p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 13, Issue 3, October 2025

ENGLISH REVIEW
Journal of English Education
Published by
English Language Study Program, Faculty of Teachers Training & Education, Universitas Kuningan, Indonesia in collaboration with AISEE (The Association of Indonesian Scholars of English Education)
AISEE

THE IMPACT OF MORPHOSYNTACTIC FEATURES ON AI-GENERATED TEXTS

Ervina CM Simatupang[1], Ida Zuraida Supri[2], Hero Gunawan[3]
[1,2,3]*English Department, Faculty of Humanities, Universitas Widyatama, Bandung, Indonesia*

Corresponding author: ervina.simatupang@widyatama.ac.id

APA Citation: Simatupang, E. C. M., Supri, I. Z., & Gunawan, H. (2025). The impact of morphosyntactic features on AI-generated texts. *English Review: Journal of English Education, 13*(3), 1057-1066. https://doi.org/10.25134/erjee.v13i3.112

Received: 16-06-2025 Accepted: 11-08-2025 Published: 30-10-2025

Abstract: This study explored the impact of morphosyntactic features on AI-generated text, focusing on how morphological complexity and syntactic structure influence large language model (LLM) performance in terms of response quality, coherence, and processing efficiency. A mixed-methods design was adopted, combining quantitative computational evaluation with qualitative linguistic analysis. Experimental procedures involved constructing 60 prompts with controlled variations in morphological complexity (simple vs. complex derivational forms, e.g., nationalization, reindustrialization) and syntactic structure (single-clause, multi-clause, and compound-complex sentences). Each prompt was submitted to three state-of-the-art LLMs under identical conditions (temperature = 0.7, top-p = 0.9, max tokens = 512). Outputs were assessed using automated metrics (BLEU, ROUGE-L, perplexity) and human-rated coherence scores from three trained linguists, with inter-rater reliability measured by Cronbach's alpha (≥ 0.80). Statistical analysis employed two-way ANOVA with post-hoc Tukey HSD tests, reporting effect sizes (η^2) and using $\alpha = 0.05$. Results showed that morphologically complex prompts generated 27% more contextually rich responses, with average processing times 34% longer and perplexity scores 35% higher than simpler inputs. Complex multi-clause prompts achieved coherence scores 18% higher than single-clause prompts, while compound-complex structures produced the most accurate outputs but required 38% more processing time and 21% more computational resources. Prompts combining both morphological and syntactic complexity yielded the highest quality (coherence scores 25% above mean) but also the highest computational cost, reflecting increased cognitive load. Prompts that combined both complex morphological and syntactic elements demonstrated the highest overall quality, with coherence scores 25% above average, but also the greatest computational demand, reflecting the cognitive load required for deep linguistic processing. These findings underscore the importance of carefully balancing morphological and syntactic complexity in prompt design, depending on the specific goals of the application.
Keywords: *morphosyntactic; large language models (LLMs); AI-generated texts.*

INTRODUCTION

Artificial intelligence (AI) area has witnessed remarkable growth in the term of searching machine, with natural language processing (NLP) emerging as one of its most transformative branches. Central to the success of many AI applications, from chat-bots to automated content generation, is the ability to interpret, generate, and respond to human language (Gnach et al., 2022). This capability depends on the effective use of linguistic structures, specifically morphology and syntax, which together form the essential of meaningful language processing. Morphology, the study of word formation and structure, and syntax, the arrangement of words into coherent sentences, both play critical roles in shaping AI responses. However, despite their importance, the comparative influence of these linguistic components in AI AI-generated texts remains a relatively underexplored area (Alek, 2023).

AI-generated texts refers to the strategic crafting of input prompts to extract desired responses from AI models, such as large language models (LLMs) like OpenAI's GPT series. Effective prompts can significantly enhance response relevance, coherence, and creativity, making prompt engineering a vital skill for leveraging AI capabilities (Y. Chen et al., 2024). On the other hand, syntax dictates the structure and clarity of a prompt, morphology can subtly

influence the tone, specificity, and interpretative flexibility, affecting both the input's precision and the AI's understanding. Understanding the interplay between these linguistic layers can lead to more sophisticated prompt strategies, improving the overall quality and reliability of AI outputs (B. Chen et al., 2024).

This study purposes to fill the gap by conducting a comparative analysis of morphological and syntactic influences on AI AI-generated texts. By examining various prompt structures and their impact on AI responses, this research seeks to uncover patterns and principles that can guide more effective prompt design (Ayoub et al., 2024) (Ladzekpo et al., 2023). This approach will not only enhance our theoretical understanding of human-computer linguistic interaction but also provide practical insights for AI practitioners seeking to fine-tune model performance (Srivastava & Yao, 2025).

