AI, Technology, and the Future of Language Studies: Quantitative Evidence from LLaMA-Generated English Text

M. Lenin Babu
Assistant Professor, Department of Electronics and Communication Engineering, Bapatla Engineering College, Bapatla, Andhra Pradesh, India.

Abstract: The growing integration of Artificial Intelligence (AI) is reshaping language studies by influencing how language is produced, interpreted, and evaluated through human–AI interaction. This study examines how variations in English prompt formulation affect the linguistic characteristics and task alignment of AI-generated text. Using a transformer-based Large Language Model Meta AI (LLaMA) accessed through a web-based interface, sixteen controlled prompts were designed across four linguistic dimensions: syntactic variation, lexical choice, pragmatic framing, and contextual specificity. The generated responses were evaluated using five metrics: word count, mean sentence length, lexical density, Academic Word List (AWL) percentage, and a rubric-based task fulfilment score. The results show that prompts with explicit academic framing and contextual specificity consistently yield higher task fulfilment and greater Academic Word List (AWL) percentage, independent of response length. In contrast, general descriptive prompts tend to produce longer but less academically aligned outputs. These findings provide quantitative evidence that prompt formulation functions as a key linguistic control mechanism, highlighting prompt literacy as an emerging competence in the future of language studies.

Keywords: Artificial Intelligence, Language Studies, Prompt Design, LLaMA, Academic Word List (AWL), Text Analysis.

1. Introduction

The increasing use of Artificial Intelligence (AI) in academic communication has transformed how language is produced and evaluated, requiring language studies to account for machine-generated text alongside human-authored discourse. Unlike conventional writing, where authors directly control lexical and syntactic choices, interaction with AI systems is mediated through prompts, which indirectly shape the generated output. Prompt formulation therefore functions as a linguistic mechanism that links human intent to machine-generated discourse.

Although recent transformer-based language models can generate fluent and context-sensitive English text, their outputs remain highly sensitive to prompt wording. Small variations in syntactic structure, lexical choice, pragmatic framing, or contextual constraints can substantially affect register, relevance, and task alignment. Prior research on prompt design and prompt literacy has highlighted the importance of user–AI interaction but has largely remained descriptive, offering limited quantitative evidence from a language studies perspective [1]–[4].

From a linguistic standpoint, this represents a clear research gap. Existing work on AI-generated language has focused predominantly on model architecture, training paradigms, or ethical considerations, with comparatively less attention to measurable linguistic outcomes such as lexical density, Academic Word List (AWL) percentage, and task fulfilment [2], [5], [6]. In language studies, these measures are closely associated with register variation and genre conventions in formal writing, making them suitable indicators for evaluating academic discourse.

This paper presents a quantitative analysis of how controlled variations in English prompt formulation affect the linguistic characteristics and academic alignment of AI-generated text. Focusing on syntactic variation, lexical choice, pragmatic framing, and contextual specificity, it

conceptualizes prompt formulation as a form of linguistic control and supports prompt literacy as a competence in AI-mediated academic communication. Section 2 outlines the experimental design, prompt construction, model configuration, and evaluation metrics; Section 3 reports the quantitative results; Section 4 interprets the findings in relation to research on prompt literacy and academic register; and Section 5 summarizes the main contributions and future research directions.

2. Methods and Methodology

This study adopts a controlled, quantitative research design to examine how English prompt formulation influences the linguistic characteristics and academic alignment of AI-generated text. The methodology isolates prompt-level linguistic variation while keeping the AI system, topic domains, and interaction conditions constant, enabling systematic comparison. Given the specialized nature of the linguistic variables being isolated, this research is structured as a pilot exploratory study (N=16). This approach allows for a granular, ablation-style analysis of specific prompt-level modifications while providing a replicable framework for future large-scale validation.

All textual outputs were generated using a transformer-based large language model from the LLaMA family, specifically the **LLaMA-4-Maverick-17B-128E Instruct-Turbo** model (Meta), accessed through the **DeepInfra** web interface. The model employs a **17-billion-parameter mixture-of-experts architecture with 128 experts** and was used solely for text generation. No fine-tuning, system-level parameter modification, or iterative prompting was performed, ensuring that observed output differences could be attributed exclusively to prompt formulation.

