Text Simplification for German

Uliana Eliseeva
Saarland University / 66123 Saarbrücken, Germany
ulel00001@stud.uni-saarland.de

Abstract

Automatic text simplification has drawn quite a lot of attention in recent years. In the rule-based approach to this task, the applied rules are normally uncompromisingly strict such that an original text is broken down into basic syntactic structures such as one clause per sentence or exclusively active voice. This kind of simplified output is either produced for people with reading disabilities or further passed to an automatic summarizer. In the present project, we build a simplification system for German language that utilizes a set of statistically-motivated rules that are meant to bring the syntactic structure of a complex Wikipedia article closer to that of the children-targeted Klexikon. The pipeline performs transformations of the syntactic dependencies such as passive voice, coordinated clauses, and AMS (measure argument of adjective, e.g. "3 meters"). In this report, we show and discuss the quality of the transformations and outline their potential application.

1 Introduction

Automated text simplification is being actively researched these days. A bulk of this research is being conducted for English though, while simplification for non-English languages is rarely in the center of attention. In this project, we aim to build an architecture for syntactic text simplification for German focusing on general domain articles such as Wikipedia. Our goal is to transform original articles into a statistically-motivated syntactic structure that would be more clear to children.

The task of text simplification aims to make texts easier to comprehend and faster to access. In text simplification the objective is usually to rewrite a text into a *simplified language* to make it accessible and understandable for non-native speakers, learners of the language, or people with reading disabilities. Crucially, the level of complexity of the original text and the intended complexity need to be taken into consideration. However, simplification may also refer to removing excessive information from an original text. Therefore, the task of simplification may also be seen as the process of making a complex, lengthy and highly specialized text, such as a scientific article, accessible to laymen.

As evident from the task definition, simplification may imply a certain level of summarization since the length of a linguistically simple text can still impede its comprehension. This task overlap also establishes itself in the expected result. To a greater or lesser extent, both simplification and summarization inevitably lead to information loss in the output.

As to execution, simplification can be performed on a syntactic or word level. The former expresses complex structures with their simpler alternatives, while the latter replaces domain-specific terms or infrequent words with more frequent ones. Those kinds of changes in a text can be done with a set of rewrite rules, which is normally referred to as a rule-based approach.

Traditionally, rule-based simplification is focused on performing adaptations for people with reading disabilities. More specifically, this kind of text transformation requires to completely avoid certain syntactic structures like passive voice, embedded relative clauses, reverse word order, or pronouns. As a result, the simplified text may be even more difficult to process for other categories of readers who are used to more standard coherence and cohesion devices.

As mentioned before, a more accessible discourse may benefit to not only dyslexic audience but also non-experts, children, learners of language etc. From this angle, the rule-based simplification has not received much attention to the best of our knowledge, especially for German language. Sets of rules are typically applied to modify input texts as an intermediate step before passing it to a summarizer but not in a stand-alone task.

2 Background

Traditional rule-based methods can be especially useful for text simplification because they follow predefined guidelines rewriting texts into a simplified language. Suter et al. (2016) built such a system for *simplified German* for people with reading disabilities following guidelines developed by Maaß (2015).

These guidelines define simplified language on the character, word, sentence, text, and layout level. On the character and word level, the software removes parentheses with their contents, expands abbreviations, rewrites numbers as words, segments compounds visually, uses look-up tables to find easier synonyms for some words. On the sentence level, it changes past simple to past perfect, rewrites passive to active, splits clauses so that they result in one clause per sentence. On the text level and layout, it adds explanations for difficult words and presents text one sentence per line.

To the best of our knowledge, Suter et al.'s (2016) system is the only one freely available rule-based simplification software for German. However, the nature of their task - simplifying texts for disabled readers - leads to a very strict set of rules at all levels of text production and presentation which can be confusing for readers like children, language learners, or non-experts. These reader categories obviously have different complexity needs.

Complexity-reducing models have also been developed for non-native German readers who speak the language at different levels. Thus, Rios et al. (2021) train a neural machine translation model to transform Standard German into A1-B2 levels. Their corpus (which is not publicly available at the moment) contains around 10,000 documents self-scraped from public resource websites adapted for different CEFR levels. Among the findings of the study, the researchers highlight that with the decreasing complexity of the target texts, from B2 to A1, the performance of the model decreases as well.