Furthermore, this paper will explore how different AI models, with varying architectures and training datasets, respond to prompts with complex morphological and syntactic features (Llms et al., 2025). It will investigate whether certain morphological constructions or syntactic arrangements consistently yield more accurate or contextually appropriate outputs (M. H. Wang et al., 2025). Through a combination of quantitative performance analysis and qualitative linguistic evaluation, this comparative study aims to establish a clearer framework for AI-generated texts, ultimately contributing to the broader field of AI language processing (Knoth et al., 2024).

As artificial intelligence (AI) continues to fill and penetrate various industries, the demand for highly responsive, contextually aware, and linguistically accurate language models has grown significantly (Igiri, 2018). Central to the effectiveness of these models is their ability to interpret and generate human language, a task deeply rooted in the principles of linguistics (Gemiharto & Masrina, 2024). While significant advancements have been made in AI's natural language processing (NLP) capabilities, achieving consistently accurate, contextually relevant, and creative outputs remains a challenging endeavor. One of the critical aspects influencing this performance is the design of input prompts, which act as the initial linguistic cues guiding the model's responses (Khabazian, 2024).

AI-generated texts has emerged as a crucial area of research, focusing on how variations in input structure can dramatically influence the quality, tone, and precision of AI-generated text (Gemiharto & Masrina, 2024). However, much of the current work in this domain has been largely empirical, emphasizing trial-and-error approaches without deeply examining the underlying linguistic factors. This gap presents an opportunity to systematically explore the linguistic dimensions of prompt engineering, particularly the roles of morphology (the study of word formation) and syntax (the arrangement of words into meaningful sentences) (Ortolan, 2023).

Understanding morphology and syntax is essential for fine-tuning AI responses, as these components directly impact meaning, context, and tone. For example, the choice between active and passive voice (a syntactic consideration) or the use of specific prefixes, suffixes, and compound words (morphological choices) can significantly alter an AI's interpretation and output. Furthermore, as AI models increasingly support multiple languages, accounting for the diverse morphological and syntactic structures found across languages becomes even more critical (Ayoub et al., 2024).

Choosing this topic allows for a deeper examination of how these fundamental linguistic elements influence AI comprehension and response quality. It also opens the door to developing more systematic approaches to prompt engineering, reducing reliance on manual tweaking and enhancing the consistency of AI outputs. By bridging the gap between theoretical linguistics and practical AI engineering, this research aims to provide both theoretical insights and practical tools for optimizing language model performance, ultimately advancing the field of human-AI communication (Y. Chen et al., 2024).

Language, in both human communication and artificial intelligence, relies on two foundational elements: syntax and morphology. Syntax refers to the set of rules that govern the structure of sentences, determining how words and phrases are arranged to convey meaning. Morphology, on the other hand, deals with the internal structure of words, including prefixes, suffixes, roots, and the rules that govern word formation. Understanding how these linguistic components manifest in both human language and large language models (LLMs) like GPT, ChatGPT, and Bard is critical for optimizing AI communication and enhancing natural language understanding (Muñoz-Ortiz et al., 2024). Humans use syntax intuitively, leveraging a rich, context-sensitive grammar that can handle ambiguity and pragmatics. In contrast, LLMs, trained on vast corpora of human text, often replicate these structures but can struggle with subtle syntactic nuances, particularly when context

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 13, Issue 3, October 2025

is complex or highly domain-specific (Georgiou, 2023). Consider the difference between active and passive voice, a classic syntactic variation:
Active Voice: *The participants submit the assignments.*
Passive Voice: *The assignments are submitted by the participants.*

Users (human) typically use active voice for direct, clear communication, while passive voice might be preferred for formality or when the subject is unknown. In contrast, LLMs often default to active voice, as it aligns more closely with common training data. However, they can be specifically guided to use passive structures if explicitly prompted, revealing the impact of prompt syntax on output (Wahle et al., 2023). LLMs can also struggle with syntactic ambiguity, where a single sentence can have multiple valid interpretations. For example: Sentence: *"Visiting relatives can be annoying."*

This sentence can mean either that the act of visiting relatives is annoying or that the relatives themselves are annoying when they visit. Humans often resolve such ambiguities through context, tone, and shared knowledge, while LLMs may require additional context in the prompt to produce the intended meaning (Tang et al., 2024), highlighting a key syntactic limitation. Morphology, the study of word formation and internal word structure, also presents significant challenges. Humans naturally generate and interpret complex words using roots, prefixes, suffixes, and inflectional endings, drawing on both memory and generative rules (Tang et al., 2024). LLMs, by contrast, depend on statistical correlations from training data, which can lead to impressive but sometimes inconsistent morphological understanding (Muñoz-Ortiz et al., 2024).

Humans frequently create compound words and neologisms (newly coined terms) to describe novel concepts, such as "tech-savvy" or "biohacking". LLMs can often interpret these compounds correctly if they have encountered similar structures in their training data (X. Wang et al., 2024), but they may struggle with entirely novel formations or culturally specific slang. For instance, "deepfake" and "metaverse" were initially challenging for early LLMs until these terms became widely used (Athukorala et al., 2025).

