Multilingual Text Simplification and its Performance on Social Sciences Coursebooks

Tim Schlippe and Katharina Eichinger

IU International University of Applied Sciences, Germany.

Email: tim.schlippe@iu.org;

Abstract

Text simplification is an essential task in today's society. It has the potential to help minorities get information, the broad masses have access to higher education, and assist in learning a new language. Through text simplification, we achieve an inclusive world with less language barriers. However, for training automatic artificial intelligence based systems for text simplification, no much high-quality text resources exist—particularly for various languages. Consequently, we investigate leveraging data from other languages to enhance the performance of text simplification systems. As simplification helps better understanding learning content, we finally apply our best German and English systems—which are based on a state-of-the-art Text-to-Text Transfer Transformer (T5) [1]—on sentences of German and English social sciences coursebooks from our university and perform a human evaluation in a survey. Furthermore, a comparison of text simplified with our best systems to text simplified with ChatGPT demonstrates that our English model significantly outperforms ChatGPT in 3 of 5 criteria evaluated, while our German model even outperforms ChatGPT in 4 of the 5 criteria.

Keywords: text simplification, AI in education, natural language processing, multilingual, plain language, simple language, ChatGPT

1 Introduction

Despite the global increase in literacy rates, nearly 15% of the world's population still struggle with reading and writing [2]. Furthermore, text presented in various forms such as coursebooks, public announcements, legal documents, and medical letters is not always composed in a way that is easily comprehensible—even for people who can read and write well. The moment technical terminology is introduced into a text, its complexity level automatically increases. The problem is exacerbated when individuals with restricted language proficiency or cognitive disabilities attempt to read the text. This increases the issue of this group being unable to access crucial information [3]. Even scientific texts—which are inherently more complex than average—are being written in an increasingly complicated way. [4] analyzed the evolution of texts over the past 130 years. They analyzed over 700,000 English study abstracts and assessed readability levels. Their findings indicated an increase in the average number of syllables per word since the late 19th century, and longer sentence structures since the 1960s. Furthermore, the research revealed an expansion in the use of technical terminology, including words deemed non-essential to the subject matter by the researchers. The issue of text readability extends beyond social concerns and also has economic implications. Effective communication of information requires both the sender and the recipient to possess adequate reading skills to comprehend texts [5]. This involves not only the reader's ability but also the text being composed in a clear and comprehensible manner. Consequently, there is a pressing need for developing strategies and tools to simplify these texts, ensuring that they are accessible and easily understood by a wider audience, no matter how their reading abilities are or how familiar the reader is with specialized terminology.

There are around 6,500 to 7,000 languages spoken worldwide, according to [6]. Europe is home to approximately 150 languages. The precise count is difficult to determine due to underexplored areas of linguistics and the absence of clear criteria for distinguishing between individual languages. Text simplification is not only a comparatively niche topic, but becomes much more difficult with so many languages involved. As a result, there is a scarcity of training data for this task, and support is limited to a few languages. Moreover, even for languages that have text corpora the quality of data may not be optimal due to the intricate process of corpus creation, which is time-consuming and potentially costly. Digital resources for text simplification from government publications, Wikipedia, disability facilities, and other sources are relatively scarce, and parallel content of text and its simplified version is rarely available [7]. This further complicates the automated generation of corpora and raises the following two questions:

- How can we automatically enrich training corpora for text simplification without human effort?
- How can we leverage multilinguality for text simplification?

Therefore, our goal was to generate synergies with other languages and to use a cross-lingual approach to train a natural language processing (NLP) model that requires only a small manually created high-quality dataset translated into many other languages before training. This turns out to be a relatively straightforward and resource-saving approach, as only the Google Translate API[1] with generous free tier is required. In addition to these benefits, the resulting models perform well as shown in Section 4. In human evaluation, our best German and English model even outperforms the simplifications generated by ChatGPT[2] as shown in Section 5. OpenAI's

[1] https://pypi.org/project/googletrans
[2] https://chat.openai.com/chat

ChatGPT has become one of the most commonly utilized chatbots. The fact, that the chatbot was able to reach over 1 million users in only five days [8] undermines this statement.

In the following section, we will describe related work regarding text simplification. Section 3 shows the pipeline to artificially create multilingual training text pairs for enhanced text simplification. The experiments and the results of our text simplification systems trained with target language data and varying amounts of translations are presented in Section 4. To further evaluate our best text simplification systems in the field of education, we conducted a survey which we will delineate in Section 5. We will conclude our work in Section 6 and suggest further steps.

2 Related Work

There are several techniques to transferring or machine-translating text from one language to another—or in the our case—to a simplified language (also known as *plain language* or *simple language*). First, we will discuss the traditional approaches: rule-based and statistical. Then, we will describe the modern approaches which are based on artificial intelligence and deep learning, more specifically have a Transformers architecture. Finally, we will discuss evaluation metrics and corpora used for this task.

2.1 Traditional Approaches to Text Simplification

In rule-based machine translation (RBMT) like [9], words or phrases are translated from the source language to the target language using a set of predefined rules. RBMT necessitates the involvement of proficient professionals who possess knowledge in both the source and target languages to create crucial syntactic, semantic, and morphological rules for translation. [10] developed a rule-based text simplification system for German. They implemented some of the rules defined by Maaß [11] to achieve simplification at the word, character, text, and sentence levels. Rule-based systems appear outdated in comparison to modern NLP approaches. The reliance on manually written rulebooks, in particular, is a drawback since they require a significant amount of time to create and may not encompass the full complexity of natural language.

Consequently, the focus of [12] was on utilizing statistical machine translation (SMT) to simplify words or phrases (lexical simplification) and, to a lesser degree, simplifying the input using rewrite rules (syntactic simplification). By utilizing machine learning, SMT converts the task of machine translation into a computationally solvable problem [13]. SMT algorithms acquire translation skills by analyzing a vast collection of human-generated translations. This technique involves aligning the source language text with its corresponding translation in the target language, resulting in the creation of a bilingual corpus of texts. By analyzing bilingual corpora, SMT algorithms identify patterns and connections, and use these findings to construct statistical models that enable automated translation. [12] modified the core components of SMT to fit the needs of text simplification. In this context, they also developed the SARI Score, which we will use additionally for the evaluation of our text simplification systems. [14] employed word embeddings and word frequencies in their phrase tables for

SMT-based text simplification. According to [15] and [16], SMT systems have the ability to learn and recognize common word order patterns and grammatical structures. But they may not always encompass the complete range of syntactic or grammatical principles within a language.

2.2 Neural Network Based Approaches to Text Simplification

[17] suggest a long short-term memory (LSTM) encoder-decoder model to perform sentence-level text simplification. [18] describe a new text simplification system called TST (Text Simplification by Tagging), which leverages pre-trained Transformers. [19] propose an approach to sentence simplification using a combination of an encoder-decoder model and deep reinforcement learning. The objective of their model is to identify the most effective sentence simplification while maximizing a reward function that prioritizes simplicity, fluency, and preservation of the original meaning. [19] state that deep reinforcement learning aids in incorporating prior knowledge into the task of sentence simplification. SimpLex is an architecture for text simplification, that uses "either word embeddings (i.e., Word2Vec) and perplexity or sentence Transformers (i.e., BERT, RoBERTa, and GPT2) and cosine similarity" to generate simplified English sentences [20]. The researchers discovered that Transformer models achieved a better SARI performance than other models. On the other hand, the models based on word embeddings achieved the greatest reduction in perplexity.

For our text simplification model, we decided to use a Transformer architecture. In particular, the self-attention mechanism seems to be crucial for our task: Text simplification often involves understanding long and complex texts. Frequently, there are dependencies within sentences that run the whole length of the sentence. Transformers are designed to handle these dependencies and are also able to focus on different aspects of the input. This capability enables a Transformer model to extract the significant parts of the input and condense the output accordingly. Additionally, Transformer models exist that offer multilingual support, which is essential for our cross-lingual strategy.

2.3 Cross-Lingual Approaches to Text Simplification

Leveraging the *cross-lingual transfer* of a multilingual NLP model leads to improvements in several NLP tasks. For example, [21] are able to improve automatic short answer grading by augmenting training data with machine-translated task-specific data for fine-tuning. However, only [22] attempted sentence simplification using a cross-lingual zero-shot approach. Their model employs an encoder-decoder architecture trained in a multilingual setting. To the best of our knowledge, there are no other approaches so far to apply cross-lingual effects to text or sentence simplification.

2.4 Corpora and Evaluation Metrics for Text Simplification

Most languages do not have the benefit of having large amounts of simplified text like the English Simple Wikipedia[3]. Developing parallel corpora for automatic text simplification systems can be quite labor-intensive and time-consuming. Many languages,

[3] https://simple.wikipedia.org

particularly those that are less widely spoken, do not have the necessary resources to invest in creating parallel corpora for text simplification. Most researchers to date work with corpora such as WikiLarge[4]. However, these corpora contain a large amount of simplifications that are not accurate and the simplifications are not always sufficient [12, 14, 23]. Figure 1 demonstrates an example from WikiLarge which illustrates the problem of inaccurate text data.

Original text:

Geneva -LRB- , ; , ; , ; ; -RRB- is the second-most-populous city in Switzerland -LRB- after Zürich -RRB- and is the most populous city of Romandie -LRB- the French-speaking part of Switzerland -RRB- .

Simplification:

The city 's main newspaper is the Tribune de Genève , a daily newspaper founded on 1 February 1879 by James T. Bates , with a readership of about 187,000 .

Fig. 1: Problem of Inaccurate Text Data in WikiLarge.

Due to the weaknesses of most corpora, we decided to use a small but high-quality corpus: The ASSET (Abstractive Sentence Simplification Evaluation and Tuning)[5] corpus. It "consists of 23,590 human simplifications associated with the 2,359 original sentences from TurkCorpus [12] (10 simplifications per original sentence)" and was created primarily for validation and testing [24]. From this dataset, we used 1,000 sentence pairs with only 1 simplification each for generating translations and fine-tuning our multilingual model. Additionally, we extracted 500 sentence pairs with 10 simplifications each from ASSET to test and evaluate our systems. Figure 2 shows an excerpt of the corpus. For our multilingual text simplification experiments, we expanded it to a larger multilingual corpus by translating it into other languages.

BLEU (Bilingual Evaluation Understudy) [25] is considered as one of the standard evaluation metrics in the field of machine translation and text generation. The BLEU score measures how similar the machine-translated text is to a set of reference translations [25]. However, [26] criticize that BLEU is inappropriate for evaluating simplified texts as it is not suitable for evaluating split sentences. Long sentences are often split into several sentences when they are simplified. Furthermore, BLEU often has a negative correlation with simplicity: Simple sentences are subject to penalties [26]. In addition, BLEU favors rather conservative models that make few changes to the input [12]. Due to these facts, fewer and fewer publications have reported BLEU as their main evaluation criterion [27–29]. Due to the shortcomings of BLEU in the area of text simplification, a new score was developed: SARI (System Output against References and against the Input Sentence) [12] compares the simplified sentences against references and the source sentence. It measures the goodness of words that are

[4] https://github.com/XingxingZhang/dress
[5] https://github.com/facebookresearch/asset

Original text:

Following his success with Star Trek, he supplemented his income and showed continued support for his fans by making numerous public appearances.

Simplification 1:

After Star Trek's success, his income increased and he supported fans by appearing in public.

Simplification 2:

After Star Trek, he made money and supported his fans by making public appearances.

Simplification 3:

Following Star Trek's success, he supplemented his income and showed support for fans by making public appearances.

Simplification 4:

He saw triumph with Star Trek. He made public appearances and supported his fans. It offered additional income.

Fig. 2: ASSET Corpus with with High-Quality Simplifications.

added, deleted and kept by the system. Higher SARI scores reflect better text simplifications. [12] found that SARI scores correlate positively with estimates of human evaluations. [30] introduced EASSE (Easier Automatic Sentence Simplification Evaluation)[6], a Python module that provides a command line interface for accessing popular automatic metrics in the evaluation of sentence simplification [30]. Since the package was initially released, more and more studies use EASSE to evaluate their SARI score [18, 31–33]. Due to its fit to text simplification and its popularity, we evaluate our system performances as well with SARI, as shown in Section 4.

[34] studied the correlation of phrase-level evaluation metrics with human judgments on different self-curated datasets. They examined various text simplification approaches and concluded that most metrics fit to test the output of neural sequence-to-sequence models, but recommend a final manual evaluation [34].

Consequently, in addition to SARI, we perform a human evaluation with different evaluation criteria for selected simplifications generated by our best cross-lingual text simpification systems, as described in Section 5.

3 Multilingual Text Simplification

While the focus of related work has been on text simplification systems built and applied for only one language, the focus of this paper is on leveraging a multilingual

[6] https://github.com/feralvam/easse

NLP model to be able to use machine-translated texts as more training data and for the application on multiple languages in the context of *cross-lingual transfer*.

Google's Flan T5 Transformer [35] is such a multilingual NLP model pre-trained from monolingual corpora in 101 languages which can be adapted to a certain task with task-specific labeled text data in 1 or more languages (*transfer learning*) and then perform this learned task in other languages (*cross-lingual transfer*). Flan stands for "Finetuning language models"[7]. It is an open-source Transformer-based architecture that uses a text-to-text approach for NLP. Flan T5 is the improved successor of Google's T5 (Text-to-Text Transfer Transformer)[8] and mT5 (multilingual T5)[9] [35]. Due to our limited computing capacities, we used the version *Flan-T5-Base*.

Figure 3 shows our pipeline to artificially create multilingual training text pairs for enhanced text simplification and evaluating the resulting systems. The pipeline's steps are as follows:

1. To enrich our training data, we generate additional artificial training data by machine-translating sentences from our original English text simplification corpus to other languages.
2. We fine-tune our NLP model using the English training data plus the translations to build a multilingual text simplification system.
3. We apply our multilingual text simplification system to the test data.
4. We evaluate the quality of the multilingual text simplification system output.

Flan-T5-Base is pre-trained for a number of tasks that are invoked by adding a prefix to the associated prompt. Since text simplification is not a pre-trained task, we decided to fine-tune and evaluate the following three tasks:

- *Translate*: treats the task as a monolingual translation, i.e. the output is a rendition of the input text in the same language.
- *Paraphrase*: tells the model to rephrase the input while preserving the meaning.
- *Summarize*: causes the output to tend to be shorter, but still capture the critical aspect of the sentence.

In the next section, we will show the SARI performances of text simplification systems trained for these three tasks with the 1,000 ASSET sentence pairs in the target language and varying amounts of translations.

4 Experiments and Results

In this section we will describe the experiments and the results of our text simplification systems trained with target language data and varying amounts of translations.

[7] https://huggingface.co/docs/Transformers/model_doc/flan-t5
[8] https://huggingface.co/docs/Transformers/model_doc/t5
[9] https://github.com/google-research/multilingual-t5

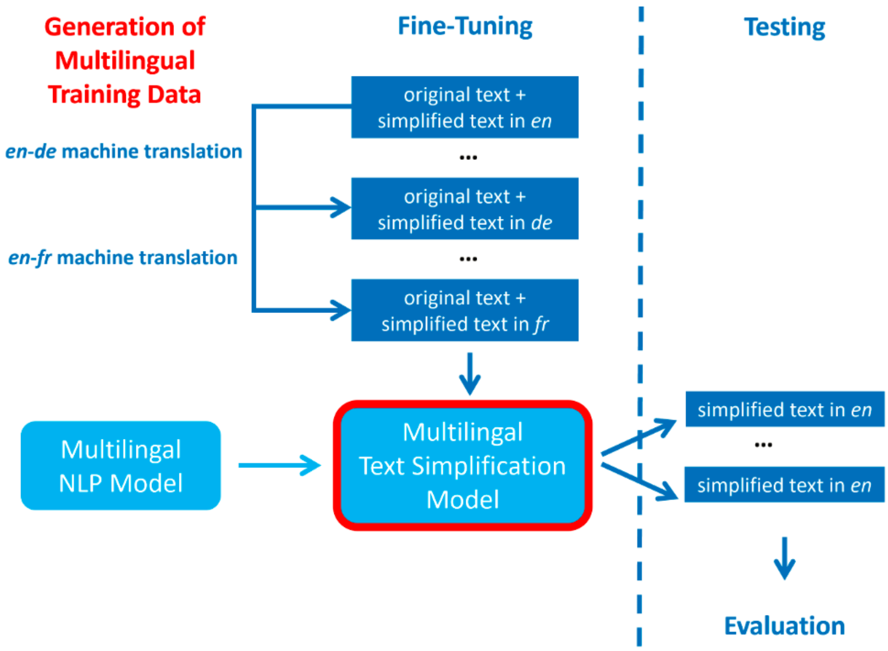

Fig. 3: System Overview.

4.1 Data for Training and Testing

For our experiments we built text simplification systems for English (*EN*), German (*DE*), Spanish (*ES*), French (*FR*) and Portuguese (*PT*) which we trained and fine-tuned with various texts in 40 languages. To retrieve the translations for fine-tuning, our 1,000 sentence pairs from the the high-quality English ASSET corpus [24] were translated into the other languages using the Google Translate API for Python. Additionally, the 500 ASSET sentences with their 10 simplifications each were machine-translated the same way to obtain sentences for testing our systems. Our reasons to use Google Translate for the translations are that the Python library Googletrans[10] provides a free service with a very generous limit. Furthermore, Google Translate can be considered a good translator: For the translation from English to the other four target languages, the following BLEU scores have been reported by [36]:

- From English to German: 0.81
- From English to Spanish: 0.80
- From English to Portuguese: 0.91
- From English to French: 0.88

[10] https://pypi.org/project/googletrans

4.2 Experiments with Tasks and Artificial Training Data

Tables 4, 5 and 6 visualize the SARI performances of *EN* text simplification systems trained for these three tasks with 1,000 sentence pairs in the target language and varying amounts of translations. The tables show that for *EN* the values for the *Paraphrase* task are generally higher than for *Translate* and *Summarize*. It can also be seen that the SARI scores with the artificial training data generated by the translations are better than the baseline systems trained with 1,000 *EN* text pairs. There is a tendency that the systems get better with more languages, however there are fluctuations: More training data through the translations does not necessarily mean better SARI scores. For example, for 2k training sentences, we observe that the impact of a translation on system performance depends on the language.

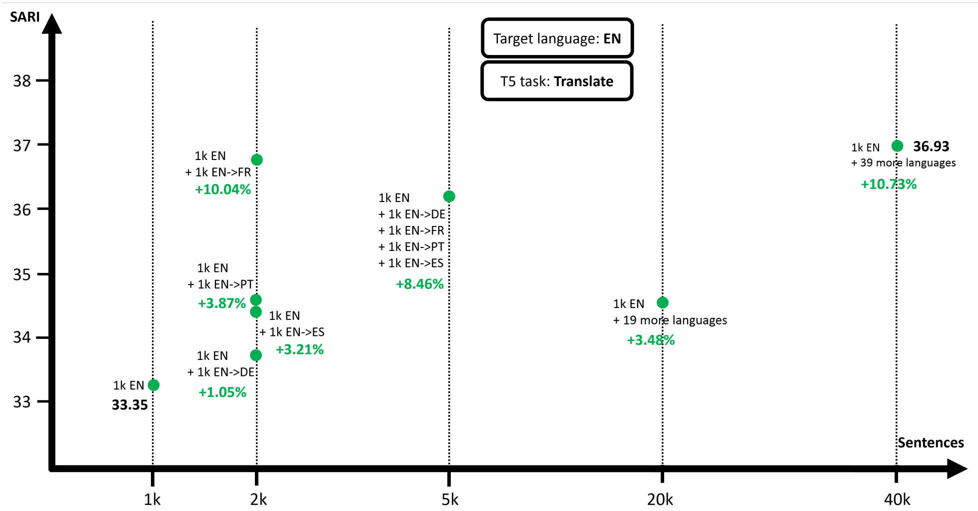

Fig. 4: English Text Simplification with T5's Translate Task.

As with *EN*, we trained text simplification systems for *DE*, *FR*, *ES*, and *PT*. Due to the limitations of this paper, we are not able present the detailed results as in Figure 4, 5 and 6. Consequently, we summarize the SARI scores of the baseline systems, the best systems and the relative improvement to the best baseline system in Table 1.

Table 1 demonstrates that for all five target languages, the task *Paraphrase* trained with 1k training sentence pairs from the target languages plus 39k translated sentence pairs performs best. The task *Paraphrase* also leads to best baseline systems. $Paraphrase_{1kEN+39moreLanguages}$ achieves with 9.28% relative most improvement compared to the best baseline system, followed by $Paraphrase_{1kDE+39moreLanguages}$, $Paraphrase_{1kES+39moreLanguages}$, and $Paraphrase_{1kPT+39moreLanguages}$ with 5.23% relative improvement each and $Paraphrase_{1kFR+39moreLanguages}$ with 2.28% relative improvement.

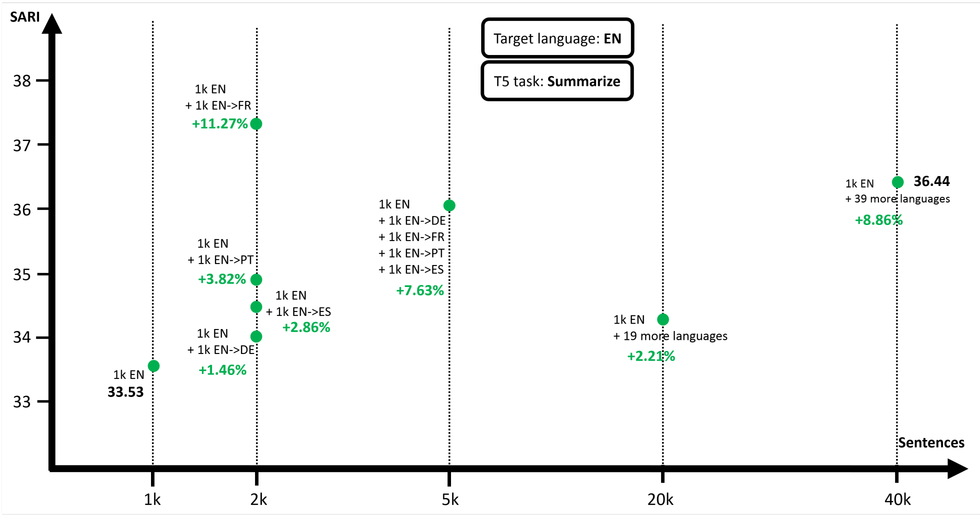

Fig. 5: English Text Simplification with T5's Summarize Task.

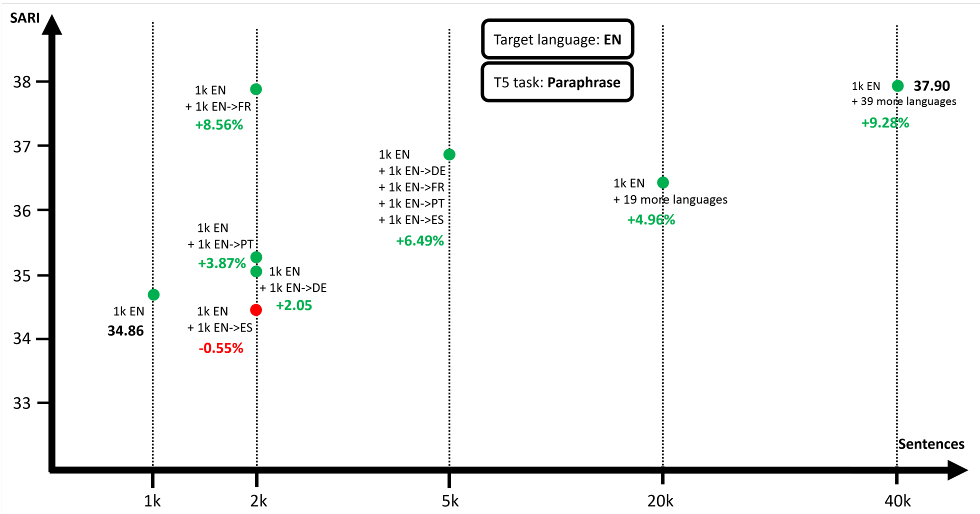

Fig. 6: English Text Simplification with T5's Paraphrase Task.

5 User Study with Social Sciences Coursebooks: Our Multilingual Text Simplification vs. ChatGPT

Since we as a university that teaches in German and English are always looking for ways to make our education more accessible to students, we wanted to find out how

		SARI
EN baselines	$Translate_{1kEN}$	33.35
	$Summarize_{1kEN}$	33.53
	$Paraphrase_{1kEN}$	34.68
Best *EN* system	$Paraphrase_{1kEN+39moreLanguages}$	37.90
Δ improvement to best *EN* baseline		**9.28%**
DE baselines	$Translate_{1kDE}$	28.13
	$Summarize_{1kDE}$	27.69
	$Paraphrase_{1kDE}$	27.36
Best *DE* system	$Paraphrase_{1kDE+39moreLanguages}$	29.60
Δ improvement to best *DE* baseline		**5.23%**
FR baselines	$Translate_{1kFR}$	22.70
	$Summarize_{1kFR}$	22.70
	$Paraphrase_{1kFR}$	22.79
Best *FR* system	$Paraphrase_{1kFR+39moreLanguages}$	23.31
Δ improvement to best *FR* baseline		**2.28%**
ES baselines	$Translate_{1kES}$	30.6
	$Summarize_{1kES}$	31.16
	$Paraphrase_{1kES}$	30.78
Best *ES* system	$Paraphrase_{1kES+39moreLanguages}$	32.79
Δ improvement to best *ES* baseline		**5.23%**
PT baselines	$Translate_{1kIT}$	31.32
	$Summarize_{1kPT}$	31.44
	$Paraphrase_{1kPT}$	31.93
Best *PT* system	$Paraphrase_{1kPT+39moreLanguages}$	33.60
Δ improvement to best *IT* baseline		**5.23%**

Table 1: Baseline Systems vs. Best Text Simplification Systems.

people respond to text from coursebooks that has been simplified using our best multilingual text simplification systems. To do this, we took excerpts from the coursebooks “Soziale Arbeit” (social work) and “Intercultural and Ethical Decision-Making”—a German and an English coursebook from the field of social sciences and had them give us feedback in a study. Since ChatGPT is becoming more famous and can simplify texts in German and English, we showed our participants the simplifications provided by ChatGPT for comparison. The prompt we used to obtain a simplified German sentence from the coursebook “Soziale Arbeit” was “Vereinfache den folgenden Satz:”. We achieved the simplification of an English sentence from the coursebooks “Intercultural and Ethical Decision-Making” with ChatGPT using the prompt “Simplify the following sentence:”. Figure 7 displays an excerpt from the German coursebook with simplifications by our best *DE* text simplification system and ChatGPT.

Our questionnaire consisted of five text excerpts from the German coursebook and five text excerpts from the English coursebook, the simplification of our best system and the simplification by ChatGPT. We did not tell our participants which text was provided by our system and which text was generated by ChatGPT. Our participants evaluated the simplified texts with a score. The score range follows the rules of a forced choice Likert scale, which ranges from (1) *strongly disagree* to (5) *strongly agree*. In order to get detailed feedback about the simplified texts, we asked for the quality of the following criteria:

Original text:

Die freizeitorientierte Soziale Arbeit mit älteren und alten Menschen ist ein Handlungsfeld der Sozialen Arbeit, das darauf zielt, freie Zeit zu gestalten und in Gesellschaft anderer zu verbringen.
(English: Leisure-oriented social work with older and elderly people is a field of action of social work that aims at shaping free time and spending it in the company of others.)

Multilingual Text Simplification:

Die freizeitorientierte Soziale Arbeit mit älteren und alten Menschen ist ein Handlungsfeld der Sozialen Arbeit.
(English: Leisure-oriented social work with elderly and old people is a field of action of social work.)

ChatGPT's Text Simplification:

Die freizeitorientierte Soziale Arbeit mit älteren und alten Menschen bezieht sich darauf, Zeit in Gesellschaft anderer Menschen zu verbringen und gemeinsam Aktivitäten zu unternehmen, um die freie Zeit zu gestalten. Dies ist ein Bereich der Sozialen Arbeit.
(English: Leisure-oriented social work with the elderly and old people refers to spending time in the company of other people and doing activities together to organize free time. This is an area of social work.)

Fig. 7: Excerpt from the German Coursebook "Soziale Arbeit" with Simplification by our Best *DE* Text Simplification and ChatGPT.

- *content*: How well does the simplified text reproduce the original text in terms of content?
- *fluency*: How smoothly does the simplified text read?
- *comprehensibility*: How understandable is the simplified text?
- *grammar*: How correct is the simplified text in terms of grammar?
- *simplification*: How well is the text simplified?

105 people participated in our survey of which 84.8% live in Germany, 6.7% in Austria, 4.8% in Switzerland, 1.9% in the Netherlands, 1% in Luxembourg and 1% in Sweden. On a scale of 1 (*no knowledge*) to 5 (*native language*), 90.5% reported that German was their native language, 3.8% gave a score of 4 and 3 respectively, and 1% each gave a score of 2 or 1. On the same scale, 6.7% indicated that English was their native language, 51.4% gave a score of 4, 38.1% rated their English proficiency at 3, 2.9% at 2, and 1% at 1. 49.5% of participants were female, 44.8% male and 5.7% preferred not to state their gender. 11.4% of participants were 18-24 years old, 43.8% were 25-34 years old, 23.8% were 35-44 years old, 9.5% were 45-54 years old, 11.4% were 55-65 years old, and 11.4% were older than 65 years old.

Figure 8 and Figure 9 visualize the participants' average feedback on the excerpts from the German coursebook "Soziale Arbeit" (social work) simplified by our best

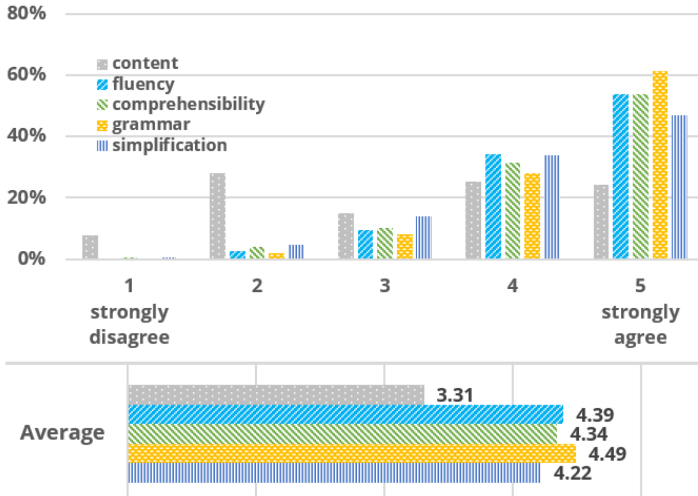

Fig. 8: Human Evaluation of *DE* Multilingual Text Simplification.

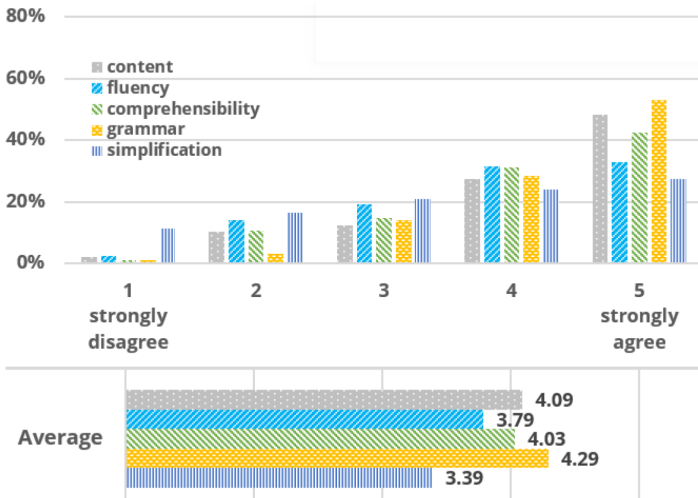

Fig. 9: Human Evaluation of ChatGPT's *DE* Text Simplification.

DE text simplification and ChatGPT. We see for our *DE* text simplification that the participants agree that the criteria *fluency*, *comprehensibility*, *grammar* and *simplification* are well covered. The comparison of Figure 8 and Figure 9 indicates that our system significantly outperforms ChatGPT in these criteria. Only for the question of how well the simplified text reproduces the original text in terms of *content* our system performs worse than ChatGPT, since it removes more content than ChatGPT for the sake of a better *simplification*. This negative correlation between *simplification* and *content* in a human evaluation is called the simplicity-adequacy tradeoff [37].

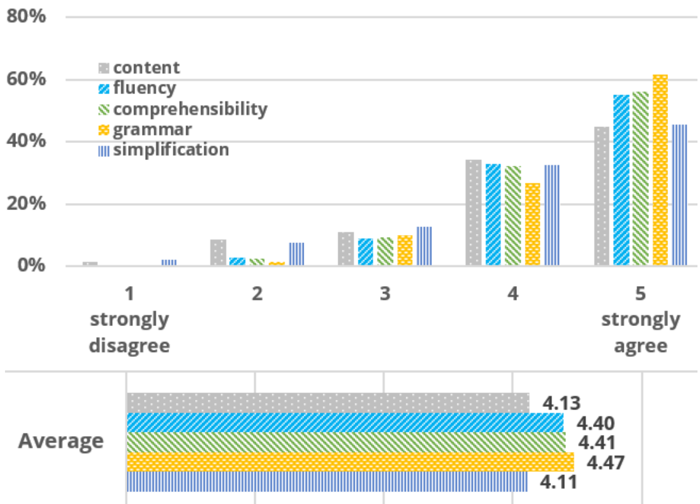

Fig. 10: Human Evaluation of *EN* Multilingual Text Simplification.

Figures 10 and 11 illustrate the participants' average feedback on the excerpts from the English coursebook "Intercultural and Ethical Decision-Making" simplified by our best *EN* text simplification and ChatGPT. We observe that the participants agree that the criteria *content*, *fluency*, *comprehensibility*, *grammar* and *simplification* are well covered in our *EN* text simplification. But also ChatGPT performs much better on our English text than on *DE* text simplification.

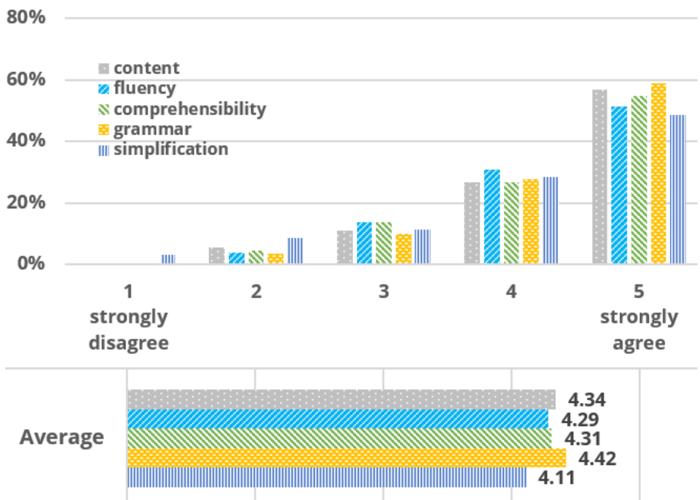

Fig. 11: Human Evaluation of ChatGPT's *EN* Text Simplification.

Comparing Figure 10 and 11 indicates that our *EN* text simplification system significantly outperforms ChatGPT on average in *fluency*, *comprehensibility*, and *grammar*. In *simplification*, the average score is equal for our system and ChatGPT. Only for the question of how well the simplified text reproduces the original text in terms of *content*, our system performs worse again. This demonstrates that ChatGPT deals with the simplicity-adequacy tradeoff [37] better in the English sentences.

6 Conclusion and Future Work

In this work, we have shown a way to leverage a multilingual NLP model and machine translation to address the problem of little data to train automatic systems for text simplification: To improve text simplification systems, it is possible to use sentence pairs which were artificially produced by machine-translating target language sentences for fine-tuning a multilingual NLP model. For all our five target languages, we achieved the highest SARI scores by adding 39k translations as training data in addition to our 1k training sentences in the target language.

Since we have found that translations from some languages have a more positive impact than from other languages, future work should include finding out which language combinations lead to optimal improvements. Moreover, it is interesting to investigate if using more translations will even further improve performance.

References

[1] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., Liu, P.J.: Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. The Journal of Machine Learning Research **21**(1) (2020)

[2] UNESCO: Literacy Rate, Adult Total (% of People Ages 15 and Above) (2022). https://data.worldbank.org/indicator/SE.ADT.LITR.ZS

[3] Charrow, V.R., Erhardt, M.K., Charrow, R.P.: Clear & Effective Legal Writing, 4th edn. Aspen Publishers, New York, NY (2007)

[4] Plavén-Sigray, P., Matheson, G.J., Schiffler, B.C., Thompson, W.H.: Research: The Readability of Scientific Texts is Decreasing over Time. eLife **6** (2017)

[5] Zaykova, I., Shilnikova, I.: Economic Translation: Theoretical and Practical Issues. In: The International Scientific and Practical Conference "Current Issues of Linguistics and Didactics: The Interdisciplinary Approach in Humanities and Social Sciences" (CILDIAH-2019) (2019). https://doi.org/10.1051/shsconf/20196900139

[6] Haspelmath, M.P.D.: Sprachen der Welt, Max-Planck-Institut für evolutionäre Anthropologie, Leipzig (2007). https://home.uni-leipzig.de/muellerg/su/haspelmath.pdf

[7] Brunato, D., Dell'Orletta, F., Venturi, G.: Linguistically-Based Comparison of Different Approaches to Building Corpora for Text Simplification: A Case Study on Italian. Frontiers in Psychology **13** (2022) https://doi.org/10.3389/fpsyg.2022.707630

[8] Taecharungroj, V.: "What Can ChatGPT Do?" Analyzing Early Reactions to the Innovative AI Chatbot on Twitter. Big Data and Cognitive Computing **7**(1), 35 (2023)

[9] Sennrich, R., Haddow, B., Birch, A.: Improving Neural Machine Translation Models with Monolingual Data. 7th International Conference on Smart Computing & Communications (ICSCC) (2015) https://doi.org/10.48550/arXiv.1511.06709

[10] Suter, J., Ebling, S., Volk, M.: Rule-based Automatic Text Simplification for German. In: The 13th Conference on Natural Language Processing (KONVENS 2016) (2016). https://www.linguistics.rub.de/konvens16/pub/35_konvensproc.pdf

[11] Maaß, C.: Leichte Sprache. Das Regelbuch. Forschungsstelle Leichte Sprache Universität Hildesheim, Berlin (2015)

[12] Xu, W., Napoles, C., Pavlick, E., Chen, Q., Callison-Burch, C.: Optimizing Statistical Machine Translation for Text Simplification. Transactions of the Association for Computational Linguistics **4**, 401–415 (2016) https://doi.org/10.1162/tacl_a_00107

[13] Chand, S.: Empirical Survey of Machine Translation Tools. In: Second International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 181–185 (2016). https://doi.org/10.1109/ICRCICN.2016.7813653

[14] Qiang, J., Wu, X.: Unsupervised Statistical Text Simplification. IEEE Transactions on Knowledge and Data Engineering **33**(4), 1802–1806 (2021) https://doi.org/10.1109/TKDE.2019.2947679

[15] Wang, Y.-Y., Ward, W.: Grammar Inference and Statistical Machine Translation. Carnegie Mellon University (1999)

[16] Zhang, M., Jiang, H., Li, H., Aw, A., Li, S.: Grammar Comparison Study for Translational Equivalence Modeling and Statistical Machine Translation. In: The 22nd International Conference on Computational Linguistics (Coling 2008), Manchester, UK, pp. 1097–1104 (2008). https://aclanthology.org/C08-1138

[17] Wang, T., Chen, P., Amaral, K.M., Qiang, J.: An Experimental Study of LSTM Encoder-Decoder Model for Text Simplification. ArXiv **abs/1609.03663** (2016)

[18] Omelianchuk, K., Raheja, V., Skurzhanskyi, O.: Text Simplification by Tagging. In: Workshop on Innovative Use of NLP for Building Educational Applications

(2021)

[19] Zhang, X., Lapata, M.: Sentence Simplification with Deep Reinforcement Learning. In: The Conference on Empirical Methods in Natural Language Processing, pp. 584–594. Association for Computational Linguistics, Copenhagen, Denmark (2017). https://doi.org/10.18653/v1/D17-1062 . https://aclanthology.org/D17-1062

[20] Truică, C.-O., Stan, A.-I., Apostol, E.-S.: SimpLex: A Lexical Text Simplification Architecture. Neural Computing and Applications (2022) https://doi.org/10.1007/s00521-022-07905-y

[21] Schlippe, T., Sawatzki, J.: Cross-Lingual Automatic Short Answer Grading. In: Proceedings of The 2nd International Conference on Artificial Intelligence in Education Technology (AIET), Wuhan, China (2021)

[22] Mallinson, J., Sennrich, R., Lapata, M.: Zero-Shot Crosslingual Sentence Simplification. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 5109–5126. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.emnlp-main.415

[23] Coster, W., Kauchak, D.: Simple English Wikipedia: A New Text Simplification Task. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, pp. 665–669. Association for Computational Linguistics, Portland, Oregon, USA (2011). https://aclanthology.org/P11-2117

[24] Alva-Manchego, F., Martin, L., Bordes, A., Scarton, C., Sagot, B., Specia, L.: ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations. In: The 58th Annual Meeting of the Association for Computational Linguistics, pp. 4668–4679. Association for Computational Linguistics, Online (2020). https://doi.org/10.18653/v1/2020.acl-main.424

[25] Papineni, K., Roukos, S., Ward, T., Zhu, W.-J.: BLEU: A Method for Automatic Evaluation of Machine Translation. In: The 40th Annual Meeting on Association for Computational Linguistics, pp. 311–318. Association for Computational Linguistics, USA (2002). https://doi.org/10.3115/1073083.1073135

[26] Sulem, E., Abend, O., Rappoport, A.: BLEU is Not Suitable for the Evaluation of Text Simplification. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 738–744. Association for Computational Linguistics, Brussels, Belgium (2018). https://doi.org/10.18653/v1/D18-1081

[27] Zhao, S., Meng, R., He, D., Saptono, A., Parmanto, B.: Integrating Transformer and Paraphrase Rules for Sentence Simplification. In: Proceedings of the 2018

Conference on Empirical Methods in Natural Language Processing, pp. 3164–3173. Association for Computational Linguistics, Brussels, Belgium (2018). https://aclanthology.org/D18-1355

[28] Lin, Z., Wan, X.: Neural Sentence Simplification with Semantic Dependency Information. Proceedings of the AAAI Conference on Artificial Intelligence **35**(15), 13371–13379 (2021) https://doi.org/10.1609/aaai.v35i15.17578

[29] Dong, Y., Li, Z., Rezagholizadeh, M., Cheung, J.C.K.: EditNTS: An Neural Programmer-Interpreter Model for Sentence Simplification through Explicit Editing. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3393–3402 (2019) https://doi.org/10.48550/ARXIV.1906.08104

[30] Alva-Manchego, F., Martin, L., Scarton, C., Specia, L.: EASSE: Easier Automatic Sentence Simplification Evaluation. In: The Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations, pp. 49–54. Association for Computational Linguistics, Hong Kong, China (2019). https://aclanthology.org/D19-3009

[31] Martin, L., Clergerie, , Sagot, B., Bordes, A.: Controllable Sentence Simplification. In: The Twelfth Language Resources and Evaluation Conference, pp. 4689–4698. European Language Resources Association, Marseille, France (2020). https://aclanthology.org/2020.lrec-1.577

[32] Clive, J., Cao, K., Rei, M.: Control Prefixes for Parameter-Efficient Text Generation. In: Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics (GEM), pp. 363–382. Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid) (2022). https://aclanthology.org/2022.gem-1.31

[33] Martin, L., Fan, A., Clergerie, , Bordes, A., Sagot, B.: MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases. In: The Thirteenth Language Resources and Evaluation Conference, pp. 1651–1664. European Language Resources Association, Marseille, France (2022). https://aclanthology.org/2022.lrec-1.176

[34] Alva-Manchego, F., Scarton, C., Specia, L.: The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification. Computational Linguistics **47**(4), 861–889 (2021) https://doi.org/10.1162/coli_a_00418

[35] Chung, H.W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., Li, E., Wang, X., Dehghani, M., Brahma, S., Webson, A., Gu, S.S., Dai, Z., Suzgun, M., Chen, X., Chowdhery, A., Valter, D., Narang, S., Mishra, G., Yu, A.W., Zhao, V., Huang, Y., Dai, A.M., Yu, H., Petrov, S., Chi, E.H.-h., Dean, J., Devlin, J., Roberts, A., Zhou, D., Le, Q., Wei, J.: Scaling Instruction-Finetuned Language Models. ArXiv **abs/2210.11416** (2022) https://doi.org/10.48550/arXiv.2210.11416

[36] Aiken, M.: An Updated Evaluation of Google Translate Accuracy. Studies in Linguistics and Literature **3**, 253 (2019) https://doi.org/10.22158/sll.v3n3p253

[37] Schwarzer, M., Kauchak, D.: Human Evaluation for Text Simplification: The Simplicity-Adequacy Tradeoff. In: SoCal NLP Symposium (2018)