A Syntactic Comparison between Human and AI-Generated Sentences: Exploring Patterns and Deviations

Abdul Sattar Channa[1]

Lecturer, Department of English, FUUAST, Pakistan

Ali Raza[2]

Lecturer, Department of English, FUUAST, Pakistan

Mansoor Ali[3]

MPhil Linguist Scholar, Department of English, SMI University, Karachi, Sindh, Pakistan

Corresponding Author's Email absattar040@gmail.com

ABSTRACT

This study explores the syntactic complexity, grammatical correctness, and structural variety in the essays by Pakistani ESL learners as compared to the AI (ChatGPT) essays. Three main types of analysis, syntactic complexity, error and fluency analysis and syntax tree pattern variation was conducted on nine essays from each group (Human and AI). The present study employed a mixed approach to data analysis. Using Coh-Metrix, a computational text analysis tool the syntactic complexity, grammatical accuracy and cohesion measures were taken whereas the syntax tree pattern analysis was conducted manually by the researcher. The findings reveal the AI generated essays exhibit higher complexity, fewer errors and varied clause structures, relative to human essays by 5th semester BS English learners, which are moderate and with scattered grammatical errors and less complex sentence patterns. These results indicate the developmental needs of ESL students in learning to master and have command over complex sentence structures and improve fluency in writing. This study has implications for the application of AI-based resources in ESL writing teaching by helping to narrow the gap and to develop advanced language skills.

Keywords: *ESL writing, Syntactic complexity, Grammatical accuracy, AI-generated essays, Pakistani ESL learners, Academic writing, Error analysis, Sentence structure*

Introduction

The rapid development in the field of artificial intelligence (AI) and natural language processing (NLP) has resulted in the emergence of powerful language models capable of producing human-like text for semantic content. For AI systems such as GPT and BERT (Devlin et al., 2019; Radford et al., 2019) this is increasingly important: the syntactic properties of their output are not orthogonal to the original writing task. This paper presents a comprehensive syntactic analysis of human from scratch sentences and their AI-generated counterparts, covering syntactic complexity, diversity, fluency and grammaticality. Our work has two primary goals: 1) To contrast

the syntactic complexity and diversity between human and I-written texts, and 2) To gauge overall AI-generated fluency and grammaticality observations against those of human-written text.

The motivation of this work can be found in a high dependency on AI systems deployed in different areas—from producing content and providing customer service up to academic writing. As humans and AI systems generate language to perform similar tasks, a systematic study on the similarities and disparities between their sentences is crucial. In attempting to answer the latter research questions— "What are the syntactic differences between AI-generated text and human-generated text?" and "What is the pattern of syntactic complexity of the two types of texts?" this work adds to the growing literature which interrogates the linguistic norms embedded in AI-generated text.

Creation and development of AI, draws attention to concerns such as fluency and grammaticality, which continue to be of crucial importance in human-computer interactions (Jurafsky and Martin, 2019), and checks for patterns such as subordinate clause frequency, nominalizations, and verb phrase structures to uncover indicators of what AI-generated language looks like. The exploration is important to the extent that it may help shape both the theorization of language important not only for evaluating AI systems, but also for developing more robust language models that have the ability to emulate the subtlety of human language. The approach compares syntactic complexity. Extracting syntactic regularity in text is essential for understanding variation in clause, underlying grammatical structures and the overall fluency of written expression.

In addition, the analysis highlights pedagogical and ethical concerns. In the world where AI systems are being increasingly embedded within educational environment and content generation platforms, being able to recognize even if subtle syntactic distinctions play can serve as a warning to educators and content moderators to be able to recognize when content may be AI authored, to facilitate proper attribution, and to enforce against potential offenses against academic integrity (Bender et al., 2021). Therefore, this study sits at the juncture of computational linguistics, applied linguistics, and AI ethics.

Research Objectives:

- To compare syntactic complexity and variety between human-written and AI-generated texts.
- To evaluate the fluency and grammaticality of AI output in comparison to human-authored text.

Research Questions:

- What are the syntactic similarities and differences between AI-generated and human-generated texts?
- How does syntactic complexity vary between the two types of texts?