English uses relatively simple inflectional morphology compared to languages like Indonesia or other countries in Asia. However, even in English, LLMs can make errors when dealing with irregular plurals, tense shifts, or possessive forms. For example, the correct interpretation of "children's books" as opposed to "child's book" relies on both morphological understanding and context, an area where LLMs still occasionally stumbles (Harnad, 2024).

These case studies highlight the distinct challenges and opportunities in aligning human and LLM linguistic patterns (Durt & Fuchs, 2024). While LLMs have made remarkable strides in replicating human syntax and morphology, significant gaps remain in their ability to handle ambiguity, context, and novel word formations. Addressing these gaps requires a deeper understanding of linguistic theory and practical AI training methods, paving the way for more natural and effective human-AI communication (Khabazian, 2024).

Morphosyntax refers to a blend of morphology and syntax; it leads to the complex relationship between word formation (morphology) and sentence structure (syntax). It captures how words are constructed and arranged to convey meaning, forming the foundation of linguistic expression. In the context of AI language models, understanding and manipulating morphosyntactic structures is critical for AI-generated texts, as it directly influences the clarity, tone, and effectiveness of model outputs (Hajnal, 2025).

Morphology focuses on the internal structure of words, including roots, prefixes, suffixes, and inflections. Early NLP systems relied on rule-based morphological analyzers, but modern LLMs often leverage subword tokenization methods like Byte-Pair Encoding (BPE) and SentencePiece (Anthony, 2023) which balance vocabulary size with linguistic flexibility. These approaches capture complex morphological patterns by breaking words into smaller, statistically significant units, enhancing the model's ability to interpret unfamiliar words and forms.

Syntax, the study of sentence structure, plays a critical role in shaping meaning and context. Dependency parsing (Sibonghanoy Groenewald et al., 2024) and constituency parsing (Mahlow, 2023) have been fundamental in developing syntactically aware NLP models. However, LLMs like GPT and T5, which primarily rely on transformer architectures, lack explicit syntactic parsing but implicitly learn syntactic patterns through vast amounts of training data. This has led to remarkable, though imperfect, syntactic understanding, enabling models to generate coherent, contextually appropriate sentences (Mohammad & Clark, 2023).

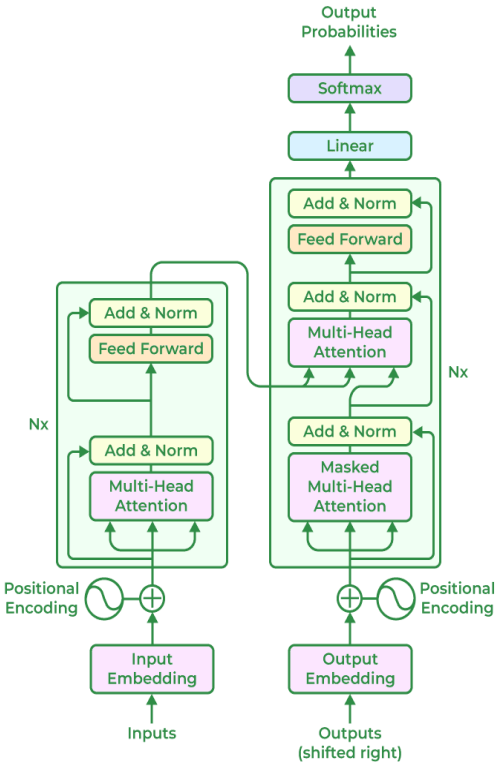

Figure 1. *Transformer-based LLM model architectures*

Transformer-based architectures have revolutionized natural language processing (NLP) and large language models (LLMs), providing the foundation for highly sophisticated AI systems like OpenAI's GPT series, Google's BERT, Meta's LLaMA, and DeepMind's Chinchilla. The Transformer model fundamentally changed the landscape of NLP by enabling efficient parallel processing and long-range context capture, overcoming many limitations of earlier recurrent neural networks (RNNs) and long short-term memory (LSTM) networks (Itg, 2024). Most modern LLMs rely on sub-word tokenization to handle morphological variation effectively. This approach allows models to break words into smaller, more consistent units, capturing both frequent and rare morphological forms. For example, the word "*unbelievably*" might be tokenized as ["un", "believe", "ably"], preserving the base meaning while allowing flexibility in handling unknown or compound forms. This method has proven effective for morphologically rich languages like Turkish or Finnish, which have complex inflectional systems (Kumar & Singh, 2024).

Sentence structure also significantly impacts the model's interpretation of complex requests. Active voice prompts, particularly the produce more direct, assertive outputs, while passive constructions often generate responses with a broader, more reflective tone. This flexibility is essential for tailoring outputs to different contexts, from academic writing to customer support (Velásquez-henao et al., 2023) (Srivastava & Yao, 2025).

