Exploring Automatic Text Simplification of German Narrative Documents

Thorben Schomacker
Hamburg Univ. of Applied Sciences
thorben.schomacker@haw-hamburg.de

Tillmann Dönicke
Univ. of Göttingen
tillmann.doenicke@uni-goettingen.de

Marina Tropmann-Frick
Hamburg Univ. of Applied Sciences
marina.tropmann-frick@haw-hamburg.de

Abstract

In this paper, we apply transformer-based Natural Language Generation (NLG) techniques to the problem of text simplification. Currently, there are only a few German datasets available for text simplification, even fewer with larger and aligned documents, and not a single one with narrative texts. In this paper, we explore to which degree modern NLG techniques can be applied to German narrative text simplifications. We use Longformer attention and a pre-trained mBART model. Our findings indicate that the existing approaches for German are not able to solve the task properly. We conclude on a few directions for future research to address this problem.

1 Introduction

1.1 Motivation

With the rise of the internet, it has become convenient and often free to access an abundance of texts. However, not all people who have access can fully read and understand the texts, even though they speak the language that the text is written in. Most often this problem originates in the complex nature of the texts. Text simplification can help to overcome this barrier.

Narrative forms are one of the primary ways humans create meaning (Felluga, 2011). Narrative texts, then, make an important contribution to how we describe and shape our environment. Simple language also contributes to involving as many people as possible in this process. Providing narrative texts in a Simple Language (*Einfache Sprache*) version, enables a large audience to read them. So, we present the first approach for the automatic text simplification of German narrative texts.

1.2 Related Works

Automatic text simplification started in 2010 (Specia, 2010) as statistical machine translation to the rule-based automatic text simplification task, using a Portuguese corpus (4500 parallel sentences). The first German text simplification dataset was created by (Hancke et al., 2012) to train a readability classifier. The dataset consisted of unaligned articles from one adult-targeting and one child-targeting journal, and was later improved and enlarged by (Weiß and Meurers, 2018), which added unaligned transcripts from one adult-targeting and one child-targeting German TV news show. Similarly, (Aumiller and Gertz, 2022) published a German document-aligned dataset with lexicon articles for adults and for children. The first sentence-aligned German simplification dataset was published in 2013 (Klaper et al., 2013) with 270 articles from five different websites, mainly of organizations that support people with disabilities. In 2016 the first (rule-based) automatic text simplification system for German was released (Suter et al., 2016). The first parallel corpus for data-driven automatic text simplification for German was introduced by (Säuberli et al., 2020). The corpus contains 3616 sentence pairs from news articles. They additionally were the first to use transformer models for German text simplification and found out that their corpus was not large enough to train them. (Battisti et al., 2020) collected a larger corpus with 378 text pairs, mostly from websites of governments, specialized institutions, and non-profit organizations. (Rios et al., 2021) investigated the usage of an adapted mBART (Liu et al., 2020) version with Longformer attention (Beltagy et al., 2020) on Swiss newspaper articles. These results have been further improved with a sentence-based approach (Ebling et al., 2022). Most recently, the first detailed surveys about German text simplification have been released (Anschütz et al., 2023; Stodden et al., 2023; Schomacker et al., 2023).

Proceedings of the 19th Conference on Natural Language Processing (KONVENS 2023), pages 139–148
September 18–22, 2023.

2 Methods

Longformer mBART Our goal was to train a document-level text generation model with a larger context (> 510 input tokens; exceeding most transformer-based models). Longformer is the only model to our knowledge, which could extend the context on a pre-trained transformer model. We searched on huggingface.co and filtered for text2text-generation models (8551), German (225), > 5000 downloads (30), and that they can perform a German-to-German translation task. This leaves only *facebook/mbart-large-50* and *facebook/mbart-large-cc25*, both introduced in (Liu et al., 2020). We decided to take *facebook/mbart-large-cc25* since it has been trained on fewer languages (25; in the CC25 dataset extracted from (Wenzek et al., 2020; Conneau et al., 2020)) in comparison to *facebook/mbart-large-50* (50). Because we reasoned that the greater the relative proportion of German in pre-training, the better. Our situation is very similar to (Rios et al., 2021), so we base our methods on their approaches. mBART uses a specific input format consisting of the sentence and a language-tag. We additionally created two tags: `de_OR` and `de_SI` for Standard German and Simple German, respectively. Both of them are derived from the original German tag `de_DE` (fifth-largest proportion in CC25) and only modified during our fine-tuning process. Similar to (Rios et al., 2021), we applied the Longformer conversion to the mBART model with a maximum input length of 1024 and 512 as the attention window size.