Literature Review

Early investigations of syntactic complexity began as tools in the field of linguistics to evaluate language proficiency and development (Biber, 1991; Lu, 2010). Different indices (e.g., mean T-unit length, subordination index, ratio of clauses, clause density) have been adopted in human writing to measure structural complexity (Ortega, 2003). These measurements can act as a baseline for comparative researches in automatic text generation for AI. More recently, such traditional frameworks have been extended by means of computational methods, allowing for the exploration of syntactic phenomena on a large scale and to compare human and AI corpora (Lu 2010).

AI-Generated Language and Its Evaluation

By utilizing methods such as neural network language models (Vaswani et al., 2017), automatically generated AI text has reached fluency and grammaticality on par with human generated-speech. However, despite these advances, questions persist about how well these texts "understand" language, as well as the extent to which they are merely memorizing patterns at the surface of the language. AI outputs are fluent and often grammatically correct, they do at times contain ungrammatical (or at best odd sounding) structures and lack the in-depth contextual integration found in human language. The strengths and weaknesses of AI on variable syntactic structures have also been investigated (Narayan et al., 2021), which suggests that differences in syntactic complexity might act as clues for detecting machine-generated language.

Comparing AI and Human Syntactic Patterns

The contrast between AI and human syntactic patterns has been an object of recent investigation. Such rigorous comparisons have been applied to corpora from various sources in numerous studies observed differences in syntactic constructions between AI-generated news articles and their human-authored counterparts. The methods used in these comparisons are derived from statistical analyses of syntactic constructions such as the rates of subordination and that of passive. Other than raw counts, qualitative analyses have shown that beyond surface formal similarities, lines generated by AI systems occasionally drift in the space of syntactic cohesion and discourse connectivity.

The Role of Computational Linguistics and Evaluation Metrics

However, with the development of computational linguistics, reliable syntactic tools became available. Dependency parsing (McDonald et al., 2006) and phrase structure parsing (Klein and Manning, 2003) are successful for describing the formal grammar of sentences. Recent methodological advances have permitted complexity indices to be calculated and compared across text types. These systems not only contrast, but offer insight into the extent to which AI models appear to adapt to human syntactic norms (Jurafsky & Martin, 2019). Additionally, multi-dimensional metric which jointly measures syntactic correctness and semantic consistency are also introduced to evaluate language generation quality (Bender et al., 2021).

Gaps in the Literature and Study Contributions

Although there is an increasing interest in syntactic analysis of AI-authored texts, to date, such a systematic and quantitative comparison between AI-generated texts and human-written texts is still lacking. While previous research has addressed fluency and grammaticality separately very little is known about a complete syntactic comparison based on both complexity and pattern usage. This is what we attempt to address in our work by taking a double perspective: The patterns of syntactic constructions and the complexity measures of sentence structure. In addition, to bridge between qualitative, data-driven and theoretical analyses, the study also aims to inform theoretical development and practical curriculum design in AI system development and educational contexts by way of the association between qualitative interpretation and quantitative measures.

Research Methodology:

The present study uses a comparative-descriptive design to investigate the syntactic constructions in human writers' texts and AI writing tool (ChatGPT). The study analyses 18 short essays written by both Humans and an AI model ranging in between 150-200 words. Nine written by Human and 9 samples generated by an AI-generated language model (ChatGPT) on the random prompts ranging in topics from technology to social media. This method provides thematic consistency and reduces topic bias. The study applies both random sampling and manual and computational analysis — that is, manual analysis of syntax tree (for grammatical hierarchies) and computational metrics (mean sentence length, T-unit length, indexes of coordination and subordination), made with Coh-Metrix tool.

The data analysis brings quantitative and qualitative methods together and presents qualitative analyses about error types, syntax awkwardness, and fluency by AI texts. Error analysis is supported by the use of tools such as the Grammarly logs and manual coding, and a fine-grained analysis of grammaticality is performed. Ethical considerations were met by obtaining consent from participants, de-identifying subjects in order to maintain confidentiality and by generating texts using AI technology only for research purpose. The study further acknowledges there are still limitations on this study such as the text length being a little bit short, the subjectivism in the manual assessments, or that the results from the proposed AI tool may not be generalizable through all the AI-generated writing.

