3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies
Vol 32(3), September 2026 http://doi.org/10.17576/3L-2026-3203-14

A Comparative Analysis of the Lexical Diversity of the Introduction Section in AI-Generated vs Human-Authored Linguistic Articles

ASIA A. ALHEETY *
Department of English, College of Education for Humanities,
University of Anbar, Ramadi, Iraq
asia90alheety@gmail.com

MEETHAQ KHAMEES KHALAF
Department of English, College of Education for Humanities,
University of Anbar, Ramadi, Iraq

HUSSAM J. MOHAMMED
Department of Artificial Intelligence, Computer Sciences and Information Technology,
University of Anbar, Ramadi, Iraq

ALA'A ISMAEL CHALLOB
Department of English, College of Education for Humanities,
University of Anbar, Ramadi, Iraq

ABSTRACT

Lexical diversity is a core indicator of academic writing quality. Despite the rapid growth of AI-assisted academic writing, limited work has compared lexical diversity in advanced linguistic articles by examining AI-generated and human-written introductions through matched quantitative indices. This gap is important because lexical diversity can help educators, researchers, and journal stakeholders understand how AI-generated academic prose differs from human-written prose in vocabulary use and information density. Accordingly, this study aimed to compare the lexical diversity and lexical density of five paired introduction sections written by ChatGPT-4 and human authors, to determine whether GPT-4 matched or exceeded human writing in lexical behaviour, to examine whether RTTR and the Maas index remained consistent across three preprocessing conditions, and to evaluate whether lexical density functioned as a clear discriminator between the two text types. For the five paired introductions used in this study, ChatGPT-generated texts tended to present greater lexical density and overall greater lexical diversity than human-written texts. However, lexical diversity results varied across RTTR and Maas index values, suggesting that authorship differences should be interpreted through multiple measures rather than a single index.

Keywords: academic discourse; comparative linguistics; large language models; lexical richness; machine-generated text

INTRODUCTION

The meteoric rise of generative artificial intelligence (AI), especially general long-form models such as ChatGPT, has fundamentally altered how text is produced in academia, work, and life (Wu et al., 2023). GPT-4 is one of the first models that, when trained or employed to generate text, will result in a high degree of closeness between what is written by an AI and what is written by a human writer (Aldosari & Altuwairesh, 2026). This unprecedented capability has ignited a fervent debate across various sectors, particularly within academic institutions and scholarly publishing, regarding the authenticity, originality, and ethical implications of AI-generated content. The ease with which sophisticated, coherent, and contextually relevant text can now be produced by

machines challenges traditional notions of authorship and intellectual contribution. With AI-generated writing content becoming more prevalent in academic writing and scholarly communication, worries have arisen about the ways it might shape language, change linguistic norms, and evolve as a genre (Dwivedi et al., 2023; Ray, 2023).

Nevertheless, the potential impact of generative AI cannot be evaluated from a linguistic standpoint without a systematic comparison of human-written and generative texts. Such an evaluation necessitates a deep dive into the intrinsic characteristics of language produced by both human intellect and algorithmic processes. Moreover, it has also been hypothesised that the increasing availability of AI-generated texts is used to train future models and drive recursive feedback loops that may affect linguistic diversity and model performance (Shumailov et al., 2023). Understanding these dynamics is crucial for mitigating unintended consequences and guiding the responsible development of AI technologies. One key conceptual dimension for this purpose, a pertinent variable when comparing human- and AI-generated texts, is lexical richness, or the degree of diversity of vocabulary use in a text or corpus (Schmitt, 2010). Lexical diversity (LD) has been a well-established measure of language proficiency, stylistic richness, and communicative effectiveness, and studies have also shown that it is sensitive to genre, task, and writer background variables (Biber et al., 1998). This sensitivity makes LD an ideal metric for discerning subtle yet significant differences between human and machine-generated prose. This and other findings related to the narrow vocabulary in AI-generated texts could be relevant not just for the study of linguistics but also for language teaching and academic writing. Only recently have researchers begun to examine the language differences, particularly in syntax, cohesion, and discourse features, that distinguish human and AI writing (Guo et al., 2023; Ray, 2023). Consequently, investigations into spatial and temporal perspectives on lexical diversity through comparative linguistic approaches remain rare for higher-complexity and advanced models such as GPT-4. Research has consistently failed to generalise either from small datasets (Herbold et al., 2023) or from low-level model versions (Zhai, 2022). These trends highlight the need for linguistically based cross-model comparisons to find out if the progress in modern AI language simply corresponds to an increase in vocabulary and richness. In this regard, the current study employs a comparative linguistic approach in the analysis of lexical diversity between human and AI writing through the medium of GPT-4. However, in contrast, this study has four aims through the comparison of ahistorical quantitative measures of lexical diversity in similar corpora. In this study, we are interested in investigating the lexical richness of academic introductions by contrasting linguistic articles written by humans against texts produced by ChatGPT-4. Secondly, it also aims to evaluate whether GPT-4 equalled or outperformed humans at the linguistic level. In particular, it examines its lexical behaviour reflecting specialised academic writing. Third, the study will investigate whether the reported values of indexes of lexical diversity, the RTTR and the Maas index, specifically, are consistent across the three clearly delineated states of text pre-processing, in particular when comparing raw texts with texts subjected to stop-word removal, as well as a content-word filter. Lastly, the present study seeks to see how salient lexical density is as a primary differentiator of AI-generated and human-written academic text. While AI text generation is rapidly evolving, the human experience in writing and communication continues to be essential and will not easily be replaced. Though AI can simulate language, craft fluid narratives and cater to various stylistic requests, it ultimately lacks the human experience, emotion, and the deeper understanding of the human condition to ever create writing that carries the same weight. AI can never possess the depth of personal experience, cultural context, ethics, and overall subjective interpretation that human writers draw upon from an almost never-ending source. This

