Arab World English Journal
INTERNATIONAL PEER REVIEWED JOURNAL ISSN: 2229-9327

Arab World English Journal (AWEJ) Special Issue on CALL Number 12. July 2026 Pp. 109-125
DOI: https://dx.doi.org/10.24093/awej/call12.7

Syntactic Complexity in AI-Generated vs. Human-Authored Linguistic and Literary Texts

Asia A. Alheety
Department of English, College of Education for Humanities, University of Al-Anbar, Ramadi, Iraq
Corresponding Author: asi24h1001@uoanbar.edu.iq

Meethaq Khamees khalaf
Department of English, College of Education for Humanities, University of Al-Anbar, Ramadi, Iraq
Email: ed.methak.khamis@uoanbar.edu.iq

Hussam J. Mohammed
Department of Artificial Intelligence, Computer Sciences and Information Technology, University of Al-Anbar, Ramadi, Iraq
Email: hussamjasim@uoanbar.edu.iq

Received:02/23/2026 Accepted:04/08/2026 Published:07/15/2026

Abstract
This paper examines how the syntactic complexity of academic writing is affected mainly by the source of authorship (AI-generated or human) or by the genre of disciplinary writing (linguistic or literary). The primary purpose is to test the syntactic-complexity differences between these variables and to establish the degree of influence of genre conventions on structural variation. The importance of the research is that it adds to the existing discussions about AI as a phenomenon in academic writing and, specifically, whether AI-generated texts are capable of syntactically reproducing the specific norms of a specific discipline of writing. To this end, the comparative corpus-based design was used. The sample consisted of 20 introduction sections: equal numbers of linguistic and literary texts and equal numbers of human-authored AI-generated texts. The Second Language Syntactic Complexity Analyzer (L2SCA) extracts fourteen syntactic complexity measures, which include length of production unit, subordination, coordination and phrasal sophistication. The results indicate that the complexity of syntax is genre-based and not source-based. Although no differences were found to be constant in both AI-generated and human academic introductions in linguistic data, there were much higher levels of subordination in literary texts of human origin. In general, the discipline genre had a more significant effect on syntax variation than the authorship source. The paper suggests the implementation of genre-sensitive models to assess AI-written academic texts and recommends additional studies that would use a bigger sample and discourse analysis.

*Keywords***:** CHATGPT, corpus linguistics, research paper introductions, syntactic complexity, AI-generated, human-authored

Cite as: 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. https://dx.doi.org/10.24093/awej/call12.7

Introduction

Syntactic complexity has been widely known to be a primary dimension of the use of language and an important feature of advanced academic writing in both computational linguistics and second language studies (Ortega, 2003; Biber & Gray, 2010). It is the variety and detailed elaboration of grammatical structures, including clausal subordination and phrasal embedding, which enables writers to compress and organize information. The capability to manipulate such structures is closely linked to the generation of research of the highest quality and to the requirements of international standards for scholarly publication (Lu, 2011). Furthermore, mastering English grammar structures is essential for enhancing language accuracy and boosting the fluency of English language speakers (Abdul Ghafoor & Challob, 2021). In this view, syntactic complexity is not an outward appearance but an expression of underlying cognitive and rhetorical command of the academic production. The recent proliferation of generative artificial intelligence, specifically large language models, hereafter denoted as AI-generated systems, has caused novel dynamics in academic writing practices. By now, AI-generated texts can generate grammatically correct and structurally sound prose that can, in many cases, be similar to the prose written by humans and used in academic writing.

Yet, a rather critical question has not been answered so far: to what degree do the texts created by AI reproduce the syntactic patterns that define expert academic writing in various disciplinary settings? According to recent research, AI-generated texts can approximate the fluency and structure of human language. Still, in numerous cases, they have certain structural regularities and biases that distinguish them among high-quality writings by humans. On the same note, there can be a claim according to which even though AI-written texts seem to be grammatically correct and scholarly in tone, they have to undergo systematic analysis based on syntactic indicators, as it will help in deciding whether they comply with the requirements of academic writing.

The research question, which is used in the present paper, arises due to the insufficient understanding of whether the syntactic complexity of academic writing is more influenced by the origin of authorship (AI-generated authorship or human-written authorship) or by the genre requirements of the discipline. Whereas former studies have focused on the role of syntactic complexity as an indicator of academic complexity, little focus has been directed on how genre-specific conventions might mediate such a correlation, especially when it comes to AI-generated text. Additionally, the available literature tends to pay attention to the superficial fluency without sufficiently investigating the structural patterns and comparing them between genres. The overall purpose of the research is to find out the variation in the complexity of syntactic features of AI-generated and human-written texts and to establish whether the variation in the authorship source or disciplinary genre is the cause of the variation.

The following research questions will be used to conduct this study: (1) Do AI-generated and human-authored linguistic and literary articles differ significantly in the area of syntactic complexity? (2) How much is the syntactic complexity affected by the discipline genre (linguistic vs. literary) in comparison with the source of the authorship (AI-generated vs. human-authored)? The theoretical foundations of the study are the known in-depth models of syntactic complexity, and in this regard, the multidimensional model of syntactic complexity by Lu (2011) can be referred to, as it divides complexity into the length of production unit, subordination, coordination, and phrasal sophistication. Moreover, Biber's multidimensional register variation provides insight into the role of genre conventions in shaping linguistic decisions across various forms of text.

According to these frameworks, the research project will assume a conceptual model where syntactic complexity will be taken as a dependent variable, which is dependent on two key factors, i.e., authorship source and disciplinary genre.

In this model, AI-generated texts are supposed to resemble structural norms that are framed by genre conventions, whereas possible variations are likely to be present in the extent and patterns of syntactic features. Even though the amount of literature on AI-written texts is increasing, there are still gaps that are essential. The past research has been more descriptive and has failed to deal adequately with critical issues like GPT authorship detection, machine and human linguistic style differences, and the role of corpus stylistics in detecting structural variation. Moreover, terminology inconsistency, e.g., using AI-generated, ChatGPT-generated, and chatbot-generated interchangeably, has added to conceptual ambiguity. To solve this problem, the current paper will always apply the AI-generated nomenclature to all texts that were created by machines.

This paper presents a systematic syntactic corpus-based comparison study of fourteen syntactic scales in parallel sets of AI-generated and human-generated linguistic and literary texts based on the L2 Syntactic Complexity Analyzer. By so doing, it will seek to establish whether syntactic complexity is dependent largely on the source of authorship or on disciplinary genre.

Literature Review

Literature review of its linguistic, pedagogical, and rhetorical ramifications. In this new area, the specific characteristics of AI-generated texts have been given special consideration, in particular, syntactic complexity, genre norms, and author. Although scientists indicate that AI systems may simulate human performance in generating grammatically correct and structurally sound texts, more recent input indicates a significant clash between statistical patterning and rhetorical intentionality. In this respect, the AIs' generated academic writing can be an indication of probabilistic regularities based on training data, as opposed to intentional stylistic and disciplinary decisions.

Casal and Kessler (2023) examined the capability of linguists to draw the line between AI- and human-generated texts in a Scopus-based study. Still, they did not identify any statistically significant results, confirming that people were successful in their attempt to detect AI-writing and human writing since the overall identification rate amounted to 38.9% (p. 1). This result contradicts the suppositions regarding distinct stylistic differences between machine writing and human writing and indicates that AI-produced texts can be quite close to traditional scholarly standards. Further, the research demonstrates that the reviewers used unreliable and unequal criteria like coherence, specificity, and perceived voice, but these characteristics did not always result in correct identification.

Under such conditions, even seasoned scholars, as the authors mention, are not specifically efficient when it comes to recognizing AI-generated texts (p. 6). This is in line with more general interests in corpus stylistics and authorship identification, with the division between machine and human linguistic styles being more and more indefinable. There are also ethical and methodological concerns that the study has brought about, such as the fact that AI tools are developed to sound like human-generated language (p. 1), which in turn makes it more challenging to define authorship and assessment in academic writing.

Besides this, Moussa and Belhiah (2024) present empirical data that AI-based writing tools can contribute to various aspects of writing, such as proficiency, creativity, and structural organization. Their results indicate that the use of AI results in the positive outcomes of language proficiency,

creativity, organizational skills, and vocabulary use (p. 138) and that AI has not been solely contributing to surface accuracy but to other additional facets of writing performance. Notably, their comparison of sentence composition reveals that students who are exposed to AI support are more likely to create more complicated syntactic frameworks, such as complex sentences, the use of subordinate clauses, and the use of a variety of punctuation (p. 143). Nevertheless, it is also noted in the study that there is variability between the levels of proficiency, with lower-level learners still depending on simpler structures even with the aid of AI. Moreover, the authors note that AI tools stimulate experimental writing style, which results in a wide scope of linguistic complexity (p. 149).

These results indicate that although AI-assisted writing can increase structural complexity, it is still uneven and conditional on the competence of the user and not necessarily dictated by the technology itself. Pedagogically, Ling Jen and Salam (2024) offer an overview of the use of artificial intelligence in essay writing by identifying both the possibilities and constraints of AI applications in the learning process (p. 90). A recent study by Pyran et al. (2025) has emphasized the linguistic performance of AI by discussing that AI tools can improve the lexical diversity and grammatical accuracy, but these tools may cause "passive acceptance" of AI outputs. This highlights the need for further exploration of syntactic structure, which this study aims to address using the L2SCA tool. Furthermore, many pedagogical studies emphasize the role of digital platforms in language acquisition. Majida Ibrahim Mukhlif and Ala'a Ismael Challob (2021) highlighted that integrating social media tools is more effective than traditional methods ,especially in enhancing the vocabulary knowledge of EFL students. Although it may be helpful in generating ideas, providing feedback, and drafting, the authors suggest that there is no empirical evidence on how AI can be effective in writing performance, and a geographical imbalance in the current research. On the same note, in an AWEJ article, the study by Al-Zubaidi et al. (2024) shows that students and faculty find AI tools useful, but their applications in academia are not clear since there are no institutional guidelines (p. 4). These results highlight the importance of having a more sensitive view of the way AI-generated writing can be complied with the established academic norms.

Further evidence is given by Chung et al. (2025), who focus on the strategies of vocabulary learning in a group of undergraduate students in an English for Academic Purposes setting. According to their findings, strategic competence is a key factor in academic language development and cannot be entirely reproduced by AI systems (p. 56). This underscores the wider weakness of AI-generated writing, namely, it can mimic structural strategies, but not the cognitive and rhetorical strategies that expert human writing entails.

Even with these improvements, the literature is more descriptive and fragmented, and there is little combination of syntactic, stylistic, and genre-based approaches. Specifically, the void in the research is the absence of syntactic complexity studies that systematically interact with genre conventions and authorship source and are conducted in controlled corpora. Furthermore, the conceptual ambiguity was also caused by the confusion in the terminology, including the usage of AI-generated and ChatGPT-generated interchangeably, as well as chatbot-generated. To solve this problem, the current research chooses to use the phrase AI-generated throughout.

According to the analyzed literature, the conceptual model according to which the syntactic complexity is considered as a multidimensional variable dependent on two major independent variables, the source of authorship (AI-generated or human-authored) and the genre of discipline (linguistic or literary). This model supposes that although AI-generated texts can be close to

structural norms, genre conventions are of primary importance in the realization of syntax. The current research aims to go beyond the descriptive explanations and offer a more empirically supported and theoretically justified study by incorporating the use of corpus-based analysis with the well-known models related to it, including Biber's multidimensional model and the measure of syntactic complexity developed by Lu.

By doing this, the research will fill an urgent gap in the literature by providing an analysis of the extent to which syntactic complexity variation is fueled by the disposition of technological authorship or deep-rooted disciplinary standards, thus joining the current discussions regarding whether or not artificial intelligence plays a role in linguistic and literary writings. The proposed conceptual model of the study, based on the reviewed literature, assumes that syntactic complexity is considered an independent and multidimensional dependent variable affected by two independent variables of primary importance: authorship source and disciplinary genre. Authorship origin is operationalized as binary between AI-generated and human-authored texts, whereas disciplinary genre is defined in terms of linguistic and literary writings.

Based on the framework of Lu (2011), syntactic complexity is conceptualized in four fields that are interdependent, which include the length of units of production, subordination, coordination, and phrasal sophistication. Also, the contribution of Biber's multidimensional model helps to inform the importance of genre conventions in influencing the linguistic variation in academic discourse. It is assumed that author source and genre both have an impact on syntactic realization, but that genre should be more dominant as it is closely linked with all norms of the discipline and all the conventions of rhetoric.

Although the text generated by AI can resemble human-textual structural patterns, it is theorized that such text represents probabilistic regularities as opposed to genre-sensitive rhetorical decisions. While AI tools can simulate complex structures, human writing, particularly by EFL learners, often exhibits distinct patterns of discourse errors. Meethaq Khamees Khalaf and Rana Abdulwahid Fadhil (2020) highlighted that syntactic and lexical errors are prevalent among learners due to mother-tongue interference, a characteristic that contrasts with the often linear and error-free output of AI. Thus, syntactic complexity can be seen as the result of a kind of interaction between the technological production (AI vs. human) and the disciplinary restrictions (linguistic vs. literary). The model predicts, therefore, (1) possible variations in AI-generated and human-written texts, (2) more vigorous differences due to genre conventions, and (3) an interaction effect between authorship source and genre that might exist. This conceptualization offers a concrete way of analyzing the fact that syntactic complexity may depend mainly on the source of authorship or mostly norms of a given field, thus filling a major gap that was identified in earlier studies.

Methodology

Research Design

The research design applied to this study is a quantitative research design that uses a corpus to examine the syntactic complexity of AI-generated and human-written academic introduction texts. Consistent with Lu (2011), corpus-based analysis offers a logical and objective guideline for comparing a variety of dimensions of linguistic structure among controlled sets of data. The term AI-generated is implemented across the board to present conceptual consistency with all the texts generated by a machine. The research is based on a comparative design that will involve two independent variables: authorship source (AI-generated and human-authored) and disciplinary genre (linguistic and literary). Syntactic complexity is the dependent variable, and is

operationalized based on several indices based on the framework of Lu (2011). The present study has a balanced corpus in both the field of linguistics and literary studies, to resolve discrepancies that were found in the previous versions of the dataset.

The texts produced by AI and those produced by humans are distributed equally in these two genres, which guarantees that there is no discrepancy between the stated area of research and the real data that will be analyzed. Since the sample size is relatively small (N = 20), multivariate statistical tests like MANOVA are not utilized carefully. Although MANOVA is kept in order to investigate general trends in more than one syntactic area, the findings of MANOVA are viewed as exploratory, but not confirmatory. In order to create analytical strength, syntactic measures are clustered into domains according to Lu (2011), thus decreasing the dimensions and creating statistical stability. The comparatively limited sample size serves as an indicator of the relative demands of the controlled and time-consuming activity of creating and matching AI-generated texts with those of human authors, which requires thorough prompt design and validation to reach comparable outcomes.

Data Sources

The research is grounded on a corpus of 20 introduction sections of research articles, including 10 human-written ones and 10 AI-generated texts. This paired-corpus design allows the syntactic complexity to be compared across the authorship sources in a controlled and multi-dimensional way, and ensures that the topic and genre are identical. The introductions of the human authors were chosen out of the Scopus-indexed journals in the area of linguistics and of literary studies based on three criteria. To select the journals, first, journals that are in Quartile 2 will be selected as they are a standard of high-quality prose and at the same time are well edited without the high level of structural density that Quartile 1 journals are known to have (González-Pereira et al., 2010). Second, the articles were chosen on the basis of journals with CiteScore values between 1.5 and 3.0 that would allow relative homogeneity and exclude low-impact and highly specific outliers. Third, the publications were conducted no later than 2023-2025 to mirror the current academic writing traditions.

The shortcoming of using only human-made data was directly solved by constructing a parallel AI-made corpus in controlled experimental conditions. In the case of every human-written introduction, an AI-generated version was created with the help of GPT-4, and there was one-to-one correspondence between them in the dataset. The prompts were designed to mimic the original topic, disciplinary setting, and the rhetorical intent, but were not to involve any type of direct copying or paraphrasing. Research focus, argumentative structure, and expected academic tone are the aspects that each prompt has explicitly spelled out to be used to generate their works. Further to increase the methodological rigor, various outputs were created on each prompt, and the most representative version was chosen according to pre-defined criteria, such as thematic consistency, structural integrity, and compliance with academic norms.

No duplication of manuals was done, and little cleaning was done to take out parts that were not completed or irrelevant without changing the syntactic structures. This process will make sure that the texts produced by the AI are original products of the model and not edited texts. This controlled generation and matching process is the reason the AI-generated texts would be considered correct counterparts to human-written introductions, and the worry that the dataset is entirely human-produced texts would be eliminated. Consequently, the research allows the comparative approach to be critically balanced and methodological, where the syntactic

complexity can be studied as the main variable, irrespective of the topic, genre, and rhetorical variation.

Research Instruments

The primary instruments of this study are textual data and the Second Language Syntactic Complexity Analyzer (L2SCA, Version 3.3) (Lu, 2010). The dataset consists of 20 introduction sections drawn from linguistic and literary research articles, including 10 human-authored texts and 10 AI-generated counterparts produced for comparative purposes. These texts were compiled into two specialized corpora to enable systematic analysis of syntactic complexity across authorship sources and disciplinary genres. Given the objectives of the study, the analysis focuses on the introduction sections of linguistic and literary texts, as these sections are typically dense in information, argumentation, and rhetorical organization, a methodological choice supported by Biber et al. (2011).

This consideration enables a finer analysis of syntactic complexity in a high-stakes academic genre. Fourteen syntactic complexity measures, which are the indicators of four key dimensions: length of production units, subordination, coordination, and phrasal sophistication, were automatically extracted using the L2SCA tool (Lu, 2010). The tool has been proven to be a valid instrument in earlier studies as a tool that can be used to measure the syntactic complexity of linguistic and literary texts (Lu, 2011). All the texts were subjected to a number of preprocessing steps before analysis so as to bring uniformity and accuracy.

These involved transforming all the documents to plain text, eliminating references, tables, and non-propositional items, as well as standardizing each text to a standard length of about 250 words. This standardization was done to guard against the variation in the text length, which could otherwise affect the syntactic complexity measures. In order to ascertain the reliability of the analysis, consistent preprocessing procedures were used. Moreover, some of the texts processed were manually checked to ensure that sentence segmentation and parsing results produced by the L2SCA tool were correct. This was done to ensure that the systematic errors were not brought about by automated extraction. On the whole, this methodological framework of combining textual data under control with a validated computational instrument offers a strong methodological basis to analyze the syntactic complexity of the AI-generated and human-written texts.

Data Collection Procedures

The procedures of the corpus collection were also designed systematically in such a way that they can be compared and be methodologically rigorous. All the documents were translated into plain text, and the unnecessary materials like references, tables, and metadata were eliminated according to L2SCA guidelines (Lu, 2010). The process of randomizing the texts written by humans was stratified. The initial stage in retrieving the eligible articles included the use of predefined criteria, followed by clustering them by discipline (linguistic and literary), followed by the selection of introduction sections at random per group, in order to maintain a good representation of different genres. To ensure that datasets are equivalent, a topic-matching process was applied. Every text written by a human was accompanied by the corresponding AI-generated one created through the same topic of research, the situation of its usage, and the purposes it serves. The prompts have been designed to mirror the main idea and the scholarly style of the initial texts without duplicating their format. Both linguistic and literary disciplines are represented in the dataset, which is also consistent with the scope of the study. The matching of samples was done in

the same disciplinary categories according to Lu (2011), and this reduced variability and offered a valid cross-source comparison.

Data Analysis Procedures

A set of procedures was applied to answer the research questions. A syntactic complexity objective analysis was then done with the L2SCA tool. The profiling of all the syntactic measures is enforced by feeding each corpus into the analyzer ,resulting in a healthy comparison according to Ortega (2003). The obtained results were used as the foundation of further comparisons, whereby every human-created corpus was compared to the corresponding ChatGPT one. These comparative results gave the basis of deep discussions and additional statistical analysis. Complexity measures and analysis tool of syntax.

Syntactic Complexity Measures and Analysis Tool

According to Bulté and Housen (2014), cognitive requirements related to text processing depend on the complexity of structures. As such, the texts collected were analyzed, employing all 14 measures that the L2SCA offers (Lu, 2010) to conduct a thorough analysis that would cover four major categories, namely, length of production unit, amount of subordination, amount of coordination, and degree of phrasal complexity. These indicators give information in detail about the degree of connectivity and sophistication in the text. L2SCA was selected because it offers a holistic approach, which considers various issues of linguistic complexity (Lu, 2011).

Methodological Procedure for Inferential Analyses

The analysis of the differences across the corpora was undertaken with the help of a one-way MANOVA that was conducted with SPSS. According to Pallant (2020), the choice of this type of option is due to the possibility of testing several dependent variables at the same time. In order to establish the core dimensions of syntactic complexity that would be considered in this analysis, four dependent variables were determined on the basis of the classification framework, which was offered by Lu (2011). The most important statistic that was applied to evaluate the importance of the multivariate effect of the independent variable on the dependent variables was the Corpus-Type independent factor (Human vs. ChatGPT) ,Wilks Lambda (Tabachnick and Fidell, 2013).

Results

Table 1 shows the complete list of syntactic complexity measures employed in this study in order to make the interpretation of the findings easier. These measures were based on the L2 Syntactic Complexity Analyzer and are the analytical basis of the structural variation of the AI-generated texts and the human-written texts. The table is used as a reference template that elucidates how every measure addresses a particular dimension of syntactic complexity that offers consistency as to the interpretation of the further outcomes.

Table 1 L2 syntactic complexity measures (Category – Label – Description)

Category	Label	Description
Length of Production Unit	MLC	Mean length of clause
Length of Production Unit	MLS	Mean length of sentence

Length of Production Unit	MLT	Mean length of T-unit
Amount of Subordination	C/T	Number of clauses per T-unit
Amount of Subordination	CT/T	Complex T-unit ratio
Amount of Subordination	DC/C	Number of dependent clauses per clause
Amount of Subordination	DC/T	Number of dependent clauses per T-unit
Amount of Coordination	CP/C	Number of coordinate phrases per clause
Amount of Coordination	CP/T	Number of coordinate phrases per T-unit
Amount of Coordination	T/S	Number of T-units per sentence
Degree of Phrasal Sophistication	CN/C	Number of complex nominals per clause
Degree of Phrasal Sophistication	CN/T	Number of complex nominals per T-unit
Overall Sentence Complexity	VP/T	Number of verb phrases per T-unit
Overall Sentence Complexity	C/S	Number of clauses per sentence

Table 1 shows all the syntactic complexity measures that were used in this research in order to analyze the findings. These are the measures, obtained with the help of the L2 Syntactic Complexity Analyzer, which is the structure of analysis of the variability of the texts generated by AI and those created by humans. The table is a reference guide, making it clear how each measure is related to a certain dimension of syntactic complexity and giving uniformity to the interpretation of other analyses.

Table 2. ***Grouping syntactic complexity measures (Ai &Lu-style domains)***

Table 2 categorizes the L2SCA indices in four main areas of syntactic complexity, which include length of production units, subordination, coordination, and phrasal sophistication. This categorization enhances the interpretation because it associates single measures with larger structural dimensions and because the further analysis is based on consistent patterns of syntactic variation instead of single measures.

Domain	Measures Included
Length of Production Unit	MLS, MLT, MLC
Subordination	C/T, CT/T, DC/C, DC/T
Coordination	CP/C, CP/T, T/S
Phrasal Sophistication	CN/C, CN/T

This table groups the individual L2SCA indices into four widely recognized domains of syntactic complexity: length of production unit, subordination, coordination, and phrasal sophistication. This classification, consistent with Ai and Lu's framework, enhances interpretability by linking multiple specific indices to coherent structural constructs. The domains were entered sequentially into the multivariate and follow-up analyses to ensure that statistical testing focused on unified dimensions of syntactic complexity rather than isolated measures.

Table 3. Descriptive Results of Syntactic Complexity (Linguistic Texts): Human vs ChatGPT
The table shows that syntactic complexity in linguistic introductions is largely similar between human-authored and AI-generated texts, with only minor differences across measures. Human texts tend to have slightly longer production units and higher subordination, indicating somewhat more complex structuring, while AI-generated texts show higher coordination, reflecting a more linear style. Most other measures, including clause density and phrasal sophistication, remain nearly identical, suggesting that overall syntactic complexity is comparable across both groups.

Measure	Human Mean	Human SD	Human N	ChatGPT Mean	ChatGPT SD	ChatGPT N
MLS	49.679	6.576	10	43.234	12.052	10
MLT	54.061	7.048	10	49.222	6.944	10
MLC	22.698	4.377	10	20.453	7.210	10
C/S	2.228	0.449	10	2.250	0.358	10
VP/T	14.251	3.390	10	12.668	3.875	10
C/T	2.229	0.323	10	2.258	0.460	10
DC/C	0.444	0.135	10	0.368	0.096	10
DC/T	0.891	0.291	10	0.856	0.127	10
CT/T	0.538	0.176	10	0.500	0.157	10
CP/C	1.050	0.275	10	1.222	0.350	10
CP/T	2.048	0.585	10	2.339	0.692	10
T/S	1.000	0.000	10	1.000	0.000	10
CN/C	0.720	0.167	10	0.686	0.207	10
CN/T	1.486	0.473	10	1.534	0.387	10

This table summarizes the descriptive indices of syntactic complexity for the Linguistic introductions using standardized 250-word samples (N=10 per group). In this dataset, Human and ChatGPT show non-identical profiles across the L2SCA measures; for example, MLS differs between Human (49.679±6.576) and ChatGPT (43.234±12.052), and CN/T also differs between Human (1.486±0.473) and ChatGPT (1.534±0.387). The remaining indices follow the same reporting structure and are used in the inferential tables below.

Table 4. Descriptive Results of Syntactic Complexity (Literary Texts): Human vs ChatGPT

To provide additional evidence for the study of linguistic texts, Table 4 presents descriptive statistics for syntactic complexity measures of literary introductions in human- and AI-written corpora. The comparison enables us to discuss whether the tendencies identified in the linguistic field can be applied to literary writing, which is usually marked by various stylistic and rhetorical norms.

Measure	Human Mean	Human SD	Human N	ChatGPT Mean	ChatGPT SD	ChatGPT N
MLS	22.018	4.323	10	20.654	1.556	10
MLT	21.394	3.723	10	21.542	1.146	10
MLC	12.999	1.972	10	15.330	2.325	10
C/S	1.873	0.402	10	1.377	0.144	10
VP/T	6.984	0.946	10	6.558	0.805	10
C/T	1.519	0.467	10	1.510	0.292	10
DC/C	0.437	0.087	10	0.284	0.092	10
DC/T	0.846	0.579	10	0.531	0.155	10
CT/T	0.772	0.285	10	0.346	0.134	10
CP/C	0.559	0.191	10	0.888	0.151	10
CP/T	1.048	0.205	10	1.192	0.159	10
T/S	1.000	0.000	10	1.000	0.000	10
CN/C	0.271	0.154	10	0.362	0.157	10
CN/T	0.526	0.318	10	0.389	0.159	10

Human Literary works are more subordinate, suggesting that their structures are more complex and hierarchical. On the contrary, AI-generated texts are more coordinated and create more linear sentence structures. The phrasal sophistication is intermingled between the two groups. Altogether, human writing often exhibits a more complex syntactic structure, whereas AI-writing is characterized by superficial development.

Table 5. Multivariate Tests (MANOVA) Results

This table illustrates the outcomes of the multivariate analysis of variance (MANOVA) for testing the influence of authorship source (human vs AI-generated), genre (linguistic vs literary), and their combination on the four areas of syntactic complexity, as shown in Table 5.

Effect	Wilks' Λ	Num df	Den df	F	p
Source	0.6731	4	33	4.0066	0.0093
Genre	0.0711	4	33	107.8540	0.0000
Source×Genre	0.7433	4	33	2.8494	0.0392

In this table, the tests (MANOVA) of the 2 x 2 design with Source (Human vs ChatGPT) and Genre (Linguistic vs Literary) as predictors of the four domains of syntactic complexity are defined and the results are assessed by using the Wilks 2 x -tests (Source p = 0.0093, Genre p = 0.0000, Source x Genre p = 0.0392). The findings imply that the authorship source and genre significantly affect the complexity of syntax; however, the influence of genre is considerably stronger. Despite the fact that the differences between the AI-generated and human-generated texts are also extremely huge, they are still relatively modest compared to the effects of disciplinary conventions. The high interactivity effect also indicates that these differences vary across genres, which are more dominant in certain contexts, especially in literary texts. In general, the syntactic complexity seems to be genre-oriented rather than authorship-oriented.

Table 6. Tests of Between-Subjects Effects (Follow-up ANOVAs for each domain)

Table 6 displays follow-up two-way ANOVAs of the influence of authorship source, genre, and their interaction on the four domains of syntactic complexity.

Dependent Variable	Effect	df	F	p
LengthPU	Source	1	4.6516	0.0378
LengthPU	Genre	1	474.7265	0.0000
LengthPU	Source×Genre	1	6.4730	0.0154
Subordination	Source	1	8.8923	0.0051
Subordination	Genre	1	28.6740	0.0000
Subordination	Source×Genre	1	5.2204	0.0283
Coordination	Source	1	8.5504	0.0059

Coordination	Genre	1	86.2762	0.0000
Coordination	Source×Genre	1	0.0009	0.9766
PhrasalSoph	Source	1	0.0146	0.9045
PhrasalSoph	Genre	1	125.6186	0.0000
PhrasalSoph	Source×Genre	1	0.0562	0.8140

Table 6 reports follow-up two-way ANOVAs of each domain to determine whether the differences could be related to the Source, Genre, or a combination of the two. These univariate tests are to be viewed alongside the descriptive profiles in Tables 3 and 4. The findings indicate that the genre effect is remarkably high and significant in all the fields, especially in the production units' length, coordination, and phrasal sophistication. By comparison, the authorship source produces a smaller effect, being significant only in length, subordination, and coordination, and not in phrasal sophistication. The interaction effect is high for length of production units and subordination, meaning that there is only a difference between AI-generated and human-written text in these areas; otherwise, they remain constant in other areas.

Table 7. Pairwise Comparisons / Post-hoc (as applicable)

Table 7 presents specific pairwise tests aimed at further examining differences between AI-generated and human-written texts in particular situations and genre-related differences in the human-written texts, using Bonferroni-adjusted p-values.

Domain	Context	Comparison		t	p	p (Bonferroni)
LengthPU	within Literary	Human ChatGPT	vs	-0.4631	0.6516	1.0000
LengthPU	within Human	Linguistic Literary	vs	20.6622	0.0000	0.0000
Subordination	within Literary	Human ChatGPT	vs	3.2256	0.0066	0.0133
Subordination	within Human	Linguistic Literary	vs	1.8135	0.0908	0.1816
Coordination	within Literary	Human ChatGPT	vs	-4.4302	0.0004	0.0008
Coordination	within Human	Linguistic Literary	vs	7.5204	0.0000	0.0000
PhrasalSoph	within Literary	Human ChatGPT	vs	0.3235	0.7506	1.0000
PhrasalSoph	within Human	Linguistic Literary	vs	6.6501	0.0000	0.0000

Table 7 presents specific contrasts in pair-wise data in accordance with the research objective (Human vs ChatGPT in Literary; Linguistic vs Literary in Human) with Bonferroni-

adjusted p-values. The findings show that the variations between human texts and AI texts are notable in the subordination and coordination in the literary genre, whereby human texts have more subordination, and AI texts have more coordination. On the other hand, the length of production units and phrasal sophistication do not differ significantly in the production units as well as in the same genre. Also, intertextual comparison in human-written sources demonstrates that there are large genre effects in most areas, especially length, coordination, and phrasal sophistication, which supports the overwhelming role of genre in the variation of syntax.

Discussion

The research was aimed at investigating whether the syntactic complexity in texts is determined more by the origin of authorship (AI-generated or human-authored) or by the discipline genre. The results openly respond to the research questions and give valuable information on the quality of the AI -created writing. In relation to the first research question, the findings reveal that the variations between AI-generated and human-written texts are not always relevant and significant in all spheres of syntactic complexity. Although there was some difference, especially in subordination and coordination, the difference was minor. This observation concurs with that of Casal and Kessler (2023), who revealed that even experienced linguists find it difficult to consistently tell the difference between AI-generated and human-written texts, implying that AI systems may be able to successfully guess the traditional academic language form. In a similar study, Moussa and Belhiah (2024) have discovered that AI-assisted writing can improve the complexity of the structure, but not in all contexts.

Contrastingly, the second research question reveals that the disciplinary genre's impact on syntactic complexity is substantially greater than the source of authorship. The findings have always revealed a significant difference in both linguistic and literary writings in various syntactic areas, especially in the length of production units, coordination, and phrasal sophistication. These results confirm the opinion that syntactic variation is significantly determined by the conventions and rhetorical demands peculiar to genres, which is suggested in the multidimensional model by Biber (2011) framework by Lu.

Syntactic complexity, in this sense, seems to be a discipline-specific phenomenon rather than the fact that the text was created by a human or an AI system.

This phenomenon also indicates that the difference between AI-generated and human-written texts is context-specific due to the interaction effect between the authorship source and genre. Particularly, the difference is more evident in literary works, particularly in the patterns of subordination, where the texts written by humans are characterized by a higher level of hierarchical organization. This helps justify the thesis that even though AIs may be able to reproduce superficial patterns in writing, it may nonetheless not be effective when it comes to other, more subtle, genre-aware rhetorical structures.

These results have significant implications for research and practice. First, they question the belief that the syntactic properties allow the easy differentiation of AI-generated writings. Rather, they propose that assessment systems are to transcend the complexities of surface and include genre-sensitive and discourse-level standards. Second, the findings reveal that the significance of disciplinary conventions in developing academic writing is important, and context-sensitive practices are important in linguistic analysis and AI rating.

On the whole, the given work can be considered a contribution to the current discussion concerning the topic of artificial intelligence in academic writing because it proves that syntactic complexity

does not rely on the source but mostly on the genre. The study should be further extended in future studies by increasing the size of the data and adding other linguistic and discourse attributes to conduct further studies on the interplay of AI-generated writing, genre conventions, and rhetorical effectiveness.

Conclusion

The primary objective of the research was to investigate whether syntactic complexity in the introduction of academic research articles is mostly determined by the source of authorship (AI-generated or human-authored) or the genre of the research article. To do so, the paper utilized a corpus-based approach of analysing standardised samples of 250-word introductions with the help of fourteen measures of syntactic complexity that were divided into four domains, which are length of production units, subordination, coordination, and phrasal sophistication.