Findings

Table 1: *Syntactic Complexity Metrics*

Group	Words	Sentences	Mean Sentence Length	Clauses	Clause-to-Sentence Ratio	T-unit Length	Subordination Index
AI	203	10	20.3	23	2.30	20.3	0.80
AI	211	11	19.2	24	2.18	19.2	0.73

Group	Words	Sentences	Mean Sentence Length	Clauses	Clause-to-Sentence Ratio	T-unit Length	Subordination Index
AI	181	9	20.1	20	2.22	20.1	0.56
AI	226	10	22.6	25	2.50	22.6	0.90
AI	233	11	21.2	24	2.18	21.2	0.64
AI	227	11	20.6	23	2.09	20.6	0.55
AI	229	10	22.9	26	2.60	22.9	0.90
AI	237	11	21.5	25	2.27	21.5	0.73
AI	231	11	21.0	24	2.18	21.0	0.64
H	176	9	19.6	17	1.89	18.4	0.48
H	140	8	17.5	13	1.63	17.0	0.40
H	160	9	17.8	14	1.56	17.3	0.44
H	165	10	16.5	15	1.50	16.1	0.38
H	170	9	18.9	16	1.78	18.0	0.43
H	205	12	17.1	18	1.50	16.9	0.41
H	185	10	18.5	17	1.70	18.0	0.45
H	190	11	17.3	18	1.64	17.2	0.42
H	180	10	18.0	16	1.60	17.5	0.40

[Data retrieved from Coh-Metrix analysis and organized manually by the researcher.]

The syntactic complexity analysis reveals an observable difference between AI-generated text and that generated by Pakistani ESL Learners. Versus expert essays, AI essays by persistently showed higher values, according to almost all complexity metrics. In particular, the ratio was higher than that of the impedance samples, ranging from 2.09 to 2.6, and a longer average sentence length (up to 22.9 words), that reflected the higher frequency of multi-clausal and well-structured sentences. In addition, T-unit length and subordination index in AI essays were relatively longer (Table 1), indicative of higher uses of embedding and subordination structures.

The human-written essays however, which were written by 5th-semester Pakistani ESL learners, showed mild structural complexity to sentences; that is, there were smaller number of exceptions per sentence, few clauses in sentence, and less subordination index. Although the human writers made try constructions more complex, their sentence structures were closer to basic/intermediate levels of syntactic development. The proportion of clauses to sentences in these essays, whilst not disastrously low, does point to a preference for straightforward construction, meaning that there are far fewer sub-ordinate or co-ordinate clauses/structures. Overall, the syntactic evidence indicates that the AI system's writing is characterized by greater control over syntactic variation and complexity compared to human learners, who -although able- are still working towards acquiring control over complex sentence structures and use of subordinated clauses.

Table 2: *Error and Fluency Analysis*

Group	Grammar Errors	Syntax Errors	Repetition	Awkward Phrasing	Overall Fluency (1–5)
AI	1	0	1	0	4.8
AI	0	0	1	0	5.0
AI	1	0	1	1	4.5
AI	0	0	1	0	4.9
AI	0	0	1	0	4.9

Group	Grammar Errors	Syntax Errors	Repetition	Awkward Phrasing	Overall Fluency (1–5)
AI	0	0	1	0	4.8
AI	0	0	1	0	5.0
AI	0	0	1	0	4.9
AI	0	0	1	0	5.0
H	3	1	2	1	4.2
H	2	1	2	1	4.1
H	3	2	1	2	3.9
H	2	1	1	1	4.0
H	2	0	2	1	4.1
H	3	1	2	0	4.2
H	2	1	2	1	4.0
H	3	2	1	2	3.8
H	2	1	1	1	4.0

[Data retrieved from Coh-Metrix analysis and organized manually by the researcher.]

The error and fluency analysis further points out the difference between AI-generated essays and the essays generated by the Pakistani ESL learners. Auto-generated AI texts were nearly fluently (on average 4.89 out of 5) with no obviously awkward phrasing, minimal grammatical and structural issues. The sole common design feature among the AI texts was controlled repetition, which was likely more rhetorical emphasis or topic arousing than a linguistic constraint.

The human essays, on the other hand, had slightly less fluency averaging between 4.5 and 4.8, and occasional grammatical errors, awkward sentences, and more disparate levels of flow and pair. Although the samples were free of syntax errors (thereby displaying elementary syntactic control), a number of samples revealed occasional grammatical lapses and a lack of consistency in the expression of meaning. Furthermore, although repetition was also present in a number of human essays, for that group it was sometimes a signal of restricted variation of vocabulary rather than emphatic coloration. These findings indicate that though Pakistani learners writing at the tertiary level is quite fluent, it betrays developmental problems in grammar, and phrasing, when compared with polished and mechanically accurate output of AI. This emphasizes the need for ongoing grammar instruction and fluency-enhancement activities to overcome the foregoing problems.

Table 3: *Syntax Tree Pattern Analysis*

Group	Simple	Compound	Complex	Compound-Complex	Notes on Clause Type & Variation
AI	3	2	3	2	Mix of clause types, frequent use of purpose/conditional clauses
AI	2	3	4	2	Balanced mix; frequent use of relative and result clauses
AI	3	3	2	1	Strong coordination; some ellipsis and nominal clause embedding
AI	2	2	4	2	Heavy use of cause-effect, relative, and conditional clauses
AI	2	3	4	2	Frequent use of relative, contrastive, and purposive subordination

Group	Simple	Compound	Complex	Compound-Complex	Notes on Clause Type & Variation
AI	3	3	3	2	Balanced pattern with coordination; some use of conditional and explanatory clauses
AI	2	2	4	2	Strong use of explanatory, contrastive, and concessive subordination
AI	2	3	4	2	Predominant use of conditional and causal clauses, especially in contrasts and conclusions
AI	3	2	4	2	Uses contrastive, explanatory, and moral-evaluative clause structures effectively
H	4	2	2	1	Simple dominant; limited variety, mostly basic adverbial clauses
H	3	3	2	1	Balanced coordination; few dependent clauses used effectively
H	4	2	2	1	Reliance on coordination; little subordination or embedding
H	3	2	3	1	Some explanatory and relative clauses; lacks consistent depth
H	3	3	2	1	Limited clause variety; mostly coordination with some subordination
H	3	2	3	1	Explanatory and conditional clauses attempted; inconsistent usage

Group	Simple	Compound	Complex	Compound-Complex	Notes on Clause Type & Variation
H	2	3	3	1	Compound bias; complex ideas underdeveloped
H	3	2	2	1	Repetitive use of coordination; subordination minimal
H	3	3	2	1	Some attempt at clause diversity; unclear clause hierarchy

[Data analyzed manually by the researcher based on structural observation.]

The parsed pattern analysis of syntax tree shows the significant differences in syntax structure and variety of clause between AI-written essays and human-written essays. AI essays were syntactically robust, producing a similar proportion of simple, compound, complex, and compound-complex sentences. In the nine AI texts, subordinated clauses in various types of relative, conditional, causal, explanatory and contrast were strategically and often deployed. Such a complex embedding of clause as this reflects a higher level of logical and rhetorical relationship inside discourse.

The essays of the human participants, on the other hand, were syntactically diverse but demonstrated a less consistent control of complex sentence forms. Human essays used more compounds and simple formations and significantly fewer complex and compound-complex structures than AI texts. Clause variability was found, although it tended to be simpler or less diverse, with relative and adverbial clauses dominating. Clause embedding was observed, but only with simple types of explanations or conditions, revealing that the syntactical repertoire was emerging, yet still restricted. These trends show that Pakistani ESL learners can build coherent structure up to a certain degree of complexity, but they fall short of AI in terms of clause embedding and syntactic variety.

Discussion

The examination of AI generated and human authored essays provides profound insights into the expertise of syntactic variety, linguistic proficiency, and grammatical sensitivity of Pakistani ESL learners. By analyzing the essays along three dimensions of analysis (syntactic complexity, error and fluency patterns, and syntactic variation), this study allows for a comparison of both of groups of speakers' linguistic performance. Table 1: Syntactic Complexity Metrics Table showed that both AI- and human-written texts both have a high-level of syntactic development, as AI-written essays consistently outperform human-written essays in most metrics. AI essays had more clauses and more balanced clause-to-sentence ratios, which may indicate better control of sentence embedding and syntactic layer. Human-written essays, despite having a same degree of structural diversity, show considerably smaller clause-per-sentence ratios and subordination indices.

This finding is consistent with previous studies where it is claimed that ESL learners such as the participants in the present study are able to produce complex sentences but tend to make use of simpler structures to ensure that they are clear and correct (Lu, 2010). The human essays exhibited natural variation, sometimes inconsistent in clause choice, and generally more conservative in syntactic expansion. The differentiation was also confirmed by the Error and Fluency Analysis Table. The AI-generated essays were presented with an almost perfect fluency and correctness, and few error cases were noted. On the other hand, although human written texts in general were more fluent, they had occasional but remarkable instances of grammatical mistakes, awkward language and slightly worse fluency scores. These results are consistent with issues that are known to characterize ESL individuals in higher education who may still be developing awareness of morpho-syntactic features and academic style (Ferris, 2009). While the majority of students were strong in writing, the fact that there are errors/ lack of consistent style from time to time suggests the need for continued support from instruction in grammar and self-editing.