unique human ability gifts us with empathy and critical thought and empowers our creative intuition to create texts that move us on the deepest and general levels. Statistical probabilities derived from massive datasets may be very efficient, but the irony, metaphor, and insight that remind us of our humanity more often come from the human mind trying to understand good and evil among the complex realities of our existence. So while we may be able to develop AI technologies that inspire human- and machine-generated text that are more and more difficult to objectively distinguish based purely on lexical diversity and other quantifiable linguistic metrics, it is possible that the ultimate Deciding Factor may be the mere presence, or lack thereof, of human spirit. Future research and practice must strive not only to detect AI writing but to capture the uniqueness that points to human writing and to ensure that uniquely human expression continues to be present as the world becomes further automated. Therefore, the present study addresses the following research questions: (1) How do ChatGPT-4-generated and human-written linguistic article introductions differ in lexical diversity under the all-words, no-stop-words, and content-word conditions? (2) Do RTTR and the Maas index produce consistent interpretations across these preprocessing conditions? (3) Does lexical density serve as a clear differentiator between AI-generated and human-written introductions? The study is significant because it links AI-assisted academic writing to measurable lexical patterns that can inform academic writing instruction, assessment, and responsible AI use (Hyland, 2004; Kasneci et al., 2023; McNamara et al., 2014).

LITERATURE REVIEW

The discovery of large language models has created a new and different dimension to research on language use and textual production. Since ChatGPT is able to produce long, continuous texts in various styles, there is an increasing curiosity to investigate the textual properties of models in comparison to those of humans (Alheety et al., 2026; Iskender, 2023; Khalil & Er, 2023). Although large language models have been previously evaluated in numerous ways, such as accuracy, task performance, or factual knowledge, lexical evaluation regarding vocabulary richness and variation has been explored to a much lesser extent. Lexical richness, which is sometimes referred to as lexical diversity, is the range and frequency of vocabulary items in a text (Nation, 2013; Schmitt, 2010). It has been viewed in the past as an important measure of language and writing quality, especially in the context of academic and argumentative writing. Studies in linguistics have revealed that lexical richness is affected by the genre of the text, the degree of familiarity with the topic, educational background, and communicative purpose (Biber et al., 1998; Laufer & Nation, 1995). Thus, a lexical richness comparison between text producers, in particular, between humans and AI systems, provides valuable insights into how closely the language produced by AI resembles human language behaviour. Lexical richness is a concept that has been operationalised using various quantitative measures. Type-Token Ratio, for example, was an extremely popular early metric (Hout & Vermeer, 2007) that was quickly critiqued for being sensitive to text length. The Root Type-Token Ratio (RTTR) and the Maas index were put forward to overcome this limitation, providing more stable measures of lexical diversity across texts of different lengths (Tweedie & Baayen, 1998). Since then, such measures have been implemented more broadly in linguistic and applied linguistics research to compare vocabulary use between different speakers, writers, and text types. Nevertheless, when it comes to AI-generated text, initial investigations have started to examine linguistic disparities between human and machine language. Studies that compare AI and human writing by examining syntactic patterns, discourse structure, and stylistic

features have found that, while AI texts often are quite fluent, they also may differ systematically from human texts (Guo et al., 2023). In particular, some studies claim that earlier iterations of ChatGPT have a more repetitive lexicon, leading to lower lexical diversity when compared to human-written text (Hamat, 2024; Herbold et al., 2023; Zhai, 2022). Nevertheless, these results currently refer to rather small data sets or older versions of the models and hence are less relevant for contemporary systems such as GPT-4. Here, small datasets refer to studies using limited text samples or narrowly defined corpora that restrict generalizability, whereas older model versions refer to pre-GPT-4 or early ChatGPT systems whose lexical outputs may not reflect the performance of newer models (Herbold et al., 2023; OpenAI, 2023; Zhai, 2022). Recent studies have increasingly shown that AI-generated writing can approximate human academic prose at the levels of fluency, structure, and coherence, but measurable linguistic differences may remain in repetition, lexical choice, and discourse organisation (Guo et al., 2023; Herbold et al., 2023; Kasneci et al., 2023; Khalil & Er, 2023; Nor & Aziz, 2026). These findings justify extending comparison beyond general detectability to specific lexical measures such as RTTR, the Maas index, and lexical density. Still, exciting new work suggests that model architecture and training data have come far enough that we may soon get improvements in lexical richness, as newer models tend to produce distributions of vocabulary usage that are much closer to those of human writers (Ray, 2023). However, the degree to which marginal reductions in lexical variation may be achieved is unknown. Further, the majority of existing studies cover either more general text generation or rephrasing tasks, and empirical investigations comparing the lexical richness of larger bodies of text (e.g., articles) are largely absent. The problem addressed in this study is that current discussions of AI-generated academic writing often emphasise detection or general quality, while less attention is given to how lexical diversity operates in comparable introductions from the same academic field. Without this comparison, it remains unclear whether AI-generated introductions merely imitate academic style or whether they display measurable differences in vocabulary range and information density (Hyland, 2004; Kasneci et al., 2023; Schmitt, 2010).

