EaseText: Enhancing Comprehension of Complex Sentences Using LLMs

Praneeth Vadlapati*

A B S T R A C T

Numerous sentences are considerably complex to understand for numerous readers, including non-native speakers of a language and readers who lack expertise in a field. Complex sentences often introduce significant challenges for readers from diverse age groups and mental capacities who might lack focus or expertise in reading. This research investigates the usage of a Large Language Model (LLM) in the simplification and explanation of sentences with complex structure and meaning. Complex sentences commonly include technical words, complicated sentence structures and uncommon vocabulary. Simplification of complex text broadens accessibility for readers to understand complex text. This paper proposes a method to leverage an LLM to replace uncommon words with standard words, break down complex sentences into understandable bullet points and generate a concise summary. A validation process is conducted by leveraging the selected LLM to ensure that the simplified outputs retain the meaning and the context of the original text. The method has numerous potential applications in education and legal interpretation of the laws and the rules. The findings reveal the potential of LLMs to improve accessibility to information for diverse readers. The code is available at github.com/Pro-GenAI/EaseText.

Keywords: Artificial Intelligence (AI), Large Language Models (LLMs), Natural Language Processing (NLP), Natural Language Understanding (NLU), Sentence Simplification, Accessibility

1. Introduction

Comprehension and interpretation of complex texts is a common challenge across a wide range of fields[1-4] in multiple use cases, including government documentation, legal paperwork and scientific literature. Complex sentences with long clauses, numerous commas, advanced technical vocabulary and multiple embedded ideas create a risk of obstructing comprehension[5,6] or redirecting the interpretation, which leads to misinformation. Challenges include language barriers. Complex sentences also affect readers who experience cognitive illness, overload or stressful conditions in which they cannot comprehend complex text with patience[7].

A. Proposed system and its benefits

The recent advancements in LLMs in artificial intelligence (AI) enabled novel opportunities to use LLMs to process natural language[8-10], which enables the generation of understandable explanations of complex sentences. Previous language models encountered challenges with accuracy in preserving meaning during simplification. Modern LLMs have been trained on vast amounts and variants of text from numerous sources[11,12]. This study utilizes a modern LLM to simplify and explain complex sentences. The process ensures the integrity of the context and enhances the accessibility of the text to a broader audience. LLMs can be used with custom instructions called prompts to customize the model according to the requirements[13-15]. The method is applicable in a wide range of contexts, such as technical documentation and legal interpretation of laws.

2. Related Work

Drozdov et al. (2022)[16] worked on compositional semantic

parsing using LLMs with a focus on structured task execution by utilizing structured queries to answer text-based questions. Srikanth et al. (2021)[17] worked on elaborative simplification using pre-trained language models. Mathews et al. (2018)[18] utilized attentive neural networks that are sequence-to-sequence models to simplify sentences. There is a considerable amount of research on making text more understandable. However, the existing research did not experiment with explaining complex technical statements by dividing them into multiple points and a concise summary using an LLM based on custom instructions. LLMs based on Transformer architecture[19] are a recent advancement. LLMs process language with fluency and contextual understanding, which ensures they are suitable for sentence simplification tasks[20]. This study covers the research gap with a focus on leveraging LLMs to explain complex text using points and a concise summary.

3. Methods

A. Selecting and loading an LLM

An accurate language model with high performance and natural language understanding is necessary for this step. Hence, ChatGPT is selected as the LLM based on its accuracy. "ChatGPT is a sibling model to Instruct GPT, which is trained to follow an instruction in a prompt and provide a detailed response." [21] "ChatGPT is fine-tuned from a model in the GPT-3.5 series, which finished training in early 2022." [21]

B. Preparation of complex sentences

Large complex sentences are written for the experiment, ensuring the inclusion of diverse structures, complex vocabulary, more commas and numerous ideas in numerous words, which are 80 to 90 words for one sentence. The criteria used for creation were to represent a range of complex sentence types that readers face across numerous real-world tasks. Three such sentences were written in this step to utilize in the steps that follow.

Table 1: Complex sentences.