The experimental design follows an ablation-style framework. Sixteen prompts were organized into eight paired comparisons across four linguistic dimensions, with each pair differing in only one aspect. **Syntactic variation** was examined using two pairs (A1–A2 and B1–B2), **lexical variation** using two pairs (C1–C2 and D1–D2), **pragmatic framing** using two pairs (E1–E2 and F1–F2), and **contextual specificity** using two pairs (G1–G2 and H1–H2). Briefly, base prompts such as "Explain climate change" were contrasted with refined versions specifying analytical structure, academic audience, or contextual constraints. Topic continuity was preserved within each pair so that observed differences could be directly attributed to the targeted linguistic modification. This controlled, pairwise approach aligns with prior work on prompt evaluation and human and AI interaction [8], [11].

Table 1 presents representative examples of base and refined prompts for each linguistic dimension. The complete prompt set followed the same design logic, with each pair differing in only one targeted linguistic feature.

Table 1. Representative Samples of Prompt Types Used in the Study

Linguistic Dimension	Base Prompt (Verbatim Sample)	Refined Prompt (Verbatim Sample)
Syntactic variation	Explain climate change.	Explain climate change using a structured analytical format.
Lexical variation	Describe the effects of climate change.	Analyze the impacts of climate change using formal academic language.
Pragmatic framing	Write about climate change.	Write an academic explanation of climate change for university students.
Contextual specificity	Explain climate change.	Explain the causes and consequences of climate change in coastal regions.

Each prompt was submitted once, and only the first generated response was collected. No follow-up prompts or response regenerations were used. All outputs were stored in plain-text format for analysis. Five evaluation metrics were employed: word count (response length), mean sentence length (syntactic complexity), lexical density (proportion of content words), and Academic Word List

(AWL) percentage based on the Academic Word List, and a rubric-based task fulfilment score. Task fulfilment was assessed on a five-point scale, where **5** indicated full satisfaction of prompt intent and academic expectations, and **4** indicated minor limitations in completeness, focus, or academic register. Specifically, the 5-point rubric was defined to ensure consistency: a score of 5 represents full adherence to academic register and prompt constraints; a score of 4 indicates minor deviations in lexical density or focus; while scores of 3 and below (though not observed in this sample) represent increasing levels of task misalignment.

Data analysis was conducted on a pairwise comparative basis, with metric changes interpreted relative to the specific linguistic modification applied, supporting precise attribution of effects and interpretation of prompt formulation as a linguistic control mechanism. To assess whether refined prompts differed significantly from base prompts, a Wilcoxon signed-rank test was applied to paired Academic Word List (AWL) percentages across the eight prompt pairs.

3. Results

This section reports the quantitative outcomes of the prompt-based experiment. The complete numerical results for all sixteen prompts are presented in **Table 2.**

Table 2. Quantitative results of LLaMA prompt analysis (W.C = word count; MSL = mean sentence length; L.D = lexical density; AWL = Academic Word List percentage; TFS = task fulfilment score).

Prompt ID	W.C	MSL	L.D	AWL (%)	TFS
A1	518	21.33	66.60	0	5
A2	491	21.91	70.26	0	5
B1	405	12.74	72.10	1.11	5
B2	380	13.88	70.53	3.75	5
C1	623	16.47	60.35	0.16	4
C2	520	13.68	59.81	1.30	5
D1	427	17.38	64.17	0.71	4
D2	525	17.79	65.14	1.15	5
E1	415	15.04	66.75	0.96	4
E2	386	15.96	71.24	1.81	5
F1	445	15.71	66.74	2.70	4
F2	500	17.75	64.80	2.59	5
G1	516	19.19	68.60	1.94	4
G2	147	20.57	70.07	6.12	5
H1	424	18.43	68.40	1.89	4
H2	409	21.42	64.06	1.71	5

Results are described strictly in terms of observed values across the five-evaluation metrics: word count, mean sentence length (MSL), lexical density (L.D), Academic Word List (AWL) percentage, and task fulfilment score (TFS).

