on critical thinking, originality of ideas, and the richness of narrative and affective expression.

This study also has several limitations. First, the size of the corpus is relatively small. Future research should expand the sample size and include a wider range of genres and topics. Second, the scope of analysis is mainly focused on the lexical level, while comparisons involving syntactic complexity and discourse cohesion remain to be further explored in future work. Third, with the continuous evolution of large language models, newer versions may exhibit different lexical distribution patterns, making longitudinal and tracking studies necessary. Future research may also consider incorporating eye-tracking experiments or cognitive load measurement techniques to further examine differences in reading experience between human-written and machine-generated texts from a processing perspective. In addition, the lexical features identified in this study could be integrated into AI detection systems to improve both detection accuracy and interpretability.

References

[1] X. Zhai (2023). ChatGPT for next generation science learning. XRDS: Crossroads, The ACM Magazine for Students, vol. 29, no. 3, p. 42-46.

[2] F. Jiang and X.L. Li (2026). A study on lexical richness and syntactic complexity of AI-generated texts: A comparison between ChatGPT and native college students' writing. Foreign Languages, vol.49, no. 2, p. 43-51.

[3] J.H. Zhu, M.Y. Wang, E.H. Yang, J.R. Nie, L.E. Yang and Y.J. Wang (2024). A comparative study of linguistic features between large language model-generated responses and human responses. Journal of Chinese Information Processing, vol. 38, no. 4, p. 17-27.

[4] E. Kasneci, K. Sessler, S. Küchemann, M. Bannert, D. Dementieva, F. Fischer et al. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, vol. 103, p. 102274.

[5] Z.H. Wang (2024). Research on identification of ChatGPT-generated texts based on semantic features. Master's thesis, Yanbian University.

[6] D. R. E. Cotton, P. A. Cotton and J. R. Shipway (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, vol. 61, no. 2, p. 228-239.

[7] C.A. Gao, F.M. Howard, N.S. Markov, E.C. Dyer, S. Wheeler, A. Leiter and A.T. Pearson (2023). Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence detector. bioRxiv.

[8] A. Shah, P. Ranka, U. Dedhia, S. Prasad, S. Muni and K. Bhowmick (2023). Detecting and unmasking AI-generated texts through explainable artificial intelligence using stylistic features. International Journal of Advanced Computer Science and Applications, vol. 14, no. 10.

[9] S. Herbold, A. Hautli-Janisz, U. Heuer, Z. Kikteva and A. Trautsch (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. Scientific Reports, vol. 13, no. 1, p. 18699.
[10] A. Housen, F. Kuiken and I. Vedder (2012). Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA. John Benjamins.
[11] X. Lu (2010). The relationship of lexical richness to the quality of ESL learners' oral narratives. The Modern Language Journal, vol. 94, no. 3, p. 474-496.
[12] P.M. McCarthy and S. Jarvis (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. Behavior Research Methods, vol. 42, no. 2, p. 381-392.
[13] T. McEnery and A. Hardie (2011). Corpus linguistics: Method, theory and practice. Cambridge University Press.
[14] D. Biber and B. Gray (2016). Grammatical complexity in academic English: Linguistic change in writing. Cambridge University Press.
[15] M.A.K. Halliday (1985). An introduction to functional grammar. Edward Arnold.
[16] C. Zhang, X.Y. Ma and Q.Y. Yan (2026). A comparative study on linguistic complexity between GenAI-generated texts and students' English argumentative essays. Shandong Foreign Language Teaching, vol. 47, no. 1, p. 77-86.
[17] X. Zhao (2025). A comparative analysis of AI-generated texts and PEP English textbook texts from the perspective of language and culture: A case study of senior high school reading materials. Master's thesis, Beijing Foreign Studies University.
[18] T. Nkhobo and C. Chaka (2023). Student-written versus ChatGPT-generated discursive essays: A comparative Coh-Metrix analysis of lexical diversity, syntactic complexity, and referential cohesion. International Journal of Education and Development using Information and Communication Technology, vol. 19, no. 3, p. 69-84.