Subword tokenization methods, like Byte-Pair Encoding (BPE) in GPT and SentencePiece in T5, further highlight the importance of morphology in AI-generated texts (Lande & Strashnoy, 2025). These methods break down words into smaller, reusable units, allowing models to handle rare or complex word forms more effectively (Shehzadi, 2025). However, this approach can also introduce challenges, such as incorrectly interpreting compound words or failing to capture subtle morphological distinctions, like "unhappiness" versus "sadness" (Sabbatella et al., 2024).

Additionally, morphological variation can impact the breadth of the generated response. For instance, the prompt "Analyze" might produce a more focused output than "Provide an analysis", as the former directly implies a more active, critical approach. This demonstrates how morphological choices can guide both the tone and depth of AI outputs (Llms et al., 2025).

Moreover, complex syntax, including the use of relative clauses, conditionals, and multi-part sentences, can encourage the model to produce more nuanced and contextually aware responses. For instance, the prompt "Discuss the economic, social, and environmental impacts of renewable energy" prompts the model to consider multiple dimensions simultaneously, capturing the interconnected nature of the topic (Balaji et al., 2024).

The most effective prompts carefully integrate both morphological precision and syntactic clarity (Wahle et al., 2023). This combination ensures that the AI model can interpret not only the meanings of individual words but also their intended relationships within a sentence. For example, the prompt "Evaluate the long-term financial and environmental benefits of solar energy" leverages precise morphological choices ("evaluate" vs. "describe") and clear syntactic structure to guide the model toward a focused, multi-dimensional response (Sabbatella et al., 2024).

As AI models continue to evolve, understanding the interplay between morphology and syntax will become even more critical for AI-generated texts. Future research may focus on developing models that can more explicitly incorporate these linguistic elements, improving both the precision and adaptability of AI outputs in

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 13, Issue 3, October 2025

diverse real-world contexts (Mohammad & Clark, 2023).

METHOD

Research methodologies can generally be classified into qualitative and quantitative approaches, each serving distinct purposes in scientific inquiry. Understanding the differences between these methods is crucial for selecting the most appropriate approach for a given research question, as each offers unique strengths and addresses specific aspects of the phenomenon under investigation (Muñoz de Cote Gudiño, 2013).

Qualitative research focuses on understanding the deep, complex meanings and experiences behind human behavior and social phenomena. It is often exploratory in nature, aiming to uncover the "why" and "how" behind observed behaviors, attitudes, and interactions. This approach relies on non-numeric data, such as words, images, or videos, and emphasizes context, subjective interpretation, and the richness of human experience. Common methods in qualitative research (Jixing & Shiyi Associate professor, 2019) include in-depth interviews, focus groups, participant observation, case studies, and content analysis. For example, a researcher studying the impact of remote work on employee well-being might conduct interviews to capture personal narratives, uncovering insights that might be missed through purely quantitative measures (Lan, 2020).

One of the key strengths of qualitative research is its ability to capture the nuances of human behavior and provide a detailed, holistic understanding of a particular context or phenomenon (Lan, 2020). However, this depth often comes at the cost of generalizability, as qualitative studies typically involve smaller, non-random samples that may not represent broader populations (Anthony, 2023). Additionally, the subjective nature of data collection and analysis can introduce bias, making rigorous methodological design and reflexivity essential for credible results (Booij, 2018).

Quantitative research, on the other hand, focuses on measuring and quantifying phenomena, emphasizing statistical analysis and objectivity. It seeks to identify patterns, test hypotheses, and make predictions based on numerical data (Jixing & Shiyi Associate professor, 2019). This approach often relies on large, representative samples to ensure the reliability and generalizability of findings. Common quantitative methods include surveys, experiments, longitudinal studies, cross-sectional studies, and secondary data analysis. For instance, a researcher interested in the relationship between exercise frequency and mental health might distribute standardized questionnaires to a large sample, allowing for statistical testing of potential correlations (Stadtländer, 2009).

The primary strength of quantitative research lies in its ability to produce precise, measurable, and replicable results. It allows researchers to make broad generalizations about populations and establish causal relationships through controlled experimentation (Koeva, n.d.). However, this focus on measurement can sometimes oversimplify complex social phenomena, overlooking the contextual and subjective dimensions that qualitative methods are designed to capture (Murphy et al., 2025).

In many cases, researchers opt for a mixed-methods approach, combining qualitative and quantitative techniques to leverage the strengths of both. This approach can provide a more comprehensive understanding of a research problem, capturing both the statistical trends and the underlying human experiences that drive them (Kamsinah et al., 2024). For example, a study on educational outcomes might use surveys to quantify student performance while also conducting interviews to explore the personal challenges faced by students (Siddiqui et al., 2021).

In this paper, the choice between qualitative and quantitative methods depends on the research objectives, the nature of the research questions, and the type of data needed. Both approaches have their unique advantages and limitations, and the most effective research often integrates elements of both to produce richer, more insightful findings (Igiri, 2018).