As shown in Figure 1, refined prompts generally produce higher Academic Word List (AWL) percentages than base prompts, with the largest increase observed for the context-specific pair G.
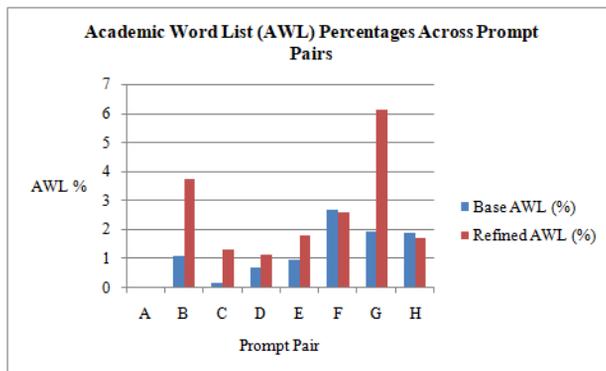
Figure 1. Comparison of Academic Word List (AWL) Percentages between Base and Refined Prompts

Responses vary widely across prompts: word count ranges from 147 (G2) to 623 (C1), mean sentence length from 12.74 (B1) to 21.91 (A2), lexical density from 59.81% (C2) to 72.10% (B1), and AWL from 0% to 6.12%. Task fulfilment scores range between 4 and 5. All prompt pairs are evaluated using the same five metrics, enabling direct comparison.

3.1 Effect of Syntactic Variation

The A pair (A1–A2) shows minimal impact on academic alignment: both yield AWL = 0% and task fulfilment = 5, with only minor differences in lexical density and sentence length. The B pair (B1–B2) shows moderate surface variation, with reduced word count, slightly increased sentence length, and higher AWL (1.11% to 3.75%), while task fulfilment remains unchanged. Thus, syntactic rephrasing mainly affects surface measures without improving task alignment.

3.2 Effect of Lexical Variation

In the C pair, reduced word count (623 to 520) coincides with higher AWL (0.16% to 1.30%) and improved task fulfilment (4 to 5). In the D pair, the refined prompt yields higher lexical density (65.14% vs. 64.17%), higher AWL (1.15% vs. 0.71%), and higher task fulfilment (5 vs. 4), despite a longer response. Lexical refinement therefore produces stronger gains than syntactic variation.

3.3 Effect of Pragmatic Framing

Pragmatic framing improves task fulfilment in both pairs. In the E pair, lexical density (66.75% to 71.24%) and AWL (0.96% to 1.81%) increase as word count decreases, raising task fulfilment from 4 to 5. In the F pair, the academically framed prompt produces a longer response and higher sentence length, with task fulfilment again improving (4 to 5) despite a slight decrease in AWL.

3.4 Effect of Contextual Specificity

Contextual specificity yields the strongest effects. In the G pair, the refined prompt produces a much shorter response (147 vs. 516 words) while increasing lexical density (70.07% vs. 68.60%) and AWL (1.94% to 6.12), with task fulfilment rising from 4 to 5. In the H pair, task fulfilment also improves (4 to 5) with a small reduction in word count and an increase in sentence length, despite slight decreases in lexical density and AWL.

3.5 Cross-Metric Observations

A Wilcoxon signed-rank test was performed on the AWL percentages across the eight pairs to evaluate the impact of prompt refinement. The analysis yielded a test statistic of W = 3.0 and a p-value of 0.063. Because the p-value exceeds the conventional threshold ($\alpha = 0.05$), the null hypothesis—stating no significant difference exists—cannot be rejected. However, a strong directional trend is observed, with AWL usage increasing in 75% of the refined pairs (notably Pair G,

which rose from 1.94% to 6.12%). These findings support the utility of the experimental framework as a pilot investigation.

4. Discussion

The findings indicate that English prompt formulation plays a central role in shaping the academic quality of AI-generated text. While syntactic variation influences surface-level features such as sentence length and lexical density, it shows comparatively smaller effects on Academic Word List (AWL) percentage and task fulfilment, suggesting that grammatical restructuring alone is insufficient to induce an academic register. In contrast, lexical refinement, pragmatic framing, and contextual specificity were observed to improve task fulfilment and academic alignment within this experimental sample.