Domain Adaptation By using domain adaptation, we aim to enrich the vocabulary with previously unseen words and adapt the existing embeddings to the narrative text domain and the historical environment of the texts. After we created the longmbart-model we started the domain adaptation process. We downloaded all documents from TextGrid (textgrid.de) in the category "prose" and randomly sampled 60 documents. In a next step, we sentence-split the documents using spaCy (spacy.io), shuffled them and masked 15% of the words. We used these masked and unmasked sentence-pairs for a single epoch training of the model. Both sides of the pair are tagged with the `de_DE` tag. We used a learning rate of $3e{-}10$, an attention window size of 512 during the conversion, a maximum input and output length of 70, and a batch size of 8.

Fine-Tuning We fine-tune our model on document-aligned German narrative texts, using three sources for Standard Language data: 1) gutenberg.org, 2) projekt-gutenberg.org, and 3) textgridrep.org. We selected *Die Bremer Stadtmusikanten* (`mils-stadtmusikanten`), *Der seltsame Fall von Dr Jekyll und Mr Hyde* (`eb-hyde`) and *Der Schimmelreiter* (`pv-schimmelreiter`) as development set because their amount of words is close to the average amount of all samples in the fine-tuning dataset and they originate from different sources. For the same reasons, we selected *Des Teufels rußiger Bruder* (`mils-bruder`), *Der Graf von Monte Christo* (`eb-christo`) and *Der Sandmann* (`pv-sandmann`) for testing. We used four sources for Simple Language texts: 1) einfachebuecher.de (`eb`), 2) kindermannverlag.de (`kv`), and 3) passanten-verlag.de (`pv`), which consist of classic novels, as well as 4) the *Märchen in Leichter Sprache* 'Fairy Tales in Simple Language' from ndr.de (`mils`). The links to the Standard Language and Simple Language version can be found in Table 2 in the appendix. The `mils` samples include the complete text, while for the novels we use only the excerpts provided in the form of free reading samples (usually the first chapter of the text). We manually cut in the end of the Standard Language version to match the extent of the Simple Language version.

Hyperparameter Setup Following (Rios et al., 2021), we set the attention mode (Beltagy et al., 2020) to sliding chunks (with overlap) and the attention window size to 512. Since our dataset is rather small, we turn gradient accumulation (`accumulate_grad_batches`) off. We use the Adam optimizer and optimize the learning rate with the PyTorch Lightning `LearningRateFinder` between $3e{-}20$ and $3e{-}1$. For Decoding we use beam search (size = 4).

3 Analysis and Evaluation

3.1 Analysis

We manually compared the three generated output sequences of our test texts to the Standard Language version and the Simple Language version. In summary, we found that 1) the model copies the input text to a very high degree without any modifications, 2) in cases where the model discarded parts of the inputs, it did not recognize the importance of the sequence, such as spelled-out antecedents for pronouns, and 3) it truncates rather randomly and without any semantic reason.

For reasons of space, we only discuss *Der Sandmann* in the appendix (section B), on which we can show all the phenomena we want to discuss.

3.2 Evaluation Measures

3.2.1 BERTscore, BLEU and ROUGE

BERTscore (Zhang et al., 2020) is currently the recommended (Alva-Manchego et al., 2021) way of comparing (generated) text simplification candidates and the (gold) references. It is a soft metric that yields high correlations with human judgments (Alva-Manchego et al., 2021). We select *google/mt5-base* as the underlying model, since it is the best performing model with Max Length > 1022, German support, and a compatible transformers version (Zhang, 2020) (*google/mt5-xl* and *google/mt5-large* did not fit our hardware resources). Following (Alva-Manchego et al., 2021) we use the BERTscore to determine the early stopping point during fine-tuning. We additionally employed two n-gram based approaches, BLEU (Papineni et al., 2002) and ROUGE (Lin, 2004), because they are the most commonly used metrics for text generation.

3.2.2 Entropy

We use two flavors of Shannon entropy as a characterization, or measurement, of redundancy. In the basic implementation, we calculate the bag-of-words (BOW) entropy: $H(W) = \sum_{w \in W} \frac{count(w)}{n} \cdot - \log_2 \left(\frac{count(w)}{n} \right)$, where w is a word in the bag of words W, $count(w)$ is the frequency of w in W, n is the total size of W, and $H(W)$ is the text-level entropy.

In addition, we calculate the shortest-unique-prefix (SUP) entropy (Kontoyiannisy, 1997), by calculating the length of the shortest prefix starting at x_i that does not appear starting anywhere in the previous i tokens $x_0, x_1, \ldots, x_{i-1}$. This prefix-length l_i can be thought of as the length of the next unique substring after the past up to position $(i-1)$ has been encoded. In other words, this metric measures the surprise value of a substring. The SUP entropy is calculated as: $\hat{H}_N = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{l_i}{log(i+1)} \right]^{-1}$ with $N < M$, where M is the largest possible index (= the sequence length $+1$). (Kontoyiannisy, 1997) do not elaborate on how N should be chosen, so we set it to $\lfloor \frac{M}{2} \rfloor$.