Whereas dyslexic and non-native readers of German have started to receive some attention in the recent years, children with their specific discourse complexity needs remain largely unaddressed. This is partially rooted in the fact that the age-specific guidelines have not been developed yet. Moreover, any other type of simplification guidelines does not have a standardized version (Suter et al., 2016), let alone audience-specific ones.

The rule-based simplification in English also shows a sparse coverage of different target groups. Mostly, the studies are focused on simplification that is meant to improve the quality of some other systems, such as summarization. This intermediate role is carried out, for example, in Vale, Lins and Ferriera's (2020) work. They built a rule-based pipeline that was used to simplify syntax before the text is summarized. They also contrasted the rule-based method to an optimization method, a supervised and an unsupervised deep learning models.

Their set of syntactic rules was quite strict as well. All sentences of the original corpus had to contain only one clause and passive voice was completely avoided. The rules were applied recursively, which means that the transformations were performed until no transformations were possible. The simplified texts were then summarized and compared to an unsimplified summary. The rule-based simplification was found to neither significantly improve nor deteriorate the summarization. On the other hand, the supervised neural syntactic simplification has shown the best results in recall and precision compared to the unsimplified baseline and the other models.

This might pose a question of how reasonable it is to choose a rule-based architecture for this specific task when neural models are proving superior. We give preference to the manually written rules in this project for a number of reasons. Firstly, supervised neural networks require abundant parallel corpora. For the moment, such corpora are not accessible or simply missing for German. Secondly, simplification as such is often not seen as a stand alone task. However, when it comes to simplification without summarization Siddharthan (2014) points out:

> "Statistical approaches are not well equipped to handle simplifications that require syntactic reordering, morphological transformations, and insertions due to lack of explicit linguistic knowledge" (from Suter et al. (2016))

The lack of unified guidelines, exclusion of children as a target group, underestimated role of syntactic simplification have all motivated the current project to tackle these gaps in the NLP research.

3 Methodology

This section describes the data set, its syntactic analysis, and a set of rewrite rules derived from the analysis.

3.1 Data

The corpus we have used was collected by Aumiller and Gertz (2022). It contains about 2900 article pairs from German Wikipedia and its version for 8-12 year-old children called Klexikon. The documents are preprocessed so that the figures, captions, external links are removed, the subtitles are marked with a double equal sign.

The articles are document-aligned, but not sentence-aligned. The reason for that is that the Klexikon entries were written independently from the main Wikipedia articles and share only the titles. Thus, the Klexikon articles may include some information blocks that are spread across multiple Wikipedia articles. This complication makes the sentence alignments impossible.

Below is an example snippet of such a document pair under ID 2361 (color added). The color indicates the shared pieces of information between the texts, while the grayed-out parts are not shared by the documents.

> *Wikipedia:*
>
> Ein Alphabet (frühneuhochdeutsch von kirchenlateinisch alphabetum, von altgriechisch ao alphabetos) ist die Gesamtheit der kleinsten Schriftzeichen bzw. Buchstaben einer Sprache oder mehrerer Sprachen in einer festgelegten Reihenfolge. Die Buchstaben können über orthographische Regeln zu Wörtern verknüpft werden und damit die Sprache schriftlich darstellen.
>
> *Klexikon:*
>
> Ein Alphabet oder Alfabet besteht aus Buchstaben. Das sind Zeichen, die zu Lauten gehören. Aus den Buchstaben kann man Wörter zusammensetzen. Das Alphabet ist eine bestimmte Art zu schreiben.

At first sight, one can notice that some Wikipedia sentences correspond to multiple Klexikon sentences and the unshared parts can either be included inside a shared sentence or form a separate one. Thematically, Klexikon omits extended knowledge (e.g. *(frühneuhochdeutsch von kirchenlateinisch alphabetum, von altgriechisch ao alphabetos)*) of the Wikipedia and includes more simple explanations for potentially difficult words (e.g. *Das sind Zeichen, die zu Lauten gehören.*).

3.2 Syntactic Analysis of the Corpus

In order to spot more systematic differences between the two corpora, we performed a syntactic analysis of the whole corpus for both Wikipedia and Klexikon articles. Analysing how Klexikon behaves syntactically is important in order to come up the meaningful set of rewrite rules for the simplification system. For this purpose, we used SpaCy library version 3.3.