Prompts incorporating analytical verbs, explicit audience specification, or contextual constraints yield responses that are more focused and academically appropriate. The strong performance of context-specific prompts further demonstrates that academic quality is determined more by prompt precision and relevance than by response length.

These results support the interpretation of prompts as linguistic control mechanisms in AI-mediated writing. From a language studies perspective, the ability to formulate effective prompts, referred to as prompt literacy, emerges as an essential competence for academic communication in AI-supported environments. This interpretation aligns with prior research that conceptualizes prompt literacy and prompt engineering as emerging competencies in AI-mediated academic communication [2], [6], [8].

The study has several limitations. The analysis is based on a single large language model accessed through a hosted inference platform and focuses exclusively on English prompts within a controlled experimental setting. In addition, the prompt set is limited to sixteen paired instances, and the evaluation relies primarily on descriptive metrics, with limited inferential statistical testing. Accordingly, the findings should be interpreted as exploratory patterns rather than broadly generalizable conclusions.

5. Conclusion

This study proposes a replicable, pilot linguistic framework for examining how prompt formulation shapes the academic characteristics of AI-generated text. The results show that lexical refinement, pragmatic framing, and contextual specificity consistently enhance task fulfilment and **Academic Word List (AWL) percentage,** whereas syntactic variation alone has a limited effect.

Overall, academic quality is influenced more by prompt precision and contextual clarity than by response length, supporting prompt literacy as a key competence in AI-mediated academic communication. The framework is replicable and can be extended to other language models, languages, and educational contexts. Future work may also explore large language models as auxiliary evaluators (judge LLMs) to complement human assessment of task fulfilment and academic register.

6. References

1. Chu, K. Y., & Chiu, P. M. H. (2025). Prompt design and AI response quality in university students' academic learning tasks. *Proceedings of the Association for Information Science and Technology*.
2. Hwang, Y., Lee, J. H., & Shin, D. (2023). What is prompt literacy? An exploratory study of language learners' development of a new literacy skill using generative AI. *arXiv preprint*.
3. Barashkina, E., & Maslenkova, N. A. (2025). Communicative strategies of user interaction with generative AI: Prompt formulation as a genre structure. *Chelyabinsk Humanities Journal*.

4. Leidinger, A., Rooij, R., & Shutova, E. (2023). The language of prompting: What linguistic properties make a prompt successful? In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

5. Kang, X., Li, X.-Z., Bai, X.-X., Zhang, Y., & Chen, S. (2025). Exploring the structural logic and learning path of prompt language in AI-assisted interaction design. *Education Journal*.

6. Knoth, N., Tolzin, A., Janson, A., & Leimeister, J. M. (2024). AI literacy and its implications for prompt engineering strategies. *Computers and Education: Artificial Intelligence, 5*, Article 100165.

7. Zakhster, E. M. (2025). Prompt engineering in English language teaching: Theory and practice. *Humanities and Social Sciences Bulletin of the Financial University*.

8. Braun, M., Greve, M., Kegel, F., Kolb, L., & Beyer, P. E. (2024). Can (AI) have a word with you? A taxonomy on the design dimensions of AI prompts. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*.

9. Deshmukh, R., Raut, R., Bhavsar, M., Gurav, S., & Patil, Y. (2025). Optimizing human–AI interaction: Innovations in prompt engineering. In *Proceedings of the 3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*.

10. Viveros-Muñoz, R., Carrasco-Sáez, J. L., Contreras-Saavedra, C., San-Martín-Quiroga, S., & Contreras-Saavedra, C. E. (2025). Does the grammatical structure of prompts influence the responses of generative artificial intelligence? An exploratory analysis in Spanish. *Applied Sciences, 15*(3), Article 1324.

11. Sclar, M., Choi, Y., Tsvetkov, Y., & Suhr, A. (2023). Quantifying language models' sensitivity to spurious features in prompt design: How I learned to start worrying about prompt formatting. In *Proceedings of the International Conference on Learning Representations (ICLR)*.