In both cases, we use spaCy to tokenize the generated output. We consider all tokens including punctuation marks and lowercase them.

3.3 Results and Discussion

We analyze the model's performance via two kinds of metrics: similarity-based (BERTscore, BLEU and ROUGE) and entropy-based (SUP and BOW). Table 1 shows that the model without fine-tuning and domain adaptation performs the best both in terms of entropy and similarity. A single epoch of fine-tuning seems not to affect the models' performance, but fine-tuning it for 11 epochs worsens it drastically. Similarly, domain adaption without and with 1 epoch of fine-tuning drops below all non-domain-adapted models. Both domain adaptation set-ups (50 and 100 documents) perform the same, so the number of domain adaptation documents seems to have no effect on the performance. Interestingly, with more fine-tuning (11 epochs) the SUP entropy is improved, while the BERTscore-similarity further drops.

The model without domain adaptation and without fine-tuning performed the best and the more we trained the model, the more frequently individual text elements are repeated—first individual clauses, then words, and in the end only characters. These are results that no longer represent meaningful texts, let alone a high-quality text simplification. We did not manage to definitively conclude on reasons why both fine-tuning and domain adaptation do not outperform the pre-trained model. We assume that the main reason could be so-called catastrophic forgetting, which can occur in all scenarios where machine learning models are trained on a sequence of tasks and the accuracy on earlier tasks drops significantly. The model in our experiments was previously trained on inter-language translation (from one language to another) and we fine-tune it on intra-language translation (from one version of a language to another version of the same language). So, domain adaptation, being an intra-language task, differs from the original mBART task. The model's general text generation capability dropped after fine-tuning and domain adaptation. (Ramasesh et al., 2021) demonstrate that forgetting is concentrated at the higher model layers and argue that it should be mitigated there. In their set-up, these layers change significantly and erase earlier task subspaces through sequential training of multiple tasks. All the mitigation methods they in-

Domain Adapt.	BERTscore$_{F1}$	ROUGE-l_{F1}	BLEU	SUP	BOW	Fine Tuning ♣	lr
-	**0.682**	**0.127**	**1.43**	**1.000**	**6.685**	0	-
-	0.682	0.127	1.43	1.000	6.685	1	7.8e-20
-	0.318	0	0	340.000	0.003	11 (100;10)	8.1e-07
50 texts	0.301	0	0	123.666	0.038	1	3e-10 ♠
50 texts	0.301	0	0	123.666	0.038	0	-
100 texts	0.301	0	0	123.666	0.038	0	-
100 texts	0.301	0	0	123.666	0.038	1	3e-10 ♠
100 texts	0.298	0	0	49.666	0.0441	11 (100;10)	3e-10 ♠

Table 1: Average performance of our models on the test texts. ♣ : Best epochs with max epochs (and early stopping patience, if used, in parenthesis). ♠ : The lr auto was unable to find an optimal learning rate; so we use a predefined value.

vestigate stabilize higher layer representations, but vary on whether they enforce more feature reuse, or store tasks in orthogonal subspaces. There are several other possible reasons for this behavior and opportunities to improve the models' performance. In the following section, we give an outlook on possible ways of adjustment.

4 Conclusion and Future Work

In this paper, we apply existing transformer-based methods to generate text simplifications on document level. Furthermore, we investigate the usage of fine-tuning and domain adaptation.

Our work contributes to the field of automatic German text simplifications. This field is understudied, and future works that want to build on top of our and other previous works' findings could research the following areas:

Catastrophic Forgetting (Yu et al., 2021) investigate catastrophic forgetting and speculate that their second phase of pre-training results in some form of catastrophic forgetting for the pre-trained model, which could have hurt the adaptation performance. They recommend to use RecAdam (Chen et al., 2020), which mitigated the problem in their abstractive text summarization study.

Repetition Problem (Fan et al., 2018) show that maximization-based approaches (such as beam search) tend to produce text that contains undesirable repetitions, and stochastic methods tend to produce text that is semantically inconsistent with the given prefix. We use beam search in our approach and experience a significant increase of repetition during training. (Xu et al., 2022) divide approaches for mitigating repetition into 1) training-based (Welleck et al., 2020; Lin et al., 2021; Xu et al., 2022) and decoding-based (See et al., 2017; Fan et al., 2018; Holtzman et al., 2020) approaches. Recently, two new decoding approaches, Nucleus (Holtzman et al., 2020) and Contrastive Search (Su and Collier, 2022), have shown promising results in terms of reducing repetition and improving the overall quality of generated text. Future work could apply these newer decoding methods to the task of document-level text simplification. However, although there is an increasing number of mitigating techniques, the causes of the repetition problem are still under-investigated.

Entropy Entropy metrics provide additional and very inexpensive guidance on the quality of generated text simplification. They can show very well to what degree the repetition problem is present in the text generation model. We encourage future research in similar tasks to measure entropy in their works.

Masking Strategies We only use the commonly used token masking strategy for BART. (Lewis et al., 2019) describes other strategies, that can be used in the future as well.

Controllability and Learning Strategies (Erdem et al., 2022, p. 1165–1168) name a few resources where adding metadata, such as named entities or parts of speech, to the input can be used as an advanced learning strategy to improve results and offer more controllability over the output. We do not add any metadata and observed in Section 3 that our model is not able to properly recognize named entities. Inserting corresponding metadata could potentially improve the performance in this regard.

Unify Designations Designations of people are interchangeably used in Standard Language. A good example is the father in *Der Sandmann*, who is mostly addressed as *Vater* ("father") but also as *Papa* ("dad") by his children and as *Herr* ("master"), as in *Herr des Hauses* ("man of the house"), by his house staff. All these words mean the same and are referring to the same person. Unifying them could help.

Limitations

The work we described in this paper investigates the automatic simplification of narrative documents in German. Our Methodology is focussed on document-level simplification and is only transferable to a limited extent to simplification that works on sentence-level or other linguistic levels. Additionally, our approach as well as the future research areas are generally applicable to document-level simplification in a broad variety of languages. The choice of quantitative evaluation can be applied to any text simplification task, with structural and linguistic limitations. The qualitative evaluation highly considers the narrative nature of our data, so it is transferable to the simplification of narrative texts in any form but hardly applicable to other text genres.

The data we used is targeted towards different audiences, children and/or people with a lower literacy. Furthermore, some are written in easy language (*Leichte Sprache*) and others in the broader category simple language (*Einfache Sprache*). Future researchers are advised to carefully check the data sources and evaluate to which degree the data can be used for the intended purpose. Due to copyright restrictions, we are only able to provide public URLs to the data, and cannot provide the data directly.

Ethics Statement

We state that our work complies with the ACL Ethics Policy.[1] Our work investigates the automatic simplification of narrative documents in German. Providing simplified versions of texts positively contributes to the inclusion of people with cognitive disabilities and lower literacy into a growing number of aspects of society. Automatically generated simplifications offer a lower cost point compared to their human-made equivalents. On the one hand, this increases the number of people that can afford to read these text, on the other hand, it can endanger the future job prospects of human translators, which specialized in simplifying texts.

Acknowledgments

We would like to thank the Norddeutscher Rundfunk (NDR) for allowing us to use the "Märchen in Leichter Sprache" [2] and make them available to a scientific audience.

References

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2021. The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification. *Computational Linguistics*, 47(4):861–889.

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:2305.12908 [cs].

Dennis Aumiller and Michael Gertz. 2022. Klexikon: A German Dataset for Joint Summarization and Simplification. In *Proceedings of the 13th Conference on Language Resources and Evaluation (LREC 2022)*, pages 2693–2701.

Alessia Battisti, Dominik Pfütze, Andreas Säuberli, Marek Kostrzewa, and Sarah Ebling. 2020. A Corpus for Automatic Readability Assessment and Text Simplification of German. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3302–3311, Marseille, France. European Language Resources Association.

Iz Beltagy, Matthew E. Peters, and Arman Cohan. 2020. Longformer: The Long-Document Transformer. ArXiv: 2004.05150.

Sanyuan Chen, Yutai Hou, Yiming Cui, Wanxiang Che, Ting Liu, and Xiangzhan Yu. 2020. Recall and Learn: Fine-tuning Deep Pretrained Language Models with Less Forgetting. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7870–7881, Online. Association for Computational Linguistics.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Sarah Ebling, Alessia Battisti, Marek Kostrzewa, Dominik Pfütze, Annette Rios, Andreas Säuberli, and Nicolas Spring. 2022. Automatic Text Simplification for German. *Frontiers in Communication*, 7:706718. Publisher: Frontiers Research Foundation.

[1] https://www.aclweb.org/portal/content/acl-code-ethics

[2] https://www.ndr.de/fernsehen/barrierefreie_angebote/leichte_sprache/Maerchen-in-Leichter-Sprache,maerchenleichtesprache100.html

Erkut Erdem, Menekse Kuyu, Semih Yagcioglu, Anette Frank, Letitia Parcalabescu, Barbara Plank, Andrii Babii, Oleksii Turuta, Aykut Erdem, Iacer Calixto, Elena Lloret, Elena-Simona Apostol, Ciprian-Octavian Truică, Branislava Šandrih, Sanda Martinčić-Ipšić, Gábor Berend, Albert Gatt, and Grăzina Korvel. 2022. Neural Natural Language Generation: A Survey on Multilinguality, Multimodality, Controllability and Learning. *Journal of Artificial Intelligence Research*, 73:1131–1207.

Angela Fan, Mike Lewis, and Yann Dauphin. 2018. Hierarchical Neural Story Generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 889–898, Melbourne, Australia. Association for Computational Linguistics.

Dino Felluga. 2011. General Introduction to Narratology.

Julia Hancke, Sowmya Vajjala, and Detmar Meurers. 2012. Readability Classification for German using Lexical, Syntactic, and Morphological Features. In *Proceedings of COLING 2012*, pages 1063–1080, Mumbai, India. The COLING 2012 Organizing Committee.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. 2020. The Curious Case of Neural Text Degeneration. In *International Conference on Learning Representations*, pages 1–16.

David Klaper, S. Ebling, and Martin Volk. 2013. Building a German/Simple German Parallel Corpus for Automatic Text Simplification. In *Klaper, David; Ebling, S; Volk, Martin (2013). Building a German/Simple German Parallel Corpus for Automatic Text Simplification. In: The Second Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR 2013), Sofia, Bulgaria, 8 August 2013.*, pages 11–19, Sofia, Bulgaria. University of Zurich.

I Kontoyiannisy. 1997. The Complexity and Entropy of Literary Styles. Technical report.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. 2019. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension.

Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Xiang Lin, Simeng Han, and Shafiq Joty. 2021. Straight to the Gradient: Learning to Use Novel Tokens for Neural Text Generation. In *Proceedings of the 38th International Conference on Machine Learning*, pages 6642–6653. PMLR. ISSN: 2640-3498.

Yinhan Liu, Jiatao Gu, Naman Goyal, Xian Li, Sergey Edunov, Marjan Ghazvininejad, Mike Lewis, and Luke Zettlemoyer. 2020. Multilingual Denoising Pre-training for Neural Machine Translation. *Transactions of the Association for Computational Linguistics*, 8:726–742. Place: Cambridge, MA Publisher: MIT Press.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Vinay Venkatesh Ramasesh, Ethan Dyer, and Maithra Raghu. 2021. Anatomy of Catastrophic Forgetting: Hidden Representations and Task Semantics. In *Proceedings of Conference on Learning Representations*, pages 1–31. International Conference on Learning Representations.

Annette Rios, Nicolas Spring, Tannon Kew, Marek Kostrzewa, Andreas Säuberli, Mathias Müller, and Sarah Ebling. 2021. A New Dataset and Efficient Baselines for Document-level Text Simplification in German. In *Proceedings of the Third Workshop on New Frontiers in Summarization*, pages 152–161, Online and in Dominican Republic. Association for Computational Linguistics. Tex.ids= riosNewDatasetEfficient2021a.

Thorben Schomacker, Michael Gille, Marina Tropmann-Frick, and Jörg von der Hülls. 2023. Data and Approaches for German Text Simplification - Next Steps toward an Accessibility-enhanced Communication.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

Lucia Specia. 2010. Translating from Complex to Simplified Sentences. In *Computational Processing of the Portuguese Language*, Lecture Notes in Computer Science, pages 30–39, Berlin, Heidelberg. Springer.

Regina Stodden, Omar Momen, and Laura Kallmeyer. 2023. DEPLAIN: A German Parallel Corpus with Intralingual Translations into Plain Language for Sentence and Document Simplification. ArXiv:2305.18939 [cs].

Yixuan Su and Nigel Collier. 2022. Contrastive Search Is What You Need For Neural Text Generation. ArXiv:2210.14140 [cs].

Julia Suter, Sarah Ebling, and Martin Volk. 2016. Rule-based Automatic Text Simplification for German. In *Proceedings of the 13th Conference on Natural Language Processing*, pages 279–287.

Andreas Säuberli, Sarah Ebling, and Martin Volk. 2020. Benchmarking Data-driven Automatic Text Simplification for German. In *Proceedings of the 1st Workshop on Tools and Resources to Empower People with REAding DIfficulties (READI)*, pages 41–48, Marseille, France. European Language Resources Association.

Zarah Weiß and Detmar Meurers. 2018. Modeling the Readability of German Targeting Adults and Children: An empirically broad analysis and its cross-corpus validation. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 303–317, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sean Welleck, Ilia Kulikov, Stephen Roller, Emily Dinan, Kyunghyun Cho, and Jason Weston. 2020. Neural Text Generation With Unlikelihood Training. In *International Conference on Learning Representations*, pages 1–18. International Conference on Learning Representations.

Guillaume Wenzek, Marie-Anne Lachaux, Alexis Conneau, Vishrav Chaudhary, Francisco Guzmán, Armand Joulin, and Edouard Grave. 2020. CCNet: Extracting High Quality Monolingual Datasets from Web Crawl Data. In *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*, page 10.

Jin Xu, Xiaojiang Liu, Jianhao Yan, Deng Cai, Huayang Li, and Jian Li. 2022. Learning to Break the Loop: Analyzing and Mitigating Repetitions for Neural Text Generation. In *Proceedings of the 36th Conference on Neural Information Processing Systems*, pages 1–36. 36th Conference on Neural Information Processing Systems. ArXiv:2206.02369 [cs].

Tiezheng Yu, Zihan Liu, and Pascale Fung. 2021. AdaptSum: Towards Low-Resource Domain Adaptation for Abstractive Summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5892–5904, Online. Association for Computational Linguistics.

Tianyi Zhang. 2020. BERTScore Default Layer Performance on WMT16 (last accessed: 2022-09-26).

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. ArXiv:1904.09675 [cs].

A Code

Our code is available on GitHub:

Pre-Processing of the Fine-Tuning Dataset based on Projekt Gutenberg, Gutenberg, and PDF-Reading Samples Texts:
github.com/tschomacker/aligned-narrative-documents

Pre-Processing of the Domain Adapation based on Textgrid Texts:
github.com/tschomacker/textgrid-domain-adaptation-dataset

Machine Learning Architecture and Implementation:
github.com/tschomacker/longmbart

B Analysis of *Der Sandmann*

The output of the model is shown in Figure 1. From line 1 to 18, the complete text is equivalent to the Standard Language input. After *Franz Moor den Daniel* (line 19) the model inserts multiple passages that occur previously in the text, e.g. three times *unglückseligen Krämers gar feindlich auf mich wirken muß* ("I can't help but think that the unfortunate grocer must have a hostile effect on me") in line 23.

Furthermore, the model changes facts in the text, e.g. in line 26 supper is served at seven o'clock, while in the Standard Language version lunch is served at the same time. The reference Simple Language version from Passanten Verlag completely discards the facts about dinner and supper, boiling this passage down to a brief introduction of the father and mentioning that he was busy with his work and that he told fascinating stories to his kids. Another aspect of line 26 is that the model output does not mention the father. This is the first introduction of this character in the story, so the model discarded an important character from this text passage. Furthermore, although the model does not fully remove the father from the text, it only refers to him via pronouns: *Er mochte mit seinem Dienst* ("He might be with his work") (line 27) refers to the father by the pronoun *Er* ("He"), despite the fact that the character was never introduced or referred to before. For a reader who has only access to the model's output, it is impossible to understand who *Er* ("He") is. A clean or complete removal of a character would show some simplification capability, even if it was an important character. In this case, it was an incomplete removal of an arguably important character.

Most of the repeated sentences do not contain information that is important to follow the story. In this respect, there is actually no need to transfer them into the simplification, let alone repeat them. Especially sentences like *So ist es in der Tat* ("So it is indeed"), are only a linguistic emphasis and arguably add linguistic complexity without additional content. If we assume that repeated sentences are perceived as important by the model [3], the model correctly recognizes an importance only in one case, namely the first mention of the barometer seller Coppolla (one of the main characters in the story) and the narrator's fear of him in lines 26–28, *unglückseligen Krämers gar feindlich auf mich wirken muß* ("unfortunate grocer must have a hostile effect on me").

[3] Profound hypothesis on the causes of repetition are sparse. We base our conjecture on the results of Xu et al. (2022), assuming a correlation between initial probability and repetition rate. If a text fragment occurs more often in the course of the document, it is more likely to be repeated. Therefore, we would say that *unglückseligen Krämers* ("unfortunate grocer") has a high initial probability for the model. In this respect, it is information that should be repeated more frequently in the text and can therefore be considered as important.

Ofel an LotharGewiß seid Ihr alle voll Unruhe, daß ich so lange - lange nicht geschrieben. Mutter zürnt wohl, und Clara mag glauben, ich lebe hier in Saus und Braus und vergesse mein holdes Engelsbild, so tief mir in Herz und Sinn eingeprägt, ganz und gar. - Dem ist aber nicht so; täglich und stündlich gedenke ich Eurer aller und in süßen Träumen geht meines holden Clärchens freundliche Gestalt vorüber und lächelt mich mit ihren hellen Augen so anmutig an, wie sie wohl pflegte, wenn ich zu Euch hineintrat. - Ach wie vermochte ich denn Euch zu schreiben, in der zerrissenen Stimmung des Geistes, die mir bisher alle Gedanken verstörte! - Etwas Entsetzliches ist in mein Leben getreten! - Dunkle Ahnungen eines gräßlichen mir drohenden Geschicks breiten sich wie schwarze Wolkenschatten über mich aus, undurchdringlich jedem freundlichen Sonnenstrahl. - Nun soll ich Dir sagen, was mir widerfuhr. Ich muß es, das sehe ich ein, aber nur es denkend, lacht es wie toll aus mir heraus. - Ach mein herzlieber Lothar! wie fange ich es denn an, Dich nur einigermaßen empfinden zu lassen, daß das, was mir vor einigen Tagen geschah, denn wirklich mein Leben so feindlich zerstören konnte! Wärst Du nur hier, so könntest Du selbst schauen; aber jetzt hältst Du mich gewiß für einen aberwitzigen Geisterseher. - Kurz und gut, das Entsetzliche, was mir geschah, dessen tödlichen Eindruck zu vermeiden ich mich vergebens bemühe, besteht in nichts anderm, als daß vor einigen Tagen, nämlich am 30. Oktober mittags um 12 Uhr, ein Wetterglashändler in meine Stube trat und mir seine Ware anbot. Ich kaufte nichts und drohte, ihn die Treppe herabzuwerfen, worauf er aber von selbst fortging.Du ahnest, daß nur ganz eigne, tief in mein Leben eingreifende Beziehungen diesem Vorfall Bedeutung geben können, ja, daß wohl die Person jenes unglückseligen Krämers gar feindlich auf mich wirken muß. So ist es in der Tat. Mit aller Kraft fasse ich mich zusammen, um ruhig und geduldig Dir aus meiner frühern Jugendzeit so viel zu erzählen, daß Deinem regen Sinn alles klar und deutlich in leuchtenden Bildern aufgehen wird. Indem ich anfangen will, höre ich Dich lachen und Clara sagen: 'Das sind ja rechte Kindereien!' - Lacht, ich bitte Euch, lacht mich recht herzlich aus! - ich bitt Euch sehr! - Aber Gott im Himmel! die Haare sträuben sich mir und es ist, als flehe ich Euch an, mich auszulachen, in wahnsinniger Verzweiflung, wie Franz Moor den Daniel. So ist es in der Tat. Mit aller Kraft fasse ich Euch aus meiner frühern Jugendzeit so viel zu erzählen, daß Deinem regen Sinn alles klar und deutlich in leuchtenden Bildern aufgehen wird. Indem ich anfangen will, höre ich Dich lachen und Clara sagen: 'Das sind ja rechte Kindereien!' - Lacht, ich bitte Euch, lacht mich recht herzlich aus! - es ist, als flehe ich Euch an, mich auszulachen, in wahnsinniger Verzweiflung, wie Franz Moor den Daniel. So ist es in der Tat. So ist es in der Tat . Mit aller Kraft fasse ich Euch aus meiner frühern Jugendzeit so viel zu erzählen, daß Deinem unglückseligen Krämers gar feindlich auf mich wirken muß , ja, daß wohl die Person jenes unglückseligen Krämers gar feindlich auf mich wirken muß, ja, daß wohl die Person jenes unglückseligen Krämers gar feindlich auf mich wirken muß. So ist in der Tat. So ist es in der Tat. - Nun fort zur Sache!Außer dem Mit aller dem Mittagsessen, das alter Sitte gemäß schon um sieben Uhr aufgetragen wurde , das alter Sitte gemäß schon um sieben Uhr aufgetragen wurde. Er mochte mit seinem Dienst . Er mochte mit seinem Dienst viel beschäftigt sein. Nach dem Abendessen. Er mochte mit seinem Dienst viel beschäftigt sein. Nach dem Abendessen, das alter Sitte gemäß, das alter Sitte gemäß, das alter Sitte gemäß, das alter Sitte gemäß von uns um sieben Uhr aufgetragen. Nach dem Abendessen, daß er aber von selbst fortging, daß er aber von selbst fortging.

Figure 1: Generated output of our best performing model, with "Der Sandmann" by E. T. A. Hoffmann as input. We did not change the format of the besides adding highlights and line numbers. The yellow highlights point the reader towards text passages, which showcase our model's shortcomings, which we discuss in Section B. For more information on the of the text, please refer to Table 2.