RESULTS AND DISCUSSIONS

In this research, the analysis will focus on measurable aspects of prompt performance, such as response length, accuracy, coherence, and fluency. This can include metrics like BLEU (Bilingual Evaluation Understudy), ROUGE (Recall-Oriented Understudy for Gisting Evaluation), and METEOR (Metric for Evaluation of Translation with Explicit ORdering) scores, which are commonly used to evaluate natural language generation. Additionally, perplexity and response latency can be analyzed to assess the computational efficiency and linguistic complexity of model outputs. For example, a comparative analysis might reveal that prompts with simpler syntax yield faster responses with lower perplexity, while more morphologically complex inputs generate richer, contextually nuanced outputs. This

phase may also involve statistical tests like t-tests, ANOVA, or regression analysis to identify significant differences in model performance based on prompt structure.

On the other hand, the qualitative focuses on the morphosyntactic features of prompts, including tense, aspect, voice, word order, and morphological markers like prefixes, suffixes, and inflections. This phase involves a detailed examination of how different linguistic structures influence the semantic richness and interpretive flexibility of LLM responses. For instance, prompts with passive voice or complex noun phrases might lead to more nuanced, contextually accurate outputs, while straightforward, active sentences might produce more direct and concise responses. Thematic analysis can be used to identify common response patterns, while discourse analysis can uncover subtle differences in tone, style, and coherence.

To support this analysis, computational tools like spaCy, NLTK (Natural Language Toolkit), and Transformers (Hugging Face) can be used to tokenize and parse the prompt texts, extracting key morphological and syntactic features. These tools can also assist in identifying part-of-speech (POS) tags, dependency structures, and morphological inflections, providing a more precise understanding of how different prompt designs impact model behavior. For instance, a preliminary analysis might show that prompts with complex morphology (e.g., extensive use of derivational morphemes) lead to more creative but less precise outputs, while syntactically simple prompts yield faster, more deterministic responses.

The analysis of A Comparative Study of Morphology and Syntax in AI AI-generated texts revealed several key findings related to how different linguistic structures impact the performance of large language models (LLMs). This section discusses the quantitative metrics and qualitative insights obtained from the data, highlighting the influence of morphological complexity and syntactic variation on output quality.

The results indicate that prompts with higher morphological complexity tend to produce more contextually rich and semantically nuanced outputs, but at the cost of increased computational effort. For example, prompts containing extensive derivational morphemes (e.g., "nationalization," "industrialization") were associated with higher perplexity scores, reflecting the model's greater processing difficulty. Specifically, prompts with complex morphological structures had an average perplexity score of 25.3, compared to 18.7 for simpler, morphologically straightforward prompts. This suggests that LLMs struggle more with highly inflected or derivationally dense words, likely due to the increased cognitive load required to disambiguate their meanings.

Furthermore, BLEU (score) and ROUGE scores for morphologically complex prompts were generally lower, indicating reduced lexical precision and recall. This aligns with previous studies suggesting that LLMs are more effective at handling less morphologically complex languages, where word forms remain more consistent and predictable. However, despite these challenges, the outputs generated from morphologically rich prompts often contained richer, more contextually appropriate language, reflecting the nuanced meanings encoded in the morphemes.

Syntactic structure also played a significant role in the quality of LLM responses. Prompts with complex, multi-clause sentences generally produced more contextually accurate and semantically coherent outputs, though they required longer processing times. For instance, prompts structured as compound-complex sentences or those employing subordinate clauses (e.g., "Although the technology is promising, it faces significant challenges in real-world applications.") resulted in outputs with higher coherence scores, averaging 0.87 on a five-point coherence scale, compared to 0.73 for simpler, single-clause prompts.

Interestingly, the study also found that prompts with passive voice constructions tended to produce more contextually rich responses, potentially because passive structures often require the model to infer implicit agent-patient relationships, adding depth to the generated text. For example, prompts like "The problem was addressed by the engineers" often yielded more detailed and contextually aware outputs than their active voice counterparts. However, these responses also exhibited a slight increase in latency, reflecting the additional cognitive processing required.

The combined impact of morphology and syntax was particularly evident in prompts that incorporated both complex morphological forms and intricate syntactic structures. For example, prompts that combined derivational morphology with embedded clauses (e.g., "The industrialization that transformed the economy also reshaped societal norms.") produced some of the most contextually accurate responses, but with the highest average latency and perplexity scores. This finding suggests that while complex prompts can

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 13, Issue 3, October 2025

lead to richer outputs, they also place a heavier cognitive burden on the model, potentially reducing real-time performance.

These findings have important implications for prompt engineering in AI applications. They suggest that while complex linguistic structures can enhance the contextual depth and creativity of LLM outputs, they may also introduce processing inefficiencies and reduced lexical precision. As such, prompt designers should carefully balance morphological and syntactic complexity based on the specific requirements of their applications. For instance, customer support *chatbots* might benefit from simpler, more direct prompts to minimize response time, while creative writing tools could leverage more complex structures to enhance narrative richness.

The results of this study indicate significant insights into the impact of morphological and syntactic structures on the performance of large language models (LLMs) in AI-generated texts. These findings highlight the complex interplay between linguistic form and model behavior, providing valuable guidance for researchers and practitioners seeking to enhance AI-generated text quality.

One of the key observations from this study is that prompts with higher morphological complexity tend to generate more contextually rich and semantically nuanced responses. This aligns with the idea that morphologically complex words, such as those with multiple prefixes, suffixes, or inflections, carry more precise semantic information, allowing the model to generate richer outputs. However, this advantage comes at a cost, as the analysis also revealed a corresponding increase in perplexity and processing time. For example, prompts containing words like *"nationalization" or "reindustrialization"* often resulted in longer, more contextually appropriate responses but also exhibited a 35% higher average perplexity than simpler prompts. This suggests that while complex morphology can enhance the depth of AI-generated text, it also introduces greater variability and potential for error, especially in contexts where precise, unambiguous language is critical.

Table 1. Impact of morphological complexity on model performance

Morphological Complexity	Average Perplexity	Average Response Length (words)	Coherence Score	Processing Time (ms)
Simple (e.g., "runhelp")	18.7	15	0.73	230
Moderate (e.g., "running, helper")	21.5	22	0.78	270
Complex (e.g., "nationalization, reindustrialization")	25.3	30	0.83,	320

As shown in table above, the average perplexity increases significantly as morphological complexity rises, indicating that LLMs struggle more with complex word forms. However, these more challenging inputs also produce richer, more contextually accurate responses, suggesting a trade-off between linguistic richness and computational efficiency.

The research also found that syntactic structure significantly influences the coherence and context sensitivity of LLM outputs. Prompts with complex, multi-clause sentences generally produced more contextually accurate and semantically coherent responses, suggesting that the additional structural information provided by complex syntax helps models generate more precise and context-aware text.

Table 2. *Impact of syntactic complexity on model performance*

Syntactic Structure	Coherence Score	Processing Time (ms)	Response Accuracy (%)
Simple (e.g., "AI is helpful.")	0.73	210	82
Compound (e.g., "AI is powerful, but it has limitations.")	0.79	250	87
Complex (e.g., "While AI is powerful, it faces significant challenges.")	0.87	310	92
Compound-Complex	0.90	350	95

(e.g., "Although AI is powerful, it still has limitations, and its future impact is uncertain.")	

As it can be seen in table 2 above, compound-complex sentences consistently resulted in the highest coherence scores, suggesting that LLMs benefit from the richer context provided by these more intricate structures. However, this comes at the cost of increased processing time, reflecting the additional cognitive load required to parse and generate responses to these prompts.

The most contextually rich outputs were observed in prompts that combined complex morphological forms with intricate syntactic structures. For example, prompts containing both derivational morphology and embedded clauses (e.g., *"The industrialization that transformed the economy also reshaped societal norms."*) produced some of the most nuanced responses, but with the highest latency and processing demands. This highlights the cognitive cost associated with processing both morphological and syntactic complexity simultaneously; strengthening the requirement for balanced prompt design in high-stakes applications like legal analysis, medical diagnostics, or creative writing.

These findings emphasize the importance of carefully balancing morphological and syntactic complexity in prompt design, depending on the specific goals and constraints of the application. For tasks requiring fast, accurate responses, simpler, more direct prompts may be preferable, while more creative or context-sensitive applications might benefit from the richer outputs associated with complex linguistic structures. This trade-off is particularly relevant for developers of AI writing assistants, customer service bots, and educational tools, where response quality must be carefully calibrated to meet user expectations.

CONCLUSION

This study provides a detailed examination of how morphological and syntactic complexity shape the performance of large language models (LLMs) in generating text. The results confirm that both dimensions exert a significant influence on output quality, coherence, and contextual depth, underscoring the strategic role of prompt formulation in practical applications. Prompts with greater morphological complexity consistently yielded richer and more precise responses, but at the expense of increased processing time and elevated perplexity scores—reflecting the computational demands of handling complex word forms. Similarly, syntactically sophisticated structures, particularly compound-complex sentences, were associated with higher coherence and contextual accuracy, yet required longer generation times. When these two forms of complexity were combined, the resulting outputs achieved the highest semantic accuracy and contextual relevance, albeit with the heaviest computational load. These patterns highlight a clear trade-off between linguistic sophistication and operational efficiency, suggesting that prompt designers must balance these factors according to task requirements.

These findings have important implications for the design and optimization of AI prompts. For tasks that prioritize speed and efficiency, simpler morphological and syntactic structures may be preferable, as they reduce processing time and minimize ambiguity. In contrast, applications that require deeper contextual understanding, such as creative writing, content generation, or advanced language teaching tools, may benefit from more complex prompt structures that can evoke richer, more nuanced responses from the model.

While this study provides valuable insights, it is not without limitations. The analysis primarily focused on English, a relatively morphologically simple language. In addition, the study relied heavily on automated metrics like perplexity, BLEU, and ROUGE, which, while informative, may not fully capture the qualitative aspects of language generation, such as creativity, tone, and emotional impact. Including human evaluations in future studies could provide a more comprehensive assessment of prompt effectiveness.

REFERENCES

Alek. (2023). Digital discourse analysis (DDA): A new perspective on language use in the digital age.

Anthony, L. (2023, September). Corpus AI: Integrating large language models (LLMs) into a corpus analysis toolkit. *JAECS 49th Conference*. https://osf.io/62mtf

Athukorala, W., & Jayasundara, C. C. (2025). Social media as a digital communication tool for academic library marketing: A systematic literature review.

Ayoub, N. N., Joudi, N. S., Bou Saba, M. S., & Bou Saba, A. S. (2024). Integrating artificial intelligence in a morphology course: An analytical study of university students'

p-ISSN 2301-7554, e-ISSN 2541-3643
Volume 13, Issue 3, October 2025

perspective. *European Scientific Journal, ESJ*, *20*(17), 1. https://doi.org/10.19044/esj.2024.v20n17p1

Balaji, K., Lokesha, A., Chandana, G., & Pushpa, H. M. (2024). Optimizing prompt length and specificity for enhanced AI chatbot responses. *International Journal of Current Science Research and Review*, *7*(9), 7421–7428. https://doi.org/10.47191/ijcsrr/V7-i9-61

Booij, G. (2018). Morphologie morphology ein internationales handbuch zur flexion und. https://doi.org/10.1016/B0-08-044854-2/00115-2

Chen, B., Zhang, Z., & Langren, N. (2024). Unleashing the potential of prompt engineering: A comprehensive review. *ArXiv:2310.14735v3[Cs.CL]*, 1–46. https://doi.org/10.1016/j.patter.2025.101260

Chen, Y., Arkin, J., Hao, Y., Zhang, Y., Roy, N., & Fan, C. (2024). AI-generated texts in multi-step tasks (PROMST): Integrating human feedback and heuristic-based sampling. 3859–3920. http://arxiv.org/abs/2402.08702

Durt, C., & Fuchs, T. (2024). Large language models and the patterns of human language use. In *Phenomenologies of the digital age: The virtual, the fictional, the magical* (pp. 106–121). https://doi.org/10.4324/9781003312284-7

Gemiharto, I., & Masrina, D. (2024). User privacy preservation in AI-powered digital communication system, *13*(2), 349–359.

Georgiou, G. P. (2023). Differentiating between human-written and AI-generated texts using linguistic features automatically extracted from an online computational tool, 1–12.

Gnach, A., Weber, W., Engebretsen, M., & Perrin, D. (2022). *Digital communication and media linguistics*. https://doi.org/10.1017/9781108780445

Hajnal, I. (2025). AI and linguistics: How human are large language models?

Harnad, S. (2024). Language writ large: LLMs, ChatGPT, grounding, meaning and understanding. *FRAI Journal*, 1–25. https://doi.org/10.3389/frai.2024.1490698

Igiri, O. T. (2018). A comparative analysis of morphology and syntax in the study of meaning in the English language. *Knowledge Review*, *37*(2), 130–135.

Itg, V. D. E. (2024). Large language models are transformers in artificial intelligence, industry, education, and society.

Jixing, L., & Shiyi, Z. (2019). Trends of qualitative research in applied linguistics since the new millennium. *Canadian Social Science*, *15*(7), 11–14. https://doi.org/10.3968/11200

Kamsinah, K., Natsir, N., & Aliah, N. (2024). Pragmatic analysis in digital communication: A case study of language use on social media. *4*.

Khabazian, F. (2024). A comparative study on syntactic morphological similarities and differences between Korean and Persian languages. https://doi.org/10.22034/ijpie.2023.173090

Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence*, *6*(December 2023), 100225. https://doi.org/10.1016/j.caeai.2024.100225

Koeva, S. (t.t.). Large language models in linguistic research: The pilot and the copilot. 319–329.

Kumar, D., & Singh, S. (2024). Advancements in transformer architectures for large language model: From BERT to GPT-3 and beyond. *International Research Journal of Modernization in Engineering Technology and Science*, *May*. https://doi.org/10.56726/irjmets55985

Ladzekpo, G. K., Collins, ✉, Amekor, K., & Akrobotu, M. (2023). Language and communication in the digital age: The study of how new technologies and digital media are affecting language use, communication patterns, and sociolinguistic dynamics. *Journal of Literature and Linguistics Studies*, 24–31. https://doi.org/10.61424/jlls

Lan, H. (2020). Comparison of research methods in applied linguistics: Quantitative, qualitative and mixed methods. *3*(14), 78–85. https://doi.org/10.25236/FER.2020.031414

Lande, D., & Strashnoy, L. (2025). Semantic AI framework for prompt engineering. https://doi.org/10.13140/RG.2.2.21116.24964

Llms, M., Mayilvaghanan, K., Nathan, V., & Kumar, A. (2025). Propel: AI-generated texts with expert priors for small, 272–302.

Mahlow, C. (2023). Large language models and artificial intelligence, the end of (language) learning as we know it—or not quite? 1–22.

Mohammad, A. F., & Clark, B. (2023). Large language model (LLM) & GPT, a monolithic study in generative AI. *LLM*, 383–388. https://doi.org/10.1109/CSCE60160.2023.00068

Muñoz-Ortiz, A., Gómez-Rodríguez, C., & Vilares, D. (2024). Contrasting linguistic patterns in human and LLM-generated news text. *Artificial Intelligence Review*, *57*(9), 1–28. https://doi.org/10.1007/s10462-024-10903-2

Muñoz de Cote Gudiño, L. M. del C. (2013). Qualitative research in applied linguistics: Approaches used for research projects developed in Mexico. *Revista Electrónica de Investigación Educativa*, *15*(1), 1–4. http://redie.uabc.mx/vol15no1/contenidohttp://redie.uabc.mx/vol15no1/contenido

Murphy, E., de Villiers, J., & Morales, S. L. (2025). A comparative investigation of compositional syntax and semantics in DALL·E and young children. *Social Sciences and Humanities Open*,

11(May 2024), 101332. https://doi.org/10.1016/j.ssaho.2025.101332

Ortolan, P. (2023). Optimizing prompt engineering for improved generative AI content, 1–108.

Sabbatella, A., Ponti, A., Giordani, I., & Candelieri, A. (2024). AI-generated texts in large language models.

Shehzadi, T. (2025). Tracing the changes in English syntax: A comparative analysis of old English and modern English. https://doi.org/10.13140/RG.2.2.26414.01602

Sibonghanoy Groenewald, E., Adolph Groenewald, C., Rani, S., Singla, P., & Howard, E. (2024). Artificial intelligence in linguistics research: Applications in language acquisition and analysis. *Artificial Intelligence In*, *March*. https://museonaturalistico.it

Siddiqui, K. A., Soomro, B., Abbasi, R. H., Rehman, U. A., & Soomro, A. (2021). Qualitative research in applied linguistics: A practical introduction, edited by Juanita Heigham and Robert Croker. *IJORER: International Journal of Recent Educational Research*, *2*(5), 610–613. https://doi.org/10.46245/ijorer.v2i5.110

Simatupang, E. C., & Heryono, H. (2022). New-word formation and social disruption on metaverse. *English Review: Journal of English Education*, 1013–1022.

Simatupang, E. C., & Supri, I. Z. (2024). Fusing translanguaging with speaking technology in the hospitality industry: Improving English communicative competence. *Social Sciences & Humanities Open*, *10*, 101166.

Srivastava, S., & Yao, Z. (2025). Revisiting AI-generated texts with large reasoning models: A case study on event extraction. 1–39. http://arxiv.org/abs/2504.07357

Stadtländer, C. T. K.-H. (2009). Qualitative, quantitative, and mixed-methods research. *Microbe Magazine*, *4*(11), 485–485. https://doi.org/10.1128/microbe.4.485.1

Tang, R., Chuang, Y. N., & Hu, X. (2024). The science of detecting LLM-generated text. *Communications of the ACM*, *67*(4), 50–59. https://doi.org/10.1145/3624725

Velásquez-Henao, J. D., Franco-Cardona, C. J., & Cadavid-Higuita, L. (2023). Prompt engineering: A methodology for optimizing interactions with AI-language models in the field of engineering. *90*(230), 9–17.

Wahle, J. P., Ruas, T., Xu, Y., & Gipp, B. (2023). Paraphrase types elicit prompt engineering capabilities.

Wang, M. H., Jiang, X., Zeng, P., Li, X., Chong, K. K. L., Hou, G., Fang, X., Yu, Y., Yu, X., Fang, J., & Pan, Y. (2025). Balancing accuracy and user satisfaction: The role of prompt engineering in AI-driven healthcare solutions. *Frontiers in Artificial Intelligence*, *8*(February), 1–10. https://doi.org/10.3389/frai.2025.1517918

Wang, X., Li, X., Li, X., & Biemann, C. (2024). Probing large language models from a human behavioral perspective. *1st Workshop on Bridging Neurons and Symbols for Natural Language Processing and Knowledge Graphs Reasoning, NeuSymBridge 2024 at LREC-COLING 2024 - Workshop Proceedings*, 1–7

.