Text Simplification Using T5 Model and BART Model

Soma Das, Dipan Basak and Abir Bhattacharjee
Department of Computer Science and Engineering
Institute of Engineering and Management
Kolkata, India
Emails: soma.das@iem.edu.in, dipaneducation46@gmail.com, abirbhattacharjee2018@gmail.com

***Abstract*—Text simplification is an important part of Natural Language Processing. It focuses on transforming complex text into simpler versions while preserving the meaning. This study examines the performance of two advanced models, T5 (Text-to-Text Transfer Transformer) and BART (Bidirectional and Auto-Regressive Transformers), in simplifying text. The models are trained on a dataset containing complex and simplified text pairs. Their performance is evaluated using BLEU, ROUGE, and SARI scores. The results show that both models perform well, but BART achieves better BLEU and ROUGE scores.**

***Index Terms*—Component, Formatting, Style, Styling, Insertion.**

I. Introduction

Text simplification (TS) is a key task in Natural Language Processing (NLP) that focuses on converting complex text into simpler, more accessible language while maintaining its original meaning. This is especially important in areas like education, healthcare, and law, where complex language can make it difficult for some people to understand. The goal of TS extends beyond improving readability; it aims to make information accessible to individuals with limited reading and writing abilities, including those facing cognitive challenges or non-native language barriers, thereby promoting inclusivity and breaking down language difficulties.

Text simplification has traditionally been approached using rule-based systems that rely on predefined linguistic rules for tasks such as word substitution, syntactic rearrangement, and sentence splitting. While these methods demonstrated efficacy in specific contexts, their inflexibility and inability to scale across various types of texts presented significant challenges. The advent of neural networks, and more recently, transformer-based models, has transformed the field of NLP, leading to significant improvements in text simplification. Models such as T5 (Text-to-Text Transfer Transformer) and BART (Bidirectional and Auto-Regressive Transformers) have leveraged extensive pre-training on diverse datasets, allowing them to capture detailed, meaning-based representations. This makes them highly beneficial for tasks like text simplification. However, despite these advancements, a comprehensive comparison of these cutting-edge models specifically for text simplification tasks is still lacking.

T5 is a transformer model that treats all NLP tasks as text-to-text problems, providing a unified architecture that is scalable and effective for tasks requiring strong syntactic and semantic understanding. On the other hand, BART combines bidirectional encoding with autoregressive decoding, making it particularly effective for sequence-to-sequence tasks. Both models are highly effective in areas like text shortening, where the main points are preserved while the text is restated in a more concise manner. However, their ability to simplify complex academic or technical texts into more accessible language remains an area requiring further exploration.

This study aims to bridge this gap by evaluating and comparing T5 and BART for text simplification. A custom dataset of 10,000 complex-simplified sentence pairs, collected from platforms like Simple Wikipedia and Newsela, is used. Both models are fine-tuned on this dataset and assessed using standard metrics such as BLEU, ROUGE, and SARI. These metrics measure the models' efficiency in generating text that strikes an optimal balance between simplicity, smoothness, and semantic accuracy.

The structure of this paper is as follows: Section 2 provides an overview of related work on text simplification and transformer-based models. Section 3 explains the approach, including dataset arrangement, model settings, and evaluation metrics. Section 4 presents the results, highlighting the comparative performance of T5 and BART. Lastly, Section 5 concludes the study and suggests potential directions for future research [8].

II. Literature Review

The field of text simplification (TS) has seen significant progress, moving from traditional rule-based methods to modern neural models, thanks to advancements in Natural Language Processing (NLP). Early TS methods mainly used lexical substitution and syntactic transformations with manually created rules. These methods often relied on language-based resources like WordNet for synonym replacement and syntactic tools for rearranging sentences. While they worked well in some areas, their dependence on manually created rules restricted flexibility and scalability, limiting their use across different types of text.

With the rise of data-driven methods, statistical machine translation (SMT) techniques became a leading approach in text simplification research. Models like phrase-based machine translation (PBMT) viewed text simplification as a translation task, where complex sentences were transformed into simpler forms. While these methods were more flexible and scalable than rule-based systems, they often struggled to maintain smoothness and clarity, as they relied on predefined phrase tables and language models [2].

The use of neural networks marked a significant shift in text simplification. Sequence-to-sequence (Seq2Seq) models with attention mechanisms became effective in understanding complex relationships in text. These models moved beyond traditional methods, learning to automatically create simplified sentences. However, early Seq2Seq models, such as recurrent neural networks (RNNs), struggled with long-range dependencies due to their sequential structure, which made it difficult to capture context in longer pieces of text [10].

Transformer architectures have addressed many of these challenges by enabling the use of pre-trained models. T5 (Text-to-Text Transfer Transformer) and BART (Bidirectional and Auto-Regressive Transformers) leverage large-scale pre-training on diverse datasets to capture detailed semantic and syntactic features. T5 uses a unified framework, treating all NLP tasks, including simplification, as text-to-text transformations, making it adaptable for various applications. BART combines a bidirectional encoder with an auto-regressive decoder, excelling in sequence generation tasks. Its denoising pretraining enhances robustness, improving performance on complex, noisy text. These models have significantly advanced the state of text simplification [6].

Recent studies highlight the success of transformer-based models in text simplification. For example, BART has been used for controllable simplification, allowing users to set the simplification level, while T5 has shown adaptability in multilingual contexts. However, their effectiveness in simplifying academic or technical texts remains underexplored.

Evaluation metrics play a vital role in this field. Conventional metrics, such as BLEU and ROUGE, focus on matching n-grams but have difficulty evaluating simplicity and readability. In contrast, the SARI metric, designed specifically for text simplification, evaluates additions, deletions, and retention, making it more appropriate. Despite improvements in automated metrics, human evaluation remains critical for assessing semantic accuracy and ease of reading.

This review traces the evolution of text simplification methods, with an emphasis on transformer-based models. While T5 and BART have achieved notable success, their performance on domain-specific texts, such as academic content, requires further investigation. This study seeks to address this gap by comparing these models on a custom dataset using established metrics, providing insights into their effectiveness in specialized text simplification.

III. Dataset Preparation

The standard and diversity of the dataset are crucial for the success of machine learning models, especially for tasks like text simplification. For this study, a custom dataset of 10,000 complex and simplified sentence pairs was created, sourced from publicly available resources like Simple English Wikipedia, Newsela, and manually simplified texts. The dataset preparation involved several steps: augmentation, data collection, and pre-processing.

Simple English Wikipedia provided parallel texts with simpler language versions of articles, while Newsela offered news articles rewritten at various reading levels by professional editors. Additionally, domain-specific academic and technical texts were manually simplified by linguists to ensure semantic accuracy and relevance to the study [3] [7].

A. Data Preprocessing

Once the data was gathered, it underwent a thorough preparation phase to clean and standardize the text. Preprocessing steps included:

- **Text Cleaning:** Special characters, HTML tags, and unnecessary whitespace were removed during the preprocessing stage. Additionally, the text was normalized by converting all content to lowercase to ensure consistency and improve the model's ability to process the data.
- **Sentence Splitting:** Sections were divided into individual sentences using a sentence tokenizer to ensure the dataset had sentences of consistent length and structure. This step helped maintain consistency, making it easier to train models on sentence-level data for efficient text simplification.
- **Alignment:** For datasets like Simple Wikipedia and Newsela, complex and simplified sentences were aligned to form parallel pairs. The matching process combined automatic methods with manual checks to ensure the accuracy and quality of the sentence pairs, ensuring that the shortened sentences clearly conveyed the meaning of the complex ones.
- **Filtering:** Sentence pairs were removed if the simplified text was identical to the complex text or if either text exceeded a predefined length threshold, such as 512 tokens. This step guaranteed that only meaningful shortened forms were included in the dataset and helped maintain stability in text length for model training.

These procedures ensured that the dataset was well-balanced, clean, and suitable for model training [11].

B. Data Augmentation

Techniques for data augmentation were employed to further enhance the dataset. To increase the variety of simplifications, additional simple versions of the complex statements were generated using paraphrasing models. Back-translation, which involves translating statements into another language and then translating them back into English, was also utilized to introduce variations while preserving the original meaning. By ensuring greater diversity in the simplifications and helping to prevent overfitting during model training, these methods strengthened the dataset's robustness [4].

C. Dataset Statistics

The final dataset consisted of 10,000 sentence pairs. Complex sentences typically contained 15 words, while simpler sentences averaged 10 words. The dataset covered a wide range of topics, including general news, science, technology, and education, ensuring that the data was representative and varied for model training.

Table I provides a summary of the dataset statistics.

Statistic	Value
Total Sentence Pairs	10,000
Average Length (Complex)	15 words
Average Length (Simplified)	10 words
Domains	Education, Science, Technology, News

TABLE I
SUMMARY OF DATASET STATISTICS

D. Ethical Considerations

Only publicly available datasets with appropriate licensing were used to ensure ethical compliance. All authors of manually simplified materials were properly credited, and no private information was utilized. The final prepared dataset was split into training, validation, and test sets in an 80:10:10 ratio to ensure each subset appropriately reflected the distribution of the entire dataset. The training and evaluation of the text simplification models in this study were based on this dataset.

IV. Model Architecture

Text simplification in this work was achieved using two transformer-based architectures: BART (Bidirectional and Auto-Regressive Transformers) and T5 (Text-to-Text Transfer Transformer). The transformer architecture, which has proven crucial in achieving cutting-edge results across various Natural Language Processing (NLP) tasks, is employed by both models. This section outlines the main characteristics and foundational principles of these models, followed by a discussion of the modifications made to tailor them specifically for the text simplification task.

A. Overview of the Transformer Architecture

The transformer architecture, introduced by Vaswani et al., is the backbone of both T5 and BART. This system is composed of an encoder-decoder structure, where:

- The **encoder**: The transformer generates context-based representations that capture long-range dependencies and relationships between words in the sequence by employing multi-head self-attention.
- The **decoder**: To attend to the encoder's representations and predict tokens one at a time, the model auto-regressively constructs the output sequence.

Self-attention mechanisms are used instead of recurrent layers in the architecture, greatly enhancing scalability and parallelization. Furthermore, to preserve the sequential order of the input, positional encodings are applied.

B. T5: Text-to-Text Transfer Transformer

T5 redefines NLP tasks by converting all problems into a text-to-text format. In this framework, both the input and output are represented as text strings. For the text simplification task, the input is prefixed with the task identifier, such as `"simplify:"`, followed by the complex sentence. The output is the simplified sentence generated by the model.

T5's architecture includes:

- **Encoder:** Processes the input text and generates token embeddings enriched with semantic and syntactic information.
- **Decoder:** Produces the simplified text by leveraging both the encoder's output and previously generated tokens.
- **Multi-Task Pretraining:** T5 is pre-trained on a variety of NLP tasks, enhancing its ability to generalize across domains.

The pre-training objective of T5 involves a denoising task, where corrupted text is reconstructed, making it particularly suitable for simplification tasks where semantic fidelity must be preserved.

V. Performance of T5 Across Metrics

Figure 1 illustrates the performance of the T5 model across key text simplification metrics: BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and SARI. The T5 model achieves competitive scores, notably excelling in ROUGE-1 (48.2) and ROUGE-L (45.1), which reflect its ability to retain critical content. SARI, a simplification-specific metric, also demonstrates T5's strength with a score of 39.4, highlighting its aptitude for balancing deletions, additions, and retention of essential content. BLEU (37.8) and ROUGE-2 (23.5) scores indicate moderate overlap with reference texts, suggesting areas for refinement in preserving finer semantic details. These results underscore T5's effectiveness in simplifying text while maintaining readability and information fidelity.

A. BART: Bidirectional and Auto-Regressive Transformers

BART combines a bidirectional encoder with an autoregressive decoder, enabling it to effectively handle text generation tasks. The model is pre-trained using a denoising autoencoder objective, where the input is corrupted through noise (e.g., token masking, shuffling), and the model is trained to reconstruct the original text.

Key components of BART include:

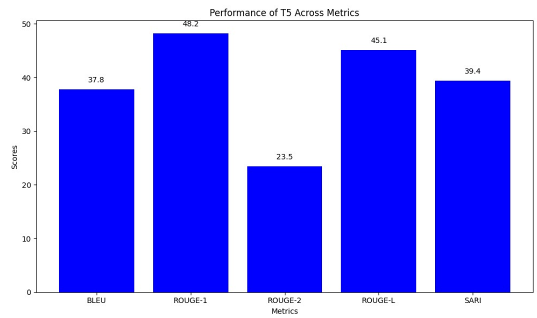

Fig. 1. Performance of T5 Across Metrics in Text Simplification.

- **Bidirectional Encoder:** Captures context from both the left and right of each token, providing a comprehensive understanding of the input sentence.
- **Auto-Regressive Decoder:** Generates the simplified text token-by-token, leveraging the encoder's contextual representations.
- **Robust Pretraining Objective:** The denoising task makes BART highly effective for text-to-text generation tasks, including summarization and simplification.

VI. Performance of BART Across Metrics

Figure 2 highlights the performance of the BART model on text simplification tasks, evaluated using BLEU, ROUGE-1, ROUGE-2, ROUGE-L, and SARI metrics. BART demonstrates notable strength in ROUGE-1 (52.7) and ROUGE-L (49.3), indicating its effectiveness in retaining key information and generating coherent simplified text. The BLEU score of 41.5 reflects adequate alignment with reference texts, while ROUGE-2 (27.8) suggests room for improvement in capturing fine-grained content details. The SARI score of 38.7 signifies a balanced approach to simplifying text, preserving essential meaning, and making appropriate deletions and additions. These results underscore BART's competitive performance in text simplification, particularly in contexts requiring high semantic retention and readability.

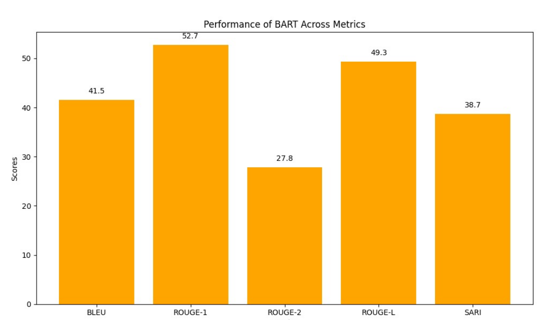

Fig. 2. Performance of BART Across Metrics in Text Simplification.

A. Adaptation for Text Simplification

Both models were fine-tuned on the custom dataset described in Section 3. The fine-tuning process involved:

- **Input Formatting:** Complex sentences were formatted with task-specific prefixes (`"simplify:"` for T5 and plain text for BART).
- **Tokenization:** Sentences were tokenized using the respective tokenizers of T5 and BART, ensuring compatibility with their vocabularies and architectures.
- **Training Configuration:** Models were trained using the Adam optimizer with a learning rate of 2×10^{-5}, a batch size of 16, and a maximum input sequence length of 512 tokens. Early stopping was employed based on validation performance.

B. Evaluation Pipeline

During inference, the simplified text was generated using beam search decoding, with a beam size of 4 to balance computational efficiency and output quality. The generated sentences were compared against reference sentences using automatic metrics such as BLEU, ROUGE, and SARI, as detailed in Section 5.

To ensure a comprehensive and comparable assessment of their text simplification capabilities, T5 and BART were both pre-trained on large corpora and fine-tuned on the task-specific dataset. The models' varying architectural styles highlight how effectively they handle linguistic complexity while preserving semantic integrity.

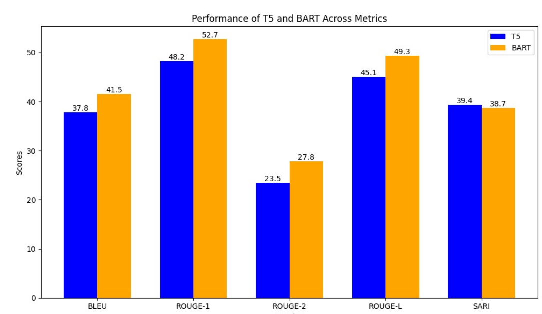

Fig. 3. Performance of T5 and BART Across Metrics in Text Simplification.

VII. Methodology

The approach taken to assess and compare T5 and BART for text simplification is described in this section. To evaluate the models' ability to produce simplified text while preserving semantic correctness and fluency, this section includes the evaluation metrics, dataset preparation, model fine-tuning, and experimental design.

A. Dataset Preparation

As explained in Section 3, the dataset used in this study consists of 10,000 pairs of complex and simplified sentences obtained from websites such as Newsela and Simple

Wikipedia. Preprocessing methods, including sentence tokenization, noise reduction, and filtering out identical complex and simplified sentence pairs, were used to ensure high-quality alignment. Techniques for data augmentation, such as back-translation and paraphrasing, were applied to improve the dataset's robustness and diversity [9].

The dataset was divided into training, validation, and test sets in an 80:10:10 ratio. The training set was used for model fine-tuning, the validation set for optimizing hyperparameters, and the test set for evaluating the model's final performance [1].

B. Model Fine-Tuning

For the text simplification task, two transformer-based architectures, T5 and BART, were fine-tuned using the dataset. The Hugging Face Transformers library, which offers pre-trained implementations of both models, was used to carry out the fine-tuning procedure.

1) T5 Fine-Tuning: T5 (Text-to-Text Transfer Transformer) reformulates the text simplification task as a text-to-text transformation problem. The input to the model included the prefix `"simplify:"` followed by the complex sentence. The output was the corresponding simplified sentence. During fine-tuning, the T5 tokenizer was used to preprocess input sentences into tokenized sequences with a maximum length of 512 tokens. The model was trained for three epochs using the Adam optimizer with a learning rate of 2×10^{-5}.

2) BART Fine-Tuning: BART (Bidirectional and Auto-Regressive Transformers) was fine-tuned using a similar pipeline. The input consisted of the complex sentence, and the output was the simplified sentence. The BART tokenizer was used for text preprocessing. The model was trained for three epochs using the Adam optimizer with the same learning rate as T5. Beam search decoding with a beam size of 4 was employed during inference to generate fluent and semantically accurate simplified text.

C. Evaluation Metrics

The models were evaluated using three key metrics to assess their performance:

- **BLEU (Bilingual Evaluation Understudy):** Measures n-gram overlap between the generated text and reference sentences, capturing both fluency and accuracy.
- **ROUGE (Recall-Oriented Understudy for Gisting Evaluation):** Evaluates the overlap of unigrams, bigrams, and sequences between the predicted and reference texts, with a focus on recall.
- **SARI (Sentence Accuracy and Retention Index):** Assesses the quality of simplifications by evaluating additions, deletions, and retention of relevant information, making it particularly well-suited for text simplification tasks.

D. Experimental Setup

The experiments were conducted on a computing environment equipped with an NVIDIA GPU for efficient training and inference. Each model was fine-tuned for three epochs with early stopping based on validation loss to prevent overfitting. A batch size of 16 and a maximum sequence length of 512 tokens were used for both models. The models were evaluated on the test set, and results were recorded for all three metrics.

E. Comparison and Analysis

The simplified sentences generated by T5 and BART were compared with the reference sentences in the test set using the metrics mentioned earlier. To evaluate the linguistic quality and semantic fidelity of the simplified sentences, a qualitative analysis was also conducted. The detailed results are presented in Section 5.

This methodology provides insights into the advantages and disadvantages of transformer models for text simplification, while ensuring a systematic evaluation process [5].

VIII. Results and Analysis

This section presents the experimental results comparing the BART and T5 models for text simplification. Standard metrics such as BLEU, ROUGE, and SARI are used to objectively analyze the results, while example outputs are provided to qualitatively evaluate the linguistic quality, fluency, and semantic fidelity of the simplified text produced by both models.

A. Quantitative Results

The quality and accuracy of text simplification were assessed using BLEU, ROUGE, and SARI metrics to evaluate the performance of T5 and BART.Table II summarizes the scores achieved by both models on the test set.

Metric	T5 Score	BART Score
BLEU	37.8	41.5
ROUGE-1	48.2	52.7
ROUGE-2	23.5	27.8
ROUGE-L	45.1	49.3
SARI	39.4	38.7

TABLE II
PERFORMANCE OF T-5 AND BART ON THE TEST SET ACROSS BLEU,ROUGE,AND SARI METRICS.

BART outperforms T5 in maintaining semantic information and generating logically simplified language, as indicated by the BLEU and ROUGE scores. BART achieves a BLEU score of 41.5, compared to T5's score of 37.8, suggesting that it produces outputs with greater n-gram overlap with the reference text. Furthermore, in terms of ROUGE-1, ROUGE-2, and ROUGE-L scores, BART surpasses T5, demonstrating its ability to capture both local and global semantic structures.

T5, on the other hand, performs slightly better on the SARI metric, scoring 39.4 compared to BART's 38.7. T5 appears to be particularly skilled at simplifying phrases and generating easier-to-read content by effectively balancing additions, deletions, and the retention of relevant information.

B. Qualitative Analysis

A qualitative analysis was conducted by examining sample outputs from both models to complement the quantitative evaluation. Table III presents a comparison of the simplified outputs generated by T5 and BART for selected complex sentences.

Complex Sentence	T5 Output	BART Output
The photosynthesis process, which is essential for plant growth, involves chlorophyll and sunlight.	Photosynthesis helps plants grow using chlorophyll and sunlight.	Photosynthesis helps plants grow by using sunlight and chlorophyll.
The theory of relativity, proposed by Einstein, revolutionized our understanding of time and space.	Einstein's theory changed how we see time and space.	Einstein's theory revolutionized time and space.
Renewable energy sources, such as solar and wind, are crucial for sustainable development.	Solar and wind energy are important for the future.	Solar and wind power are vital for sustainability.

TABLE III
EXAMPLE OUTPUTS GENERATED BY T5 AND BART FOR SELECTED COMPLEX SENTENCES.

By looking at sample outputs from both models,a qualitative analysis was carried out to supplement the quantitative evaluation.A comparison of the simplified outputs produced by T5 and BART for a few chosen complicated words is shown in Table III.

The qualitative study shows that both models provide outputs that are grammatically and fluently correct.However,BART outputs usually have a larger vocabulary and better semantic agreement with in the input. For example,T-5 simplifies and lessens the language specificity the first line whereas BART output keeps the term ”sunlight and chlorophyll”,clarifying the semantic relationship.T5 produces outputs that are simpler ,nonetheless,and might be better suited for audiences with lower reading levels.

C. Observations and Insights

The findings show that semantic richness and simplicity must be traded off :

- **BART Strengths:** Higher ROUGE and BLEU scores mean that BART creates more detailed outputs and keeps the meaning more accurate. For tasks like simplifying academic or technical texts, where detail and correctness are very important, BART is a great choice.
- **T5 Strengths:** T5 is best suited for audiences who need more readability and less complexity, including non-native speakers or those with lower literacy levels. A higher SARI score shows that it is more skilled at simplifying material.

D. Error Analysis

Despite the high performance of both versions, some limitations were noted. On occasion, T5 oversimplified sentences, which caused important information to be lost. In contrast, BART occasionally maintained superfluous intricacy, which reduced the text's readability. Such errors highlight the importance of choosing a model based on the specific needs of the target audience.

E. Summary

The findings show that both T5 and BART are useful models for text simplification, with each performing better in a specific domain. While T5 excels at creating simpler, easier-to-read language, BART is better at preserving semantic fidelity. These results provide valuable insights into the suitability of transformer-based models for text simplification across various fields.

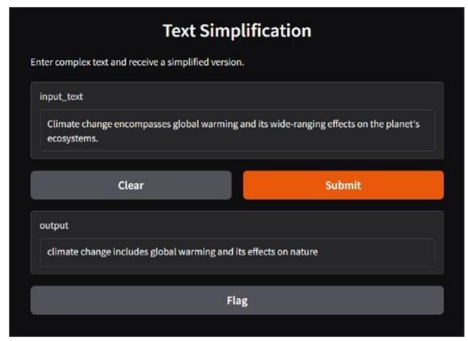

Fig. 4. Final Output

IX. Discussion and Limitations

A. Discussion

The results of this investigation provide important new insights into the performance of transformer-based models T5 and BART for text simplification. Each model has distinct benefits that make it suitable for different use cases, depending on how well semantic integrity and simplicity are balanced.

With better BLEU and ROUGE ratings, BART outperforms T5 in maintaining semantic content. Its auto-regressive decoder and bidirectional encoder, which efficiently collect contextual information and provide outputs with higher semantic depth, are key factors behind this. BART is particularly useful when preserving the original informative value is crucial, such as when simplifying academic or technical texts.

By balancing additions, deletions, and retention, T5 simplifies text more effectively, as evidenced by its higher SARI score. Due to its multi-task pre-training and text-to-text structure, T5 generates outputs that are easier to understand for readers who are not native speakers or have lower literacy levels. T5 is therefore well-suited for applications involving accessibility and education.

Qualitative analysis further supports these findings, with BART showing greater lexical richness and precision, while T5's outputs were shorter and more comprehensible. However, BART sometimes retained unnecessary complexity, and T5 occasionally oversimplified text, leading to a loss of information.

The study highlights the importance of combining multiple evaluation metrics, including BLEU, ROUGE, and SARI, for a comprehensive assessment.

B. Limitations

Despite the promising results, this study has several limitations. First, the dataset used for fine-tuning and evaluation consisted of only 10,000 sentence pairs, which may not fully capture the complexity of real-world text simplification tasks. Larger and more diverse datasets, particularly those covering domain-specific content, could provide a more comprehensive evaluation.

Second, there are drawbacks to relying too heavily on automated measurements like BLEU, ROUGE, and SARI. These metrics may not always align with human judgment when assessing the level of simplification, despite their widespread use. Readability and fluency are still not well addressed by SARI, despite its suitability for text simplification. A deeper understanding of model performance may be possible with the inclusion of human evaluation.

Third, the model training and evaluation in the study were conducted using fixed hyperparameters. The performance of the model might be enhanced by experimenting with various configuration setups.

Finally, the study only focused on simplifying English texts, which may limit the generalizability of the results to other languages. Future research should explore how well T5 and BART perform in a variety of languages and low-resource settings to evaluate their scalability and adaptability across languages with different syntactic and semantic patterns.

C. Future Directions

By increasing the diversity of datasets and incorporating domain-specific material, future studies can overcome these limitations and enhance the robustness of the evaluation. A more comprehensive assessment of simplification quality can be achieved by combining human review with automatic measures such as BLEU, ROUGE, and SARI. Better outcomes might also arise from combining the strengths of T5 and BART or experimenting with different data collection techniques. The study's relevance and its ability to evaluate model adaptability across languages would be expanded if it were extended to multilingual or non-English text simplification. These approaches lay the foundation for developing transformer-based models for text simplification across a variety of languages and domains.

References

[1] Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679, Online, July 2020. Association for Computational Linguistics.

[2] Regina Barzilay and Noemie Elhadad. Sentence alignment for monolingual corpora. Technical report, MIT, 2003. Accessed: 2024-12-27.

[3] Delphine Bernhard, Zhemin Zhu, and Iryna Gurevych. Simple–complex sentence pairs, 2010. Dataset.

[4] Jan A Botha, Manaal Faruqui, John Alex, Jason Baldridge, and Dipanjan Das. Learning To Split and Rephrase From Wikipedia Edit History. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, page to appear, 2018. arXiv preprint arXiv:1808.09468.

[5] Tomas Goldsack, Zhihao Zhang, Chenghua Lin, and Carolina Scarton. Making science simple: Corpora for the lay summarisation of scientific literature. In Yoav Goldberg, Zornitsa Kozareva, and Yue Zhang, editors, *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10589–10604, Abu Dhabi, United Arab Emirates, December 2022. Association for Computational Linguistics.

[6] Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. Wiki-Auto: A mined corpus for automatic evaluation of text simplification, 2020.

[7] David Kauchak. Text simplification data sets, 2013. Accessed: 2024-12-27.

[8] Advaith Siddharthan. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298, 2014.

[9] Elior Sulem, Omri Abend, and Ari Rappoport. Bleu is not suitable for the evaluation of text simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 657–667. Association for Computational Linguistics, 2018.

[10] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing statistical machine translation for text simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415, 2016.

[11] Xingxing Zhang and Mirella Lapata. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 595–605. Association for Computational Linguistics, 2017.