Klexikon's most frequent verbs are descriptive verbs like *gibt*, *nennt*, *heißt*, *bedeutet*, *hat* and simple action verbs like *leben*, *kommt*, *machen*, *geht*, *sieht*, *bauen* etc. In contrast to Klexikon, Wikipedia consists of less descriptive verbs (which are also more academic, e.g. *bezeichnet*, *gilt*, *verwendet*), more aspective (*begann*), causative verbs (*führte*) and other verbs describing complex processes and relations between concepts and objects (*eingesetzt*, *gegründet*, *erhielt*, *entstand*). Overall, Klexikon's most frequent verbs are less diverse in semantics (descriptive and action verbs), which makes this corpus more descriptive than explanatory.

Verbs, auxiliaries and pronouns reach a greater proportion in Klexikon compared to Wikipedia. On the other hand, adjectives and proper nouns are more present in Wikipedia. Adjectives are also more diverse in form and semantics in Wikipedia (*zahlreich*, *heutig*, *politisch*, *sogenannt*, *ehemalig*) compared to Klexikon (*ander*, *groß*, *verschieden*, *klein*, *weiter*, *ganz*). This is a sign that children's corpus utilizes more sentences and relies more on verbal rather than adjectival descriptions in which concepts are connected through SVO rather than adjectival modifiers and noun phrases as well as pronouns rather than synonyms. This is also confirmed by the relative length of the modifiers: 1.3 words for Klexikon and 1.5 words for Wikipedia.

Speaking of clause and sentence level, the average sentence length for Wikipedia is 15 words, while in Klexikon it reaches 10 words per sentence. Interestingly, coordinated sentences are slightly higher in proportion in Wikipedia, while subordinated are prominently more present in Klexikon. However, sentences with coordinated clauses tend to be shorter in Klexikon with 16 words per sen-

Table 1: Summary of the syntactic analysis of Wikipedia and Klexikon

Syntactic feature	Wikipedia	Klexikon	Resulting rule
Sentence length	15 words	10 words	
Clause number	1.5 (max. 5) clauses	1.2 (max. 3) clauses	
Passive clauses	20%	8%	transform >16-word
Passive sentence length	24 words	14 (max. 17) words	sentences to active
Coordinated clauses	3.2%	2.7%	split >22-word
Coordinated sentence length	26 words	16 (max 23) words	sentences
AMS (measure argument of adjectives)	0.07%	0.14%	deabbreviate and add
Semicolon-separated clauses	1.5%	0%	split

tence on average, which is 6 words longer than an average Klexikon sentence, compared to 26 words in Wiki, with and 11 words difference from an average Wiki sentence length.

AMS (measure argument of adjective, e.g. “7 years”, “5 centimeters”) are 2 times as frequent in Klexikon than in Wiki. This is due to Klexikon employing less implicitness ("2002" instead of "year 2002"). One more feature of Klexikon in this regard is that it uses less abbreviations compared to Wiki (*centimeters* instead of *cm*).

With respect to passive voice, Klexikon contains more than twice as little passive voice clauses as Wiki. Moreover, sentences with passive clauses in Klexikon have an average length of 14 words. In Wiki, this number is remarkably higher – 24 words in passive sentences.

Appositions are 3 times as frequent in Wiki than in Klexikon. Those in Klexikon in most cases fulfill explanatory function as how to pronounce a word or putting a difficult term in simple words (e.g. *Constitution, the main law, . . .*).

Klexikon sentences typically contain 1.2 clauses, while those of Wikipedia – 1.5. The maximum number of clauses in Klexikon reaches 3, in Wiki – 5.

3.3 Syntactic Rules

Table 1 briefly summarizes those analysis results that are relevant for the rule definition.

For the execution of rules, we used SpaCy library 3.3 with the large German model and the verb conjugation dictionary from Pattern library developed by Andreas Motl[1].

The pipeline is freely available here.

[1] gist.github.com/amotl/pattern

3.3.1 Rule 1: Expanding AMS

The AMS extending part performs a regular expression search for predefined abbreviations. Those abbreviations have been extracted from the entire corpus and manually compiled into a correspondence dictionary. It includes 132 most common abbreviations from Mathematics (*mm, lb*), Physics (*Hz, kW*), Informatics (*KB, GB*), academic life (*WS, SS*), everyday routine (*Sek, $*) etc.

After the candidates are detected, they are extended and passed to the declination module which adjusts the form to the appropriate case as in the example below.