METHODOLOGY

RESEARCH DESIGN

A controlled parallel corpus of ten introduction sections (five ChatGPT-4-generated and five peer-reviewed articles) was created to compare lexical diversity in these two modes of writing. This study adopted a controlled matched-pairs design to prioritise internal validity over broad generalisation. By using a one-to-one thematic mapping, the research effectively isolated authorship (human vs AI) as the primary independent variable. This approach neutralised potential biases from topical variations, ensuring that any observed differences in lexical richness were directly linked to the source of generation (Biber et al., 1998). Such a focused methodology was well-suited for the intricately detailed linguistic analysis required in this study (McEnery & Hardie, 2012). Methodologically, the study followed a quantitative, corpus-based comparative approach. It was quantitative because the analysis depended on measurable lexical indices, corpus-based because each introduction was treated as textual data, and comparative because the paired design allowed each ChatGPT-4 text to be evaluated against a thematically matched human-written introduction (Biber et al., 1998; McEnery & Hardie, 2012).

DATA SOURCES

The data set consists of 10 introduction sections of linguistic articles, including five human empirical research articles in linguistics selected from journals indexed by Scopus. Articles were selected based on predetermined criteria for comparability across the corpus: (1) disciplinary homogeneity (linguistics), (2) topical proximity, (3) empirical nature, and (4) length; additionally, they included an Introduction, Methods, Results, and Discussion (IMRAD) structure. To ensure temporal relevance, only literature published between 2020 and 2024 was included to ensure that the articles were written before the emergence of AI. These five human-written introduction sections were analysed alongside their five ChatGPT-generated counterparts, creating a balanced comparative corpus. An AI-generated article on the same topic was created for each human-written article using structured prompts detailing academic tone, structure of research, and a word count; therefore, GPT-4 was chosen for its aforementioned reported improvements in linguistic sophistication, coherence, and contextual reasoning over earlier large language models (Bubeck et al., 2023). GPT-4 is the latest publicly released large language model. As a result, it provided a more accurate snapshot in time of the current state of AI-assisted academic writing. Although structured prompts included a target word count, the ChatGPT-generated introductions were not forced to match the human texts exactly because exact length control could have made the texts artificial. Instead, length variation was retained and later addressed analytically through RTTR and the Maas index, which are less dependent on raw text length than the basic TTR (Jarvis & Daller, 2013; Tweedie & Baayen, 1998). Despite the small sample of articles, however, this study employed a controlled matched-pair comparative design that trades large-scale generalization for internal validity and structural comparability, though smaller and more controlled datasets can be adequately treated as long as the goal is not population-near inferences but rather a more intricately detailed analysis of linguistic features (Biber et al., 1998; McEnery & Hardie, 2012) in research in the field of corpus linguistics.

RESEARCH INSTRUMENTS

Two measures were employed to determine lexical diversity: the Root Type-Token Ratio (RTTR) and the Maas index. Both measures are established lexical richness indices and were selected because they reduce the direct effect of text length more effectively than the basic Type-Token Ratio (Jarvis & Daller, 2013; Tweedie & Baayen, 1998).

$$\text{RTTR} = \frac{t}{\sqrt{n}}$$ RTTR was calculated as RTTR = t / √n, where t represents the number of unique words (types) and n represents the total number of words (tokens). Higher RTTR values indicate greater lexical diversity (Jarvis & Daller, 2013; Tweedie & Baayen, 1998).

$$\text{Maas} = \frac{\log(n)-\log(t)}{\log n^2}$$ The Maas index was calculated as $a^2 = (\log n - \log t) / (\log n)^2$, where t represents types and n represents tokens. Lower Maas values indicate greater lexical diversity. Both metrics were calculated for three preprocessing conditions: all words, words after stop-word removal, and content words only. The datasets were preprocessed using Python (Jarvis & Daller, 2013; Tweedie & Baayen, 1998).

LEXICAL DIVERSITY ANALYSIS

All textual data were preprocessed using the Python Natural Language Toolkit (NLTK). Preprocessing steps included tokenisation, Part-of-Speech (POS) tagging using Penn Treebank and Universal POS tag sets, lemmatisation to reduce morphological variability, removal of punctuation and stop-words, and selection of content words (nouns, verbs, adjectives, and adverbs) for measuring lexical richness (McEnery & Hardie, 2012; Nation, 2013). In the following section, we expose differences in lexical diversity within the introduction sections of Scopus-indexed peer-reviewed articles and ChatGPT-4-generated texts. Based on the Root-Type-Token Ratio (RTTR) and the Maas index, lexical diversity was assessed in terms of total words, total words without stop words, and content words only. Overall, ChatGPT-4 had higher lexical diversity than all other comparison texts across the three conditions and outperformed in two of the five pairs (depending on the condition) against human-written introductions (see Table 1). One pair showed overall greater diversity for human texts in most comparisons, while the rest were mixed due to divergence between the RTTR and the Maas indices. In aggregate, these results suggest that the lexical diversity of ChatGPT-4 introductions is comparable to and often exceeds the lexical diversity of Scopus-indexed human introductions. However, pairwise variation indicates that lexical richness is responsive to context and discipline, not just the type of authorship.

RESULTS

Before presenting the full token counts, it is important to note that the five paired introductions differed in length across the three preprocessing stages. Table 1 summarises these token distributions, which provide the basis for interpreting the lexical diversity indices reported in the following tables.