Index	Complex sentence	Number of words
1	Having conducted extensive research in the field of sociolinguistics, I am interested in investigating the intricate relationships between linguistic variation and social identity, particularly with regards to the syntactic and phonological features that are indicative of regional or ethnic affiliations, in order to further comprehend the underlying mechanisms of language change and diffusion within diverse speech communities and to ultimately contribute to the development of more nuanced and sophisticated language policies that are reflective of the rich and dynamic linguistic landscapes that characterize our contemporary globalized world.	87
2	As a scholar with a specialized focus in historical linguistics, I aim to elucidate the complex processes of language evolution and linguistic diversification, by examining the interplay between cultural, demographic and geopolitical forces and their consequential impact on the development and diffusion of distinct language families and their respective protolanguages, in an endeavor to uncover the deep-seated relationships between past sociocultural contexts and the cognitive frameworks that have shaped the linguistic structures and patterns that persist across disparate linguistic systems.	80
3	In light of the recent advancements in cognitive linguistics and psycholinguistics, I am embarking on an ambitious research venture that seeks to elucidate the intricate mechanisms that underpin language processing and production, by comprehensively exploring the dynamic interactions between cognitive processes, linguistic structure and communicative functions, in order to formulate a more comprehensive understanding of the cognitive underpinnings of language use and comprehension and to pave the way for innovative interventions and applications within the realms of language education and clinical linguistics.	82

C. Simplifying the sentences using the LLM

A prompt is designed to instruct the selected LLM to simplify the prepared sentences and divide each sentence into bullet points that are followed by a concise summary. The prompt is ensured to encourage the model to break down complex parts of a sentence and use accessible and understandable vocabulary while maintaining the original context and meaning of the original sentence. This step ensures that the readers are comfortable comprehending the explanation of a complex sentence with reduced cognitive strain. The designed prompt is used with an LLM and complex sentences to generate responses that include simplified explanations.

D. Validation of the simplified text

A validation process is necessary to ensure the new LLM-generated text maintains the context and meaning of the original text. This step ensures to assess whether the context of each simplified text is lost or misrepresented. The same LLM is used for this step by utilizing a prompt to evaluate the responses. The model is provided with the original text and simplified text. The evaluation results are calculated based on this step.

4. Results

A. Simple explanations generated

Simpler explanations are generated for each complex sentence by the LLM according to the instructions provided through the prompt. The LLM followed the instructions mentioned in the prompts and generated simpler sentences that were followed by bullet points and a concise summary to conclude the meaning of the original text. The simplified explanations generated for each complex sentence are presented below.

B. Validation results

Simplified text has been compared with the original text to validate whether the original context has been lost or altered. The validation results of the simplified text are mentioned below.

5. Discussion

The findings reveal that LLMs are effective in processing natural language to allow a diverse audience to understand the text. The method used in this research provides advantages through the simplification of challenging sentences. The method enhances accessibility to information and expands the applications of LLMs. However, the LLMs might struggle with comprehending the meaning of highly complex sentences. Sometimes, the process leads to an excessive simplification of the text, leading to an increased number of words for the readers to read. Additional research on prompt engineering techniques and LLMs could solve such challenges and improve accuracy. Before implementation for real-world tasks, the system requires further testing that could be accomplished by the experts in each

field. Further experiments have yet to be conducted on potential bias added by the LLM to avoid misrepresentation of original intentions.

Table 2: Simple explanations generated.

Index	Simpler text
1	I have studied different ways people speak and how it relates to their social identity. I want to understand how language changes and spreads in diverse communities and help create better language policies. Reasons: - Studied different ways people speak - Interested in how language changes and spreads Summary: Interested in the relationships between language variation and social identity and how it influences language change and language policies.
2	As a linguist focusing on historical language changes, I aim to understand how languages have evolved and diversified over time by looking at the effects of culture, population and geography on the development and spread of different language groups and their early forms. Reasons: - Study language change - Explore impact of culture, population and geography Summary: I study how cultural, demographic and geopolitical factors have influenced the development and spread of language families and their early forms.
3	In light of new research in language and psychology, I am beginning a big research project to understand how our brains process and produce language by looking at how our thoughts, language structure and communication all work together. Reasons: - Cognitive linguistics and psycholinguistics advancements - Elucidating language processing mechanisms Summary: (in 1 line) I am starting a research project to understand how our brains process and produce language.

Table 3: Validation results.