> *Original sentence:*
>
> Am 29. März 2014 wurde das mit 48.750 m^3 Gesamtinhalt größte Aquarium der Welt, das Chimelong Ocean Kingdom eröffnet.
>
> *Transformed sentence:*
>
> Am 29. März 2014 wurde das mit 48.750 Kubikmetern Gesamtinhalt größte Aquarium der Welt, das Chimelong Ocean Kingdom eröffnet.

3.3.2 Rule 2: Adding AMS

The AMS adding rule inserts implicit “im Jahr” before years. It employs regular expressions and additionally searches for a verb or an auxiliary next to the digits. This is used to prevent incorrect insertion when the digits stand for something other than a year. As shown in the example, the second year is unchanged because its immediate context does not include a verb.

> *Original sentence:*
>
> Sie stimmte 2010 sowohl für die provisorische EFSF als auch 2012 für den

ESM mit dem Ziel der Stabilisierung des Euros.

Transformed sentence:

Sie stimmte im Jahr 2010 sowohl für die provisorische EFSF als auch 2012 für den ESM mit dem Ziel der Stabilisierung des Euros.

3.3.3 Rule 3: Splitting Semicolon

Semicolons are detected with a regular expression. The parse tree filters out verbless clauses that are separated with a semicolon and cannot function as a sentence on their own. The semicolon-separated lists inside parenthesis are also not split.

Original sentence:

Die Aare ist mit ca. 485 m^3/s der wasserreichere Fluss (Rhein: 439 m^3/s); aus hydrologischer Sicht ist also der Rhein ein Nebenfluss der Aare, nicht umgekehrt.

Transformed sentence:

Die Aare ist mit ca. 485 m^3/s der wasserreichere Fluss (Rhein: 439 m^3/s). Aus hydrologischer Sicht ist also der Rhein ein Nebenfluss der Aare, nicht umgekehrt.

3.3.4 Rule 4: Splitting Coordination

The coordination part is fully dependency-based. A sentence with 23 or more words is searched for a coordinated verb, it is combined with a subject of that verb and detached as a separate sentence together with all its children. The subject is deliberately not substituted with a pronoun to avoid coreference problems.

Original sentence:

Im März 1925 ahnte Torrio den Zusammenbruch des Syndikats voraus und entschloss sich, von diesem Mordanschlag geschwächt, auszusteigen und die Macht vollständig Capone zu überlassen.

Transformed sentence:

Im März 1925 ahnte Torrio den Zusammenbruch des Syndikats voraus. Torrio entschloss sich, von diesem Mordanschlag geschwächt, auszusteigen und die Macht vollständig Capone zu überlassen.

3.3.5 Rule 5: Changing Passive to Active

The passive module is dependency-based. It receives a sentence which is 17 words or longer and passive voice is searched for with the help of morphosyntactic features. Those features assume that the sentence should contain an auxiliary with lemma *werden* and a past participle. Potentially false detections, for example when *werden* and a verb belong to different clauses, are filtered out.

For each passive verb, it searches for either a passivized subject (expressed with *von*) or assigns *man* as an impersonal subject.

After the subject is determined, the verb is changed to the correct tense form. The module handles three tenses – Präsens, Perfekt and Plusquamperfekt. When the final conjugated form in the correct tense is determined, the sentence parts are rewritten in the correct word order. If the original sentence had a reversed word order (e.g. with the past participle in the first place for an emphasis), the resulting sentence changes it back to a regular word order with the subject in the first place.

Original sentence:

Der Gattungsname Betula wurde 1753 in Species Plantarum, 2, S. 982 f. erstveröffentlicht.

Transformed sentence:

Man hat den Gattungsnamen Betula 1753 in Species Plantarum, 2, S. 982 f. erstveröffentlicht.

3.3.6 Rules Integration

The rules are applied on word-level first with all AMS handling, and on clause-level last. On clause-level, semicolon splitting is performed first, then for every detached sentence the pipeline applies coordination split. Every separate sentence from the previous modules undergoes passive transformation if applicable.

In the end, the final text is compiled in which sentences are joined and the extra punctuation marks are removed.

3.4 Evaluation Metrics

Admittedly, the most reliable metric to evaluate text simplification remains human evaluation. However, there is a number of techniques and formulae for automatic evaluation reported in the literature. Most of them are word-based, such as SARI, which means that the source, target and predicted texts are

compared as to how many words they share. Since the corpus we have is not sentence-aