ITU NLP at TSAR 2025 Shared Task: A Three-Stage Prompting Approach for CEFR-Oriented Text Simplification

Kutay Arda Dinç
Istanbul Technical University
Istanbul, Türkiye
dinck20@itu.edu.tr

Fatih Bektaş
Istanbul Technical University
Istanbul, Türkiye
bektas18@itu.edu.tr

Gülşen Eryiğit
Istanbul Technical University
Istanbul, Türkiye
gulsenc@itu.edu.tr

Abstract

Automatic Text Simplification (TS) makes complex texts more accessible but often lacks control over target readability levels. We propose a lightweight, prompt-based approach to English TS that explicitly aligns outputs with CEFR proficiency standards. Our method employs a three-stage pipeline, guided by rule-informed prompts inspired by expert strategies. In the TSAR 2025 Shared Task, our system achieved competitive performance, with stronger results at B1 level and challenges at A2 level due to over-simplification. These findings highlight the promise of prompt-based CEFR-oriented simplification and the need for more flexible constraint design.

1 Introduction

Automatic Text Simplification (TS) is an important NLP task that aims to rewrite complex texts into simpler forms while preserving meaning. TS benefits many groups, including non-native speakers, children, and individuals with reading or cognitive difficulties. However, current systems often simplify texts without controlling the target readability level, limiting their usefulness in education and accessibility, where alignment with standardized frameworks such as the Common European Framework of Reference for Languages (CEFR) (Council of Europe, 2001) is essential.

We propose a CEFR-oriented method for English text simplification that leverages Large Language Models (LLMs) through prompt engineering rather than model training, making it lightweight, accessible, and practical. Our method is inspired by the TS strategies introduced in Özmen and Kıran (2022). It applies a three-stage process: syntactic simplification, lexical simplification, and elaboration. This method addresses a key gap in current NLP research by combining accessibility with level-appropriate simplification.

2 Related Work

CEFR is widely used to define language proficiency and guide text simplification. Most CEFR-based TS research targets lexical or sentence-level simplification, creating datasets for word substitutions or training models to generate CEFR-aligned sentences (Arase et al., 2022; Uchida et al., 2018; Li et al., 2025; Barayan et al., 2024). However, simplification at the document or paragraph level under CEFR control remains relatively underexplored and challenging.

Early text simplification research largely focused on sentence-level approaches. However, applying these methods iteratively across entire documents often fails to maintain discourse coherence and overall text integrity (Alva-Manchego et al., 2020). This limitation has encouraged research to shift towards document-level simplification, which demands broader contextual understanding and specialized evaluation methods (Sun et al., 2021).

LLMs have recently become prominent tools for document-level text simplification, addressing coherence and context issues. Approaches such as ProgDS (Fang et al., 2025) decompose the simplification process into hierarchical stages (discourse-level, topic-level, and lexical-level), guiding LLMs to progressively simplify documents while preserving logical structure and readability. Building on this line of work, our study also adopts a hierarchical paradigm, though the definition and scope of the stages differ from ProgDS. In parallel, fine-tuning of LLMs has been shown to yield significant improvements in document-level text simplification tasks, effectively improving model accuracy and output relevance (Nozza and Attanasio, 2023; Alkaldi and Inkpen, 2023; Anschütz et al., 2023). Unfortunately, none of these approaches are explicitly oriented toward CEFR-based simplification.

Proceedings of the Fourth Workshop on Text Simplification, Accessibility and Readability (TSAR 2025), pages 149–154
November 4-9, 2025 ©2025 Association for Computational Linguistics

3 Method

Our method of simplification consists of three stages that are carried out in a progressive order: syntactic simplification, lexical simplification, and elaboration.

Syntactic simplification involves restructuring complex grammatical constructions to align with the requirements of the target CEFR level. Lexical simplification focuses on replacing difficult vocabulary with simpler alternatives appropriate for the target proficiency level. Elaboration involves adding explanatory words or clarifying information to make the text more comprehensible and suited to the intended CEFR level. In Özmen (2019), case-by-case rules were provided for all three types of simplifications at different proficiency levels, along with examples demonstrating how to execute these different simplification cases. Although the examples were presented in Turkish, the underlying simplification principles are not language-specific and reflect patterns observed across multiple languages (Özmen and Kıran, 2022). Building on this foundation, Bektaş et al. (2024) incorporated these case-specific rules into Turkish prompts and reported promising results. Informed by both studies, we follow these rules when designing prompts and apply them sequentially in three stages[1], in a manner inspired by Fang et al. (2025).

All stages utilize **GPT-4o** with the following parameters:

- **Model version**: `gpt-4o-2024-11-20`
- **Temperature**: `0.5`
- **Top-p sampling**: `0.95`

3.1 Syntactic Simplification

The first stage addresses the grammatical complexity by aligning the syntactic structures with the requirements of the target CEFR level. To implement this stage, we first compiled a comprehensive set of grammatical structures from the Cambridge English Grammar Profile (Cambridge University Press, 2024), which documents grammatical constructions across levels A1 to C2.

Then, the syntactic simplification rules from Özmen (2019) were translated to English, along with illustrative examples showing how complex structures should be simplified for each target level.

[1]The full workflow, including all prompts used at each stage, is available at `https://github.com/kutayardadinc/itunlp_tsar`.

The syntactic simplification process employs a structured prompt that includes:

i) A system instruction defining the model's role as a syntactic simplification expert,

ii) The complete set of permissible grammatical structures for the target CEFR level,

iii) Specific syntactic simplification rules with examples.

The core prompt structure follows this pattern: *A text at the [LEVEL] level should not contain any grammatical structures that are not provided to you below. If the given text contains any structure other than these grammatical structures, simplify the given text to the [LEVEL] level using syntactic simplification rules. [GRAMMAR_STRUCTURES]. [SIMPLIFICATION_RULES]. Do not perform any action other than syntactic simplification. Preserve the content and details of the text. Do not delete sentences from the text.*

3.2 Lexical Simplification

For lexical simplification, we first established vocabulary inventories for each CEFR level to be able to identify words that exceed the target proficiency level. We utilized two primary vocabulary resources: Oxford 3000 (Oxford University Press, a) and Oxford 5000 (Oxford University Press, b) word lists. These two combined create a vocabulary list of 5000 words from levels A1 to C1.

Our system automatically identifies lexically complex words within the syntactically simplified text by tokenizing it using spaCy (Honnibal et al., 2020) and checking both surface forms and lemmas against our compiled word-to-CEFR-level mapping database. Words that exceed the target CEFR level are explicitly included in the prompt, with instructions to replace them with simpler alternatives. These are followed by level-specific lexical simplification rules, accompanied by translated examples adapted from Özmen (2019).

The lexical simplification prompt structure includes:

i) A system instruction defining the model's role as a lexical simplification expert,

ii) Level-specific lexical simplification rules,

iii) Identified words that exceed the target level.

The core prompt structure follows this pattern:

Simplify the given text to the [LEVEL] level using lexical simplification rules. [SIMPLIFICATION_RULES]. In the following text, the word(s) [IDENTIFIED_WORDS] are above the [LEVEL] level. Without changing the meaning of the text, replace these words with simpler words appropriate for the [LEVEL] level. Do not change any words that have no simpler equivalent or are important for the meaning of the text.

3.3 Elaboration

The final elaboration stage enhances the comprehensibility of the text by adding explanatory information and clarifying potentially ambiguous content for the target level. This stage applies elaboration rules, translated from Özmen (2019), specific to each CEFR level. At this stage, the model is instructed to expand rare or complex concepts with explanatory phrases, clarify implicit meanings by adding context, and highlight key ideas through selective repetition when appropriate.

The elaboration employs the prompt below:

Using elaboration rules, make the given text more understandable for a student at the [LEVEL] level. [ELABORATION_RULES] Do not perform any action other than elaboration. Preserve the content and details of the text. Do not delete sentences from the text.

The elaboration stage ensures that the simplified text not only uses appropriate grammar and vocabulary for the target level but also provides sufficient contextual information to support comprehension for the target proficiency level.

4 Results and Discussion

Our three-stage progressive simplification pipeline yielded mixed results on the test data provided by the TSAR 2025 Shared Task (Alva-Manchego et al., 2025), with overall MeaningBERT scores of 0.797 for both original-output (MeaningBERT Original) and reference-output (MeaningBERT Reference) comparisons, and a CEFR compliance RMSE of 0.632.

Analysis of the highest and lowest scoring texts revealed that our top-performing cases were predominantly B1-targeted, while most of the underperforming examples were A2-level simplifications. The five lowest MeaningBERT scores for both metrics came from our A2-level productions, whereas the top five scores for MeaningBERT Original were all from B1-targeted simplifications. This performance gap stems largely from over-simplification induced by the strict A2-level linguistic constraints in the prompts, such as limited grammatical structures and restrictions on vocabulary exceeding A2 proficiency.

As detailed in the methodology, our syntactic simplification step constrains GPT-4o to avoid grammatical structures outside the target CEFR level, and the lexical simplification stage involves prompting GPT-4o to replace vocabulary that exceeds the target level. However, this approach can cause over-simplification in some cases.

During the syntactic simplification process, the system output can sometimes distort the original meaning by avoiding grammatical structures that are not included in the target level's grammar list. An example of this can be seen in the A2-level simplification of text 73:

Text ID: 73-A2
MeaningBERT Reference Score: 0.5836

Reference:

> *The two men looked at the north side of the land. It was so big that they could hardly see the fences...*

Simplified:

> *The two men looked at the far end of the land. The land was very big, and it stretched very far. The walls around the land were hard to see because the men were standing far away...*

The reference employs the "so...that" construction with the modal "could" which, while arguably appropriate for the A2 level, is a structure our system is forced to avoid. Because the prompt strictly limits the output to the provided A2 grammar list, which does not include the "so...that" construction for clauses of result, the system fragments the sentence into simpler, explicitly permitted coordinate clauses, resulting in reduced cohesion.

The lexical simplification stage shows a tendency to make unnecessary substitutions in A2 texts, as shown in the following example with text ID 96-A2:

Text ID: 96-A2
MeaningBERT Reference Score: 0.5954

Original:

No matter what kind of house it is, the basement is a scary place in horror films...

Reference:

In horror films, the basement is always a scary place...

Simplified:

The lower floor is a scary place in scary films...

Here “basement”, which is listed at the B2 level in our vocabulary list, is replaced with “lower floor” and “horror films” is changed to “scary films” because “horror” is listed at the B1 level. In the B1-targeted simplification, however, “horror” is correctly preserved, while “basement” is still replaced. Although this version also does not fully match the reference at B1, it achieves a higher MeaningBERT Reference score of 0.7328 compared to 0.5954 at A2 production, as making fewer lexical substitutions in the entirety of the text allows the output to maintain a higher semantic similarity to the reference.

Beyond systematic over-simplification, some failures result from meaning shift rather than linguistic constraints. Since our pipeline utilizes an LLM three times in sequence, each stage introduces the risk of semantic drift, which can hinder meaning preservation. As the simplification with text ID 120-A2 illustrates:

Text ID: 120-A2
MeaningBERT Original Score: -0.0006

Original:

...Ensure there are times when positive feedback is given for its own sake and resist the temptation to offer constructive criticism.

Reference:

...Sometimes, give good feedback only, and try not to add any criticism.

Simplified:

...Sometimes, give only good comments. Do not always give advice to help.

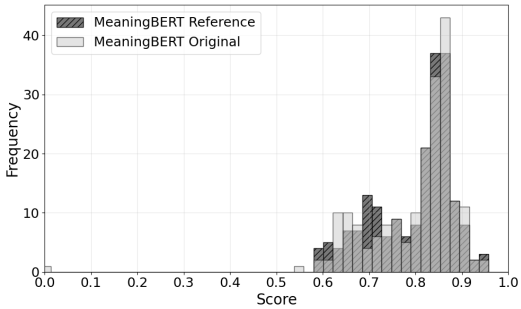

Figure 1: MeaningBERT score distributions of system outputs.

While not all poor simplifications stem from A2-level constraints, our analysis shows that B1-targeted simplifications performed significantly better. When only B1 outputs are evaluated, performance improved across all metrics: MeaningBERT Original and MeaningBERT Reference scores rose from 0.797 to 0.829 and 0.816, respectively, and RMSE decreased from 0.632 to 0.556.

Another issue about the performance of our method is its variance. As illustrated in Figure 1, the variance analysis reveals performance inconsistencies. For instance, the MeaningBERT Original score ranges from a high of 0.9424 down to an extremely low -0.0006, with a standard deviation of 0.1031. Similarly, the MeaningBERT Reference score varies widely from 0.9453 to 0.5826. This indicates that while our method can achieve very good results, it suffers from a lack of reliability in all cases.

Overall, the three-stage progressive simplification pipeline shows potential to achieve strong results, especially in B1-level targeted simplifications, without the need for costly model training and with a total cost of only $1.75. However, it can also lead to unsatisfactory outcomes due to the strict constraints imposed in the prompts.

5 Conclusion

This paper introduced a three-stage, CEFR-oriented text simplification method using syntactic simplification, lexical simplification, and elaboration through prompt engineering with GPT-4o. The approach avoids costly model training while producing controlled simplifications aligned with language proficiency levels. The results indicate a more promising performance at B1, where the output preserved meaning and coherence, but also reveal limitations at A2, where rigid grammar and vocabulary constraints often led to over-simplification or unnecessary substitutions. These findings suggest that future work should have greater flexibility in constraint design to balance accessibility with fidelity.

6 Limitations

Despite its potential, our approach has several limitations:

The most significant is its reliability. As discussed earlier, the system's performance exhibits high variance across inputs, indicating a lack of robustness.

Another practical limitation is the strict linguistic constraints imposed by the prompts, which often lead to over-simplification and meaning distortion. Additionally, the three-stage pipeline introduces a cumulative risk of semantic drift, which can further compromise meaning preservation.