The results show that the syntactic complexity is much more affected by the disciplinary genre than by the authorship source. Although some slight dissimilarities were found between texts generated by AI and texts written by people, they were not uniform or regular in the reviewed aspects. Rather, syntactic variation was also genre-motivated to a great extent. Such findings lead to the conclusion that AI-generated texts can be used to reproduce the structural norms of academic writing, especially with the help of disciplinary expectations, but structural similarity does not always mean the rhetorical one.

Comprehensively, the research shows that syntactic complexity is mainly genre-based as opposed to source-based writing, which is part of the continuing debates of the use of artificial intelligence in academic writing. It also emphasizes the need to take genre-sensitive methods when assessing AI-generated texts, suggesting that more studies should be conducted on larger corpora and by adding discourse-level analysis.

Funding: This research is not funded
Acknowledgment: Not applicable
Conflicts of Interest: The authors declare no conflict of interest
AI and AI-assisted technologies were not used.

About the author:

Asia Alheety is a teacher in Hanaa al Alshaibani Secondary School for Girls, Iraq. ition for several years. She also participated in English curriculum development courses with UNESCO to develop the English language curriculum in Iraq in 2002. In addition, she took part in training courses for English language teachers and instructors in Ramadi. She is also interested in publishing research in the field of Applied linguistics. ORCID ID: https://orcid.org/0009-0001-0099-7838

Meethaq Khamees Khalaf is an Assistant Professor and the Head of the Department of English at the University of Anbar, Iraq. With a focus on English Language and Linguistics, he holds a B.A. and an M.A. from the College of Education for Humanities at the same university. Over his academic career, he has taught various subjects, including Linguistics, Morphology, Discourse Analysis, and composition, and has supervised several impactful theses in his field.
ORCID ID:https://orcid.org/0000-0002-9352-3999

Hussam J. Mohammed is a Doctor of Philosophy specializing in Digital Forensics and Artificial Intelligence. He earned his BSc and MSc degrees in computer science from the University of Anbar, Iraq, before pursuing his PhD in digital forensics and cybersecurity at the University of Plymouth, UK. Currently, Hussam is a faculty member in the Al Department at the University of Anbar, where he is actively involved in research and academic development.
ORCID ID : https://orcid.org/0000-0001-9925-3945

References

Abdul Ghafoor, N. H., & Challob, A. I. (2021). The influence of using online discussion on developing EFL students' grammar knowledge. *Korean Journal of English Language and Linguistics, 21, 837–855*.http://doi.org/10.15738/kjell.21..202109.837.

Al-Zubaidi, K., Jaafari, M., & Touzani, F. Z. (2024). Impact of ChatGPT on academic writing at Moroccan universities. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, 4–25. https://dx.doi.org/10.24093/awej/ChatGPT.1

Biber, D., & Gray, B. (2010). Challenging stereotypes about academic writing: Complexity, elaboration, explicitness. *Journal of English for Academic Purposes, 9(1)*, 2–20. https://doi.org/10.1016/j.jeap.2010.01.001.

Biber, D., et al. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *TESOL Quarterly*, *45(1)*, 5–35. https://doi.org/10.5054/tq.2011.244483.

Bulté, B., & Housen, A. (2014). *Conceptualizing and measuring short-term changes in L2 writing complexity*. *Journal of Second Language Writing, 26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005.

Casal, J. E., & Kessler, M. (2023). Can linguists distinguish between ChatGPT/AI and human writing? A study of research ethics and academic publishing. *Research Methods in Applied Linguistics, 2(3)*, 100068. https://doi.org/10.1016/j.rmal.2023.100068.

Chung, E. et al. (2025). A mixed-methods study of strategy use in academic vocabulary learning among first-year undergraduates in Hong Kong. *Journal of English for Academic Purposes*, *74*, 101482. https://doi.org/10.1016/j.jeap.2025.101482.

González-Pereira et al. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics, 4(3*), 379–391. https://doi.org/10.1016/j.joi.2010.03.002.

Khalaf, M. K., & Fadhil, R. A. (2020). Analysis of written discourse errors committed by Iraqi EFL learners. *Dirasat: Human and Social Sciences, 47(2), 521–532.*

Ling Jen, S., & Salam, A. R. (2024). Using artificial intelligence for essay writing. *Arab World English Journal (AWEJ) Special Issue on ChatGPT*, 90–99. https://dx.doi.org/10.24093/awej/ChatGPT.5

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45(1)*, 36–62. https://doi.org/10.5054/tq.2011.240859.

Mukhlif, M. I., & Challob, A. I. (2021). Enhancing vocabulary knowledge among secondary school EFL students by using Facebook. *Turkish Online Journal of Distance Education (TOJDE), 22(3), 246–261.* https://doi.org/10.17718/tojde.961834.

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24(4)*, 492–518. https://doi.org/10.1093/applin/24.4.492.

Pallant, J. (2020). *SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS* (7th ed.). Routledge.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics*. Pearson Education.

Pryma, V., Pelivan, O., Teletska, T., Tsobenko, O., & Zagrebelna, N. (2025). AI-writing assistants and student competence: A linguistic aspect. *Arab World English Journal (AWEJ), Special Issue on Artificial Intelligence,319–329*. https://dx.doi.org/10.24093/awej/AI.18