TABLE 1. Summary of the data used in the evaluation

Category	Type	Elements	All Words	After Stop-Word Removal	Parts-of-Speech Content Words
Pair 1	Human	1	944	577	550
	ChatGPT	1	1100	735	712
Pair 2	Human	1	941	577	560
	ChatGPT	1	379	258	258
Pair 3	Human	1	1001	573	557
	ChatGPT	1	740	505	494
Pair 4	Human	1	555	478	461
	ChatGPT	1	474	307	301
Pair 5	Human	1	976	558	538
	ChatGPT	1	606	413	403

Table 1 shows the number of tokens in each paired introduction in consecutive preprocessing stages, thus presenting the contextual length required in the interpretation of the Root-Type-Token Ratio (RTTR) and the Maas index. For Pair 1, the human-written introduction had 944 tokens, and the ChatGPT-generated introduction had 1,100 tokens for the all-words condition. After removing the stop words, the human-written introduction had 577 tokens, and the ChatGPT-generated introduction had 735 tokens. When it came to Part-of-Speech (POS)-based content words, the human-written introduction had 550 tokens, and the ChatGPT-generated introduction had 712 tokens. For Pair 2, when using the all-words condition, the human-written

introduction had 941 tokens, and the ChatGPT-generated introduction had 379 tokens; upon removal of the stop words, the human-written introduction had 577 tokens, and the ChatGPT-generated introduction had 258 tokens; moreover, when using POS words, the human-written introduction had 560 tokens, and the ChatGPT-generated introduction had 258 tokens.

When it came to the all-words condition in Pair 3, the human-written introduction had 1,001 tokens, and the ChatGPT introduction had 740 tokens. However, after the removal of the stop words, the human-written introduction had 573 tokens, while the ChatGPT introduction had 505 tokens. Moreover, when it came to content words based on POS, human-written introductions had 557 tokens, while ChatGPT-generated introductions had 494 tokens, respectively.

Concerning Pair 4, in the all-words condition, the human-written introduction had 555 tokens, and the ChatGPT-generated introduction had 474 tokens. After the removal of the stop words, the human-written introduction had 478 tokens, while the ChatGPT introduction had 307tokens. Furthermore, the human-written introduction had 461 tokens, and the ChatGPT-generated introduction had 301 tokens in the content words of POS. Finally, regarding Pair 5, with the all-words condition, the human-written introduction had 976 tokens, and the ChatGPT-generated introduction had 606 tokens. After stop-word removal, the human-written introduction had 558 tokens, and the ChatGPT-generated introduction had 413 tokens. Lastly, the POS-based content words in the human-written introduction had 538 tokens, and the ChatGPT-generated introduction had 403 tokens. These distributions prove that the length of text differs among pairs and is just one of the many factors that must be considered when comparing lexical diversity indices.

The first lexical diversity comparison was conducted on the complete texts before any stop-word removal. Table 2 therefore reports the all-word condition and shows how each pair differs in number of words, distinct words, RTTR, and Maas index values.

TABLE 2. Number of words, number of distinct words, RTTR, and Maas index metrics for each category

Category	Type	Number of Words	Distinct Words	RTTR	Maas
Pair 1	Human	944	282	9.1783	0.0593
	ChatGPT	1100	396	11.9398	0.0480
Pair 2	Human	941	348	11.3445	0.0489
	ChatGPT	379	203	10.4274	0.0408
Pair 3	Human	1001	289	9.1344	0.0599
	ChatGPT	740	310	11.3958	0.0459
Pair 4	Human	555	404	17.1488	0.0183
	ChatGPT	474	213	9.7834	0.0485
Pair 5	Human	976	380	12.1635	0.0458
	ChatGPT	606	302	12.2679	0.0391

Table 2 above indicates the lexical diversity measures of Condition I: 'All Words'. The lexical diversity of every pair was measured using the indices of the Root-Type-Token Ratio (RTTR) and the Maas index; the higher the RTTR index, the higher the diversity, and the lower the Maas index, the higher the diversity.

In Pair 1, the RTTR for the human-written introduction was 9.1783, while the RTTR for the ChatGPT-generated introduction was 11.9398. Meanwhile, the Maas index for the ChatGPT-generated introduction was 0.0480, while the Maas index for the human-written introduction was 0.0593. As a result, ChatGPT had more lexical diversity in this pair. In Pair 2, the RTTR for the human-written introduction was 11.3445, while the RTTR for the ChatGPT-generated introduction was 10.4274. For the Maas index, the human-written introduction was 0.0489, and the Maas index for the ChatGPT-generated introduction was 0.0408. Here, the indices of both the RTTR and Maas

show an inconsistent pattern; as a result, no consistent benefit can be identified with this pair. In Pair 3, the RTTR for the human-written introduction was 9.1344, and the RTTR for the ChatGPT-generated introduction was 11.3958. Meanwhile, concerning the Maas index, the human-written introduction was 0.0599, and the ChatGPT-generated introduction was 0.0459. Thus, the ChatGPT-generated introduction demonstrated greater levels of lexical diversity in this pair.

Pair 4 demonstrates that the RTTR for the human-written introduction was 17.1488, and for the ChatGPT-generated introduction, the RTTR was 9.7834. At the same time, the Maas index for the human-written introduction was 0.0183, while for the ChatGPT-generated introduction, the Maas index was 0.0485. Consequently, for Pair 4, the RTTR-Human showed more lexical diversity in general between the pair.

Finally, when it came to Pair 5, the RTTR for the human-written introduction was 12.1635, and the RTTR for the ChatGPT-generated introduction was 12.2679. In regard to the Maas index for Pair 5, the human-written introduction was 0.0458, and for the ChatGPT-generated introduction, the Maas index was 0.0391. Therefore, the ChatGPT-generated introduction had a greater overall lexical diversity between the pair.

Taken together, Condition I: 'All Words' shows that the ChatGPT-generated introductions had greater lexical diversity in three of the five pairs, whereas the human-written introductions had greater lexical diversity in one pair, and one pair showed mixed evidence between the indices.

The second comparison removed stop-words in order to focus more directly on lexical items that carry greater semantic information. Table 3 presents the resulting word counts and lexical diversity values under this condition.

TABLE 3. Results after removing stop-words

Category	Type	Number of Words	Distinct Words	RTTR	Maas
Pair 1	Human	577	235	9.7832	0.0512
	ChatGPT	735	345	12.7255	0.0400
Pair 2	Human	577	299	12.4475	0.0374
	ChatGPT	265	175	10.7502	0.0307
Pair 3	Human	573	238	9.9426	0.0502
	ChatGPT	505	274	12.1928	0.0363
Pair 4	Human	461	378	17.6052	0.0122
	ChatGPT	301	185	10.6632	0.0344
Pair 5	Human	558	325	13.7583	0.0311
	ChatGPT	413	265	13.0398	0.0282

Above, Table 3 presents the lexical diversity for Condition II: 'No Stop Words'. The lexical diversity was measured between each pair using the Root-Type-Token Ratio (RTTR) and the Maas index. Again, the higher the RTTR index, the larger the diversity, and conversely, the lower the Maas index, the larger the diversity.

For Pair 1, the RTTR for the human-written introduction was 9.7832, and the RTTR for the ChatGPT-generated introduction was 12.7255. Regarding the Maas index, it was 0.0512 for the human-written introduction, and for the ChatGPT-generated introduction, the Maas index was 0.0400. This result means that the overall lexical diversity was higher in the ChatGPT-generated introduction than it was in the human-written introduction.

The results for Pair 2 showed that the RTTR score for the human-written introduction was 12.4475, while the score for the ChatGPT-generated introduction was 10.7502. Concerning the Maas index scores, the human-written introduction scored 0.0374, while the ChatGPT-generated introduction scored 0.0307. These scores reveal an ambivalent pattern across the indices, which excludes the existence of a single advantageous result for a single speaker in this pair.

For Pair 3, the RTTR score for the human-written introduction was 9.9426, and the RTTR score for the ChatGPT-generated introduction was 12.1928. For the Pair 3 Maas index, the human-written introduction was 0.0502, and the ChatGPT-generated introduction was 0.0363. This data shows that the ChatGPT-generated introduction had more overall lexical diversity than the human-written introduction in Pair 3.

Moving on to Pair 4, the RTTR score for the human-written introduction was 17.6052, and the RTTR score for the ChatGPT-generated introduction was 10.6632. Furthermore, the Maas index score for the human-written introduction was 0.0122, and for the ChatGPT-generated introduction, it was 0.0344, meaning that the human-written introduction exhibited greater overall lexical diversity than the ChatGPT-generated introduction in this pair.

For Pair 5, the RTTR value for the human-written introduction was 13.7583, and the RTTR score for the ChatGPT-generated introduction was 13.0398. Moreover, the score of the Maas index for the human-written introduction was 0.0311, while the Maas index score for the ChatGPT-generated introduction was 0.0282. These values create an ambiguous effect across indices, making it impossible to clearly favour one participant over the other in this pair.

Under Condition II: 'No Stop Words', the ChatGPT-generated introductions scored higher in two pairs, the human-written introduction scored higher in one pair, and the remaining two pairs showed mixed evidence across the RTTR and Maas indices.

The third comparison further restricted the data to content words identified through POS filtering. Table 4 reports the lexical diversity values after retaining nouns, verbs, adjectives, and adverbs only.

TABLE 4. Results after removing stop words and selecting only nouns, verbs, adverbs, and adjectives

Category	Type	Number of Words	Distinct Words	RTTR	Maas
Pair 1	Human	550	228	9.7220	0.0509
	ChatGPT	712	333	12.4797	0.0406
Pair 2	Human	560	289	12.2125	0.0380
	ChatGPT	258	170	10.5837	0.0312
Pair 3	Human	557	230	9.7454	0.0509
	ChatGPT	494	267	12.0129	0.0368
Pair 4	Human	442	361	17.1710	0.0126
	ChatGPT	295	180	10.4800	0.0352
Pair 5	Human	538	312	13.4513	0.0317
	ChatGPT	403	260	12.9515	0.0280

Table 4 presents the results for the lexical diversity of Condition III: 'Content- and Parts-of-Speech-Based Only'. As in the previous tables, lexical diversity was measured in each pair using both the Root-Type-Token Ratio (RTTR) and the Maas index. Once again, it is important to note that the higher the RTTR is, the more lexical diversity there is, and the lower the Maas index is, the higher the lexical diversity is. For Pair 1, the RTTR value for the human-written introduction was 9.7220, and the RTTR value for the ChatGPT-generated introduction was 12.4797. For the Maas index, the human-written introduction was 0.0509, while the ChatGPT-generated introduction was 0.0406. In accordance with the rule that the higher the RTTR is, the more lexical diversity there is, and the lower the Maas index is, the higher the lexical diversity is, the ChatGPT-generated introduction was revealed to have superior overall lexical diversity than the human-written introduction in this pair.