C Additional Tables

Full Title	Source-ID	Published	Source Texts (Standard and Simple)
Die Abenteuer von Tom Sawyer	eb-sawyer	English (1876)	gutenberg.org/ebooks/30165 einfachebuecher.de/Die-Abenteuer-von-Tom-Sawyer/978-3-947185-33-7
Moby Dick	eb-moby	English (1851)	projekt-gutenberg.org/melville/mobydick/ einfachebuecher.de/Moby-Dick/978-3-944668-86-4
Der Graf von Monte Christo	eb-christo	French (1846)	projekt-gutenberg.org/dumasalt/montchr1/ einfachebuecher.de/Der-Graf-von-Monte-Christo/978-3-944668-53-6
Die Abenteuer von Huckleberry Finn	eb-huckleberry	English (1885)	gutenberg.org/ebooks/64482 einfachebuecher.de/Die-Abenteuer-von-Huckleberry-Finn/978-3-947185-34-4
Der seltsame Fall von Dr Jekyll und Mr Hyde	eb-hyde	English (1886)	projekt-gutenberg.org/stevenso/jekyhyde/ einfachebuecher.de/Der-seltsame-Fall-von-Dr-Jekyll-und-Mr-Hyde/978-3-944668-54-3
In 80 Tagen um die Welt	eb-welt	French (1873)	projekt-gutenberg.org/verne/80tage/ einfachebuecher.de/In-80-Tagen-um-die-Welt/978-3-944668-32-1
Aus Kinderzeiten (in *Diesseits*)	eb-hesse	German (1907)	gutenberg.org/ebooks/47818 einfachebuecher.de/Erzaehlungen-von-Hermann-Hesse/978-3-944668-85-7
Sherlock Holmes. Das gesprenkelte Band	eb-band	English (1892)	projekt-gutenberg.org/doyle/getupfte/chap002.html einfachebuecher.de/Sherlock-Holmes.-Das-gesprenkelte-Band/978-3-944668-36-9
Sherlock Holmes. Das Zeichen der Vier	eb-vier	English (1890)	projekt-gutenberg.org/doyle/zeichen4/ einfachebuecher.de/Sherlock-Holmes.-Das-Zeichen-der-Vier/978-3-944668-39-0
20.000 Meilen unter dem Meer	eb-meer	French (1870)	projekt-gutenberg.org/verne/zwanzig1/ einfachebuecher.de/20.000-Meilen-unter-dem-Meer/978-3-947185-56-6
Die Verwandlung	eb-verwandlung	German (1912)	gutenberg.org/ebooks/22367 einfachebuecher.de/Die-Verwandlung/978-3-947185-99-3
Wolfsblut	pv-wolfsblut	English (1906)	projekt-gutenberg.org/london/wolfsblu/ passanten-verlag.de/lesen/#wolfsblut
Der Schimmelreiter	pv-schimmelreiter	German (1888)	projekt-gutenberg.org/storm/schimmel/ passanten-verlag.de/lesen/#schimmelreiter
Undine	pv-undine	French (1811)	projekt-gutenberg.org/fouque/undine/ passanten-verlag.de/lesen/#undine
Hiob	pv-hiob	German (1930)	projekt-gutenberg.org/roth/hiob/ passanten-verlag.de/lesen/#hiob
Der Sandmann	pv-sandmann	German (1816)	gutenberg.org/ebooks/6341 passanten-verlag.de/lesen/#sandmann
Weiße Nächte	pv-naechte	Russian (1848)	projekt-gutenberg.org/dostojew/novellen/chap01.html passanten-verlag.de/lesen/#naechte
Der glückliche Prinz	pv-prinz	English (1888)	projekt-gutenberg.org/wilde/maerche1/chap001.html passanten-verlag.de/lesen/#prinz3
Der Sandmann	kv-sandmann	German (1816)	gutenberg.org/ebooks/6341 kindermannverlag.de/produkt/der-sandmann/
Der Schimmelreiter	kv-schimmelreiter	German (1888)	projekt-gutenberg.org/storm/schimmel/ kindermannverlag.de/produkt/der-schimmelreiter/
Kinder- und Hausmärchen - Brüder Grimm	All mils-documents	German (1858)	projekt-gutenberg.org/grimm/khmaerch/ ndr.de/fernsehen/barrierefreie_angebote/leichte_sprache/Maerchen-in-Leichter-Sprache, maerchenleichtesprache100.html

Table 2: All documents in our corpus from `einfachebuecher.de` (eb) which are classified as "Klassiker" (classic novel) (snapshot from 07/14/2022), and Passanten Verlag (pv) (snapshot from 07/14/2022), Kindermann Verlag (kv) (snapshot from 07/14/2022) and *Märchen in Leichter Sprache* (`mils`) (snapshot from 07/14/2022).