7 Lay Summary

Text simplification helps make complex texts easier to read, which is especially useful for language learners and people with reading difficulties. Traditionally, this process is done manually by experts, which takes a lot of time and effort. While some automatic systems exist, they often do not adjust the output based on the reader's language level and are not designed around well-known frameworks like the Common European Framework of Reference for Languages (CEFR).

Training advanced systems for text simplification requires large datasets and expensive resources. This makes it difficult to build effective tools for many languages, especially those with limited data. Even when these systems work well, they often fail to match texts to specific learning levels, which is important in education and accessibility.

To address these challenges, we propose a three-stage simplification pipeline that directs a language model with carefully designed instructions instead of training it from scratch. Our approach simplifies grammar, replaces difficult words, and adds helpful explanations. This lightweight and flexible method can make reading materials more accessible and suitable for different language levels without the need for costly training or large datasets.

References

Wejdan Alkaldi and Diana Inkpen. 2023. Text simplification to specific readability levels. *Mathematics*, 11(9):2063.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020. Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, 46(1):135–187.

Fernando Alva-Manchego, Regina Stodden, Joseph Marvin Imperial, Abdullah Barayan, Kai North, and Harish Tayyar Madabushi. 2025. Findings of the TSAR 2025 shared task on readability-controlled text simplification. In *Proceedings of the Fourth Workshop on Text Simplification, Accessibility, and Readability (TSAR 2025)*, Suzhou, China. Association for Computational Linguistics.

Miriam Anschütz, Joshua Oehms, Thomas Wimmer, Bartłomiej Jezierski, and Georg Groh. 2023. Language models for German text simplification: Overcoming parallel data scarcity through style-specific pre-training. *arXiv preprint arXiv:2305.12908*.

Yuki Arase, Satoru Uchida, and Tomoyuki Kajiwara. 2022. CEFR-based sentence difficulty annotation and assessment. *arXiv preprint arXiv:2210.11766*.

Abdullah Barayan, Jose Camacho-Collados, and Fernando Alva-Manchego. 2024. Analysing zero-shot readability-controlled sentence simplification. *arXiv preprint arXiv:2409.20246*.

Fatih Bektaş, Kutay Arda Dinç, and Gülşen Eryiğit. 2024. LLMs for document-level text simplification in turkish foreign language learning. In *2024 9th International Conference on Computer Science and Engineering (UBMK)*, pages 1–5. IEEE.

Cambridge University Press. 2024. English grammar profile online. https://englishprofile.org/. Accessed: 2025-08-17.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Press Syndicate of the University of Cambridge.

Dengzhao Fang, Jipeng Qiang, Yi Zhu, Yunhao Yuan, Wei Li, and Yan Liu. 2025. Progressive document-level text simplification via large language models. *arXiv preprint arXiv:2501.03857*.

Matthew Honnibal, Ines Montani, and Sofie Van Landeghem. 2020. Boyd. *A. spaCy: industrial-strength natural language processing in Python*, 10.

Guanlin Li, Yuki Arase, and Noel Crespi. 2025. Aligning sentence simplification with ESL learner's proficiency for language acquisition. *arXiv preprint arXiv:2502.11457*.

Debora Nozza and Giuseppe Attanasio. 2023. Is it really that simple? Prompting large language models for automatic text simplification in italian. In *Proceedings of the 9th Italian Conference on Computational Linguistics (CLiC-it 2023)*, pages 322–333.

Oxford University Press. a. The American Oxford 3000. https://www.oxfordlearnersdictionaries.com/external/pdf/wordlists/oxford-3000-5000/American_Oxford_3000.pdf. Accessed: 2025-08-18.

Oxford University Press. b. The American Oxford 5000. https://www.oxfordlearnersdictionaries.com/external/pdf/wordlists/oxford-3000-5000/American_Oxford_5000.pdf. Accessed: 2025-08-18.

Ceren Özmen. 2019. *Yabancı Dil Olarak Türkçe Öğretiminde Metin Değiştirim Teknikleriyle Öykülerin Yeniden Oluşturulması*. Ph.D. thesis, Hacettepe University, Institute of Turkish Studies, Ankara, Turkey. Tez No: 588269.

Ceren Özmen and Ayşe Kıran. 2022. A study on text modification in the field of teaching turkish as a foreign language. *International Journal of Languages' Education and Teaching*, 10(2):19–42.

Renliang Sun, Hanqi Jin, and Xiaojun Wan. 2021. Document-level text simplification: Dataset, criteria and baseline. *arXiv preprint arXiv:2110.05071*.

Satoru Uchida, Shohei Takada, and Yuki Arase. 2018. CEFR-based lexical simplification dataset. In *Proceedings of International Conference on Language Resources and Evaluation*, volume 11, pages 3254–3258. European Language Resources Association.