Index	Result
1	Success
2	Success
3	Success

6. Conclusion

This research contributes an unexplored methodology to utilize LLMs for the process of simplification of sentences. By breaking down complex text into comprehensible parts, the models can significantly enhance comprehension for the readers, particularly for readers who struggle to comprehend dense technical language. This approach has immediate applications in making complex laws more understandable to new readers without expertise in the laws. Another application includes making technical documentation understandable to new readers in the field who lack expertise in reading or writing technical documentation. Future research could include interactive methods in which the text gets relatively less simplified initially and the user gets an opportunity to ask queries to the LLM, which answers the queries based on the context of the sentence and the entire document in which the sentence exists. With continued advancements in LLM technology, this method could be adapted to various languages and industries to bridge linguistic gaps and support more precise technical communication across diverse global audiences.

Appendix

1. Prompt templates used to process using LLMs

```
Sentence:
{large_sentence}
---

Act as an expert linguist and simplify the sentences.
Simplify the sentences by simplifying difficult words and phrases.
Try to make bullet points for the complex sentence.
Return inside tags.

Example response:
<simplified_sentence>
Line 1
Reasons:
 - Reason 1
 - Reason 2
Summary: (in 1 line)
Summary of the sentence.
</simplified_sentence>
```

Figure A1: Prompt template to simplify complex text.

```
Large Sentence:
{large_sentence}
---
Simplified Sentence:
{simplified_text}
---

Act as an expert linguist and validate the meaning of the simplified sentence.
Return comments inside tags and then return the final validation result
(True/False) inside tags.
Example response:
<comments>
The simplified sentence does not capture the part about doctors.
</comments>
<validation>False</validation>
```

Figure A2: Prompt template to validate a simplified text

7. References

1. Hackemann T, Heine L and Höttecke D. "Challenging to Read, Easy to Comprehend? Effects of Linguistic Demands on Secondary Students' Text Comprehension in Physics," International Journal of Science and Mathematics Education, 2022;20(1):43-68.

2. Platas-García, J. M. Castro-Manzano and V. Reyes-Meza, "Reading comprehension as a complex phenomenon. An approach based on assessment, inference and foreign language," Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education, 2020;15(2):17-37.

3. Spencer M, Gilmour AF, Miller AC, Emerson AM, Saha NM, Cutting LE. "Understanding the influence of text complexity and question type on reading outcomes," Reading and Writing, 2019;32(3):603-637.

4. Spencer M and Wagner RK, "The Comprehension Problems of Children With Poor Reading Comprehension Despite Adequate Decoding: A Meta-Analysis," Review of Educational Research, 2018;88(3):366-400.

5. Ostadghafour S and Bialystok E. "Comprehension of complex sentences with misleading cues in monolingual and bilingual children," Applied Psycholinguistics, 2021;42(5):1117-1134.

6. de Ruiter LE, Theakston AL, Brandt S and Lieven EVM, "Iconicity affects children's comprehension of complex sentences: The role of semantics, clause order, input and individual differences," Cognition, 2018;171:202-224.

7. https://www.healthline.com/health/mental-illness-and-reading.

8. https://arxiv.org/abs/2208.11857

9. https://cloud.google.com/blog/products/ai-machine-learning/google-cloud-supercharges-nlp-with-large-language-models

10. y Arcas, "Do Large Language Models Understand Us?" Daedalus, 2022;151(2):183-197.

11. https://www.nature.com/articles/d41586-022-01191-3

12. https://stpp.fordschool.umich.edu/sites/stpp/files/2022-05/large-language-models-TAP-2022-final-051622.pdf

13. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

14. https://arxiv.org/abs/2205.11916

15. https://arxiv.org/abs/2211.01910

16. https://arxiv.org/abs/2209.15003

17. Srikanth N and Li JJ, "Elaborative Simplification: Content Addition and Explanation Generation in Text Simplification," in Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, C. Zong, F. Xia, W. Li and R. Navigli, Eds., Online: Association for Computational Linguistics, 2021:5123-5137.

18. https://arxiv.org/abs/1805.05557

19. Vaswani et al. "Attention is all you need," in Proceedings of the 31st International Conference on Neural Information Processing Systems, in NIPS'17. Red Hook, NY, USA: Curran Associates Inc., 2017;6000-6010.

20. Wolf T et al., "Transformers: State-of-the-Art Natural Language Processing," in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Q. Liu and D. Schlangen, Eds., Online: Association for Computational Linguistics, 2020:38-45.

21. https://openai.com/index/chatgpt/