For Pair 2, the RTTR value for the human-written introduction was 12.2125, and the RTTR value for the ChatGPT-generated introduction was 10.5837. Concerning the Maas index, the value for the human-written introduction was 0.0380, and the value for the ChatGPT-generated introduction was 0.0312. Thus, for Pair 2, the hybrid trend in all indices makes it impossible to favour either the human-written introduction or the ChatGPT-generated introduction in this pair with one unidirectional benefit.

For Pair 3, the RTTR value for the human-written introduction was 9.7454, while the RTTR value for the ChatGPT-generated introduction was 12.0129. For the Maas index of Pair 3, the human-written introduction value was 0.0509, and the ChatGPT-generated introduction was 0.0368. Thus, in this pair, the ChatGPT-generated introduction showed a greater overall lexical diversity than the human-written introduction did.

For Pair 4, the RTTR value for the human-written introduction was 17.1710, and the RTTR value for the ChatGPT-generated introduction was 10.4800. For the Maas index of this pair, the human-written introduction value was .0126, while the Maas index value for the ChatGPT-generated introduction was .0352. Therefore, the results showed that the human-written introduction presented more lexical diversity than the ChatGPT-generated introduction in this pair.

Lastly, for Pair 5, the RTTR value for the human-written introduction was 13.4513, and the RTTR value for the ChatGPT-generated introduction was 12.9515. Concerning the Maas index values, the human-written introduction value was 0.0317, and the Maas index value for the ChatGPT-generated introduction was 0.0280. The conflicting positions of the indices lead to the absence of a single benefit to either of the members of this pair.

Across Condition III: 'Content- and Parts-of-Speech-Based Only', the ChatGPT-generated introductions showed greater lexical diversity in two pairs, the human-written introductions showed greater lexical diversity in one pair, and the remaining two pairs showed mixed evidence because the RTTR and Maas index did not produce a single consistent direction.

As a qualitative illustration of the POS-based filtering, examples of content-word categories include nouns such as corpus, introduction, vocabulary, diversity, model, and authorship; verbs such as compare, generate, examine, indicate, and evaluate; adjectives such as lexical, academic, human-written, AI-generated, comparable, and quantitative; and adverbs such as systematically, comparatively, and consistently. These examples clarify that the content-word condition focused on semantically informative words rather than all tokens equally (Nation, 2013; Schmitt, 2010).

Finally, lexical density was examined to determine the proportion of content words in relation to total words. Table 5 reports this ratio for each human-written and ChatGPT-generated introduction.

TABLE 5. Results of lexical density for each category

Category	Type	Content Words	Total Words	Lexical Density
Pair 1	Human	550	944	0.5826
	ChatGPT	712	1100	0.6473
Pair 2	Human	560	941	0.5951
	ChatGPT	258	379	0.6807
Pair 3	Human	557	1001	0.5564
	ChatGPT	494	740	0.6676
Pair 4	Human	442	555	0.7964
	ChatGPT	295	474	0.6224
Pair 5	Human	538	976	0.5512
	ChatGPT	403	606	0.6650

Above, Table 5 shows the lexical density of five pairs of introductions, with each pair including one human-written introduction and one ChatGPT-generated introduction. The table presents the ratio of part-of-speech-based content words in comparison to the overall number of tokens; high scores indicate a higher percentage of information-carrying lexical units. This metric is particularly crucial in academic discourse, where conciseness and the efficient conveyance of complex ideas are highly valued. A higher lexical density often correlates with a more information-rich text, suggesting that the author, or in this case, the AI, is using a greater proportion of substantive words rather than grammatical or functional words (Biber et al., 1998; McNamara et al., 2014).

In this analysis, substantive words refer to information-bearing content words, including nouns, main verbs, adjectives, and adverbs. Functional words refer to grammatical items such as articles, prepositions, conjunctions, auxiliary verbs, and pronouns, for example, the, a, of, in, and, to, is, are, and it. This distinction helps explain why lexical density is useful: it measures how much of a text is carried by semantically meaningful vocabulary rather than grammatical support words (Nation, 2013; Schmitt, 2010).

In Pair 1, the lexical density for the human-written introduction was 0.5826, and for the ChatGPT-generated introduction, it was 0.6473. Based on this data, the ChatGPT-generated introduction had a larger lexical density in comparison to the human-written introduction. This initial observation sets a trend, indicating that the AI model is capable of generating text that is quantitatively more dense with content-bearing vocabulary from the outset.

In Pair 2, the lexical density for the human-written introduction was 0.5951, while the lexical density for the ChatGPT-generated introduction was 0.6807. Once again, based on the data for Pair 2, the ChatGPT-generated introduction had a larger lexical density in comparison to the human-written introduction. The consistent pattern across these initial pairs suggests a systematic characteristic of ChatGPT's output: an inclination towards a higher concentration of meaningful terms, which could be interpreted as a more direct and less verbose style, at least in terms of lexical loading.

In Pair 3, the lexical density for the human-written introduction was 0.5564, and the lexical density for the ChatGPT-generated introduction was 0.6676. Yet again, the ChatGPT-generated introduction had a larger lexical density in comparison to the human-written introduction. This repeated finding reinforces the notion that ChatGPT, in these instances, is optimising for informational efficiency, potentially by selecting words that carry more semantic weight within the context of academic introductions.

In Pair 4, the lexical density for the human-written introduction was 0.7964, and the lexical density for the ChatGPT-generated introduction was 0.6224. Here, in Pair 4, it can be seen that the human-written introduction had a higher lexical density than the ChatGPT-generated introduction. This outlier is significant, as it demonstrates that human writing can, under certain circumstances, achieve a higher informational density than AI-generated text. It prompts further inquiry into the specific stylistic choices or content nuances present in this particular human-written introduction that led to its elevated lexical density, potentially reflecting a highly specialised or condensed form of expression.

Finally, in Pair 5, the lexical density for the human-written introduction was 0.5512, while the lexical density for the ChatGPT-generated introduction was 0.6650; once again, the ChatGPT-generated introduction had a larger lexical density in comparison to the human-written introduction. The return to the prevailing trend in Pair 5 solidifies the general observation that ChatGPT-generated introductions tend to exhibit higher lexical density.

Overall, lexical density was higher in the ChatGPT-generated introductions than in the human-written introductions in four of the five comparisons.

Figure 1, below, shows the Root Type-Token Ratio (RTTR) and the Maas index comparisons across the pairs under Condition I: 'All Words', Condition II: 'No Stop Words', and Condition III: 'Content- and Parts-of-Speech-Based Only'. The top row presents RTTR values, while the bottom row presents Maas index values. Overall, the figure reinforces the table-based results by showing that ChatGPT-generated introductions frequently achieved higher lexical diversity, although the pattern was not uniform across all pairs and conditions. This variation supports the need to interpret AI-generated and human-written academic texts through both quantitative indices and qualitative lexical examples (Jarvis & Daller, 2013; Tweedie & Baayen, 1998).

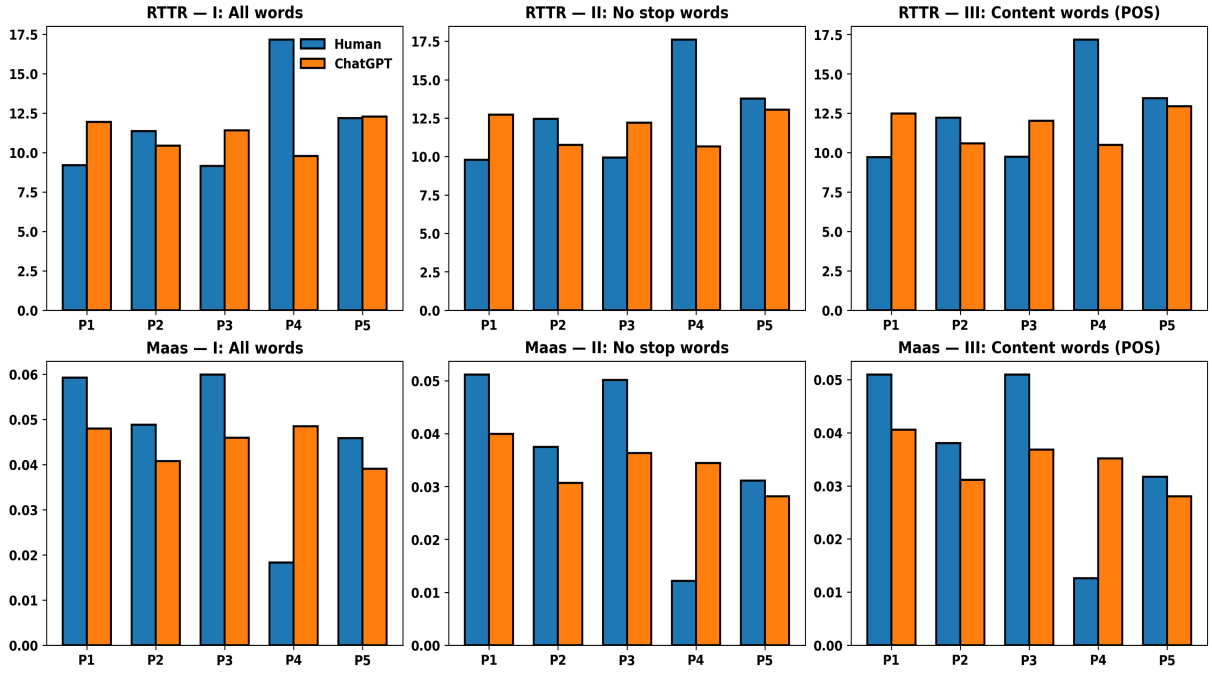

FIGURE 1. RTTR and Maas index comparisons across the pairs under Conditions I, II, & III

DISCUSSION

The lexical patterns indicate that not only does the number of lexical diversity differ, but also the distribution of lexical choices between AI-generated and human-written introductions. The ChatGPT-generated introductions were more likely to use academic and methodological vocabulary explicitly in the introduction, while the introductions written by humans exhibited greater variation in both topical and rhetorical words. This represents the difference between the texts created by ChatGPT and those produced by humans, and why certain human texts achieved the best results in some of the RTTR and Maas comparisons (Guo et al., 2023; Herbold et al., 2023; Ray, 2023).

The examples of content words also illustrate that the lexical diversity results are not a mere word count. The nouns that provide the conceptual focus of the texts (corpus, vocabulary, diversity, model, authorship); the verbs that represent analytical or evaluative action (compare, generate, examine, evaluate); the adjectives that represent disciplinary meaning (lexical, academic, comparable, quantitative); and the adverbs that add evaluative precision (systematically, consistently). The lexical categories contribute to the explanation of the different ways in which academic meaning is built in the two text types (Biber et al., 1998; Nation, 2013; Schmitt, 2010).

In contrast, functional words like articles, prepositions, conjunctions, auxiliaries, and pronouns are words that aid in the grammar of a sentence but are not directly related to the disciplinary content. The lexical density, on the other hand, is a higher value in most of the ChatGPT-generated introductions, suggesting that these introductions contain a higher density of information-carrying words, rather than being simply longer. This finding reinforces the notion that lexical density is a better discriminator than lexical diversity alone, as has been suggested by some (McNamara et al., 2014; Schmitt, 2010).

CONCLUSION

AI tools have been increasingly influencing academic writing, and lexical analysis provides a quantifiable way of investigating this change. This study does not consider AI-generated and human-written texts to be the same or completely different, but rather that they can be differentiated in terms of certain indices of lexical diversity and density. The results of the five paired introductions demonstrated that, in general, the introductions generated by ChatGPT exhibited better lexical performance, particularly in lexical density, compared to the original ones. When evaluating the lexical diversity of introductions generated by ChatGPT versus those written by humans under Condition I: ‘All Words', three pairs of the latter produced more lexical diversity than the former, and three pairs of the former showed more lexical similarity. In three pairs, under Condition I: 'All Words', ChatGPT-generated introductions demonstrated greater lexical diversity, while human-written introductions exhibited greater lexical diversity in one pair, and one showed mixed evidence. The same patterns emerged in two pairs of conditions (No Stop Words, Condition II; Content- and Parts-of-Speech-Based Only, Condition III), with ChatGPT-generated introductions being more lexically diverse than human-written ones in two cases, the human ones being more lexically diverse than ChatGPT in one case, and two cases being mixed due to a lack of full consensus between RTTR and the Maas index. The most apparent difference was lexical density, which was higher in four of the five pairs of introductions created by ChatGPT. The results indicate that ChatGPT-4 has the ability to produce academic introductions that contain a high percentage of information-bearing vocabulary and, in some instances, a high degree of lexical diversity. The study is, however, restricted by the number of matched texts (five pairs) and focused on introduction sections only. Larger corpora, other academic fields, complete sections of an article and qualitative analysis of rhetorical and semantic choice of words should all be considered in future research. In the context of writing pedagogy for academic writing, the result suggests the use of lexical diversity indices in addition to qualitative reading when assessing the writing aided by AI systems, as numbers alone do not accurately reflect human originality, academic subjectivity, and rhetorical control.

ETHICAL CONSIDERATIONS

All AI-generated texts were created for research purposes only. Human-written articles were used in compliance with copyright and Scopus database terms. The study did not involve sensitive or personally identifiable information.

REFERENCES

Aldosari, L. A., & Altuwairesh, N. (2026). Assessing Legal Translations Generated by GPT-4 Turbo Using MQM: A Comparative Study. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, *32*(1), 172-186. http://doi.org/10.17576/3L-2026-3201-11

Alheety, A. A., khalaf, M. K., & Mohammed, H. J. (2026). Syntactic Complexity in AI-Generated vs. Human-Authored Linguistic and Literary Texts. *Arab World English Journal (AWEJ) Special Issue on CALL 17.* (12) 109-125. DOI: https://dx.doi.org/10.24093/awej/call12.7

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S. M., Nori, H., Palangi, H., Ribeiro, M., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. ArXiv, abs/2303.12712.

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *Int. J. Inf. Manag., 71*, 102642.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., Yue, J., & Wu, Y. (2023). How Close is ChatGPT to Human Experts? Comparison Corpus, Evaluation, and Detection. ArXiv, abs/2301.07597.

Hamat, A. (2024). The language of AI and human poetry: A comparative lexicometric study. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, *30*(2), 1-20.

Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). AI, write an essay for me: A large-scale comparison of human-written versus ChatGPT-generated essays. ArXiv, abs/2304.14276.

Hout, R. V., & Vermeer, A. (2007). *Lexical richness and the language proficiency of native and non-native speakers*. John Benjamins Publishing Company.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.

Iskender, A. (2023). Holy or unholy? ChatGPT and the future of universities. *World Journal of English Language, 13*(4), 1-13.

Jarvis, S., & Daller, M. (2013). *Vocabulary knowledge: Human ratings and automated measures*. John Benjamins Publishing Company.

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.

Khalil, M., & Er, E. (2023). Will ChatGPT get you a PhD? On the creative abilities of large language models. arXiv preprint arXiv:2302.05287.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307-322.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.

McNamara, D. S., Graesser, A. C., Kurby, C. A., & Louwerse, M. M. (2014). *The Coh-Metrix Common Core Text Ease and Readability Assessor*. Cambridge University Press.

Nation, I. S. (2013). *Learning vocabulary in another language.* Cambridge University Press.

Nor, N. F. M., & Aziz, J. (2026). ChatGPT as a Tool in Developing Research and Language Skills in a Research Methodology Course. *3L: Language, Linguistics, Literature® The Southeast Asian Journal of English Language Studies*, *32*(1), 187-200.

OpenAI. (2023). GPT-4 technical report. arXiv.

Ray, P. P. (2023). ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. *Internet of Things and Cyber-Physical Systems*, 3, 121-154.

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2023). The Curse of Recursion: Training on Generated Data Makes Models Forget. ArXiv, abs/2305.17493.

Tweedie, F. J., & Baayen, R. H. (1998). How variable may a constant be? Measures of lexical richness in perspective. *Computers and the Humanities, 32*(5), 323-352.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., & Tang, Y. (2023). A brief overview of ChatGPT: The history, status quo and future. *IEEE/CAA Journal of Automatica Sinica, 10*(5), 1122-1136.

Zhai, X. (2022). ChatGPT user experience: Implications for education. *SSRN Electronic Journal,* 1-18. https://doi.org/10.2139/ssrn.4312418