Finally, the Syntax Tree Pattern Analysis Table analysis table showed more about the variation of sentence structure and use of clauses. AI texts also showed a remarkable level of mastery in the combination of clause types for specific purposes—from relative and conditional to causal, concessive, and explanatory—forming a coherent and rhetorically effective discourse.

Human essays, on the other hand, structurally quite cohesive, include larger degree of compound (and simple) sentences was more widespread while more complicated subordinating structures were less used. This correlates with a developmental feature and handwriting native of the students' writing at this stage where structural proficiency begins to be established though not yet differentiated. The underuse of complex-compound sentences and inserted clauses may reflect a lack of proficiency in expressing relationships between ideas, a major feature of academic and persuasive writing.

In a nutshell, it can be said that these results indicate that Pakistani ESL learners have a basic control of English syntax and fluency, however, they lack proficiency in advanced writing features regarding syntactic variation, clause complexity, and grammatical precision. In the case of AI-authored texts, the tent-pole is that of the ideal explication: coherent, fluent, and stylistically accomplished.

Conclusion

This research aimed to compare and contrast the syntactic complexity, grammatical accuracy and structural variety of AI and Pakistani ESL learners' essays in a sample of Pakistani ESL learners studying in the fifth semester of a BS English program. Through investigating three nucleus components – lexical, error, and syntax tree – the study intended to both demonstrate the potentials of AI-produced writing, and to evaluate tertiary ESL learners' current level of proficiency and areas of development. The results showed that AI-produced texts were characterized by a consistently greater degree of syntactic complexity, as indicated by longer sentence and T-unit lengths, higher densities of clauses, and more elaborate subordination. AI writings also possessed excellent grammatical accuracy and near-native fluency, and contained few, if any, errors and unnatural sentences. Human-written essays, on the other hand, tended to have moderate complexity and remained relatively well-structured, though they exhibited some grammatical and lexical issues. The use of syntax patterns by students demonstrated a preference for less complex sentences both in advanced subordination and embedded clause types.

These results indicate that Pakistani ESL users have a robust frame in formal writing but still need to acquire finer control of sentence patterns and editing for accuracy and smoothness. Yet the comparison with AI-authored writing also underscores both the possibilities and the

shortcomings of human performance, thus providing invaluable feedback on where instructional emphasis can be fed up—specifically, in promoting more structural variety, more accuracy in grammatical forms, and greater confidence in risk-taking with complex structures. In essence, this type of research adds to the literature on ESL academic writing by benchmarking learner production, with the goal of using such benchmarks to inform and improve pedagogy. Future research could consider the influence on using AI-generated samples for learning models or examine the longitudinal development of student writing with exposure to targeted syntactic instruction.

References

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Biber, D. (1991). *Variation across speech and writing*. Cambridge University press.

Chua, T. S., Fuhr, N., Grefenstette, G., Järvelin, K., & Peltonen, J. (2018). User-generated content in social media (Dagstuhl Seminar 17301). *Dagstuhl Reports, 7*(7).

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Klein, D., & Manning, C. D. (2003, July). Accurate unlexicalized parsing. In *Proceedings of the 41st annual meeting of the association for computational linguistics* (pp. 423-430).

Lu, X. (2010). Automatic analysis of syntactic complexity in second language writing. *International journal of corpus linguistics*, *15*(4), 474-496.

Martin, J. H., & Jurafsky, D. (2019). Vector Semantics and Embeddings. *Speech Lang. Process*, 1-31.

McDonald, R., Lerman, K., & Pereira, F. (2006, June). Multilingual dependency analysis with a two-stage discriminative parser. In *Proceedings of the Tenth Conference on Computational Natural Language Learning (CoNLL-X)* (pp. 216-220).

Narayan, S., Zhao, Y., Maynez, J., Simões, G., Nikolaev, V., & McDonald, R. (2021). Planning with learned entity prompts for abstractive summarization. *Transactions of the Association for Computational Linguistics*, *9*, 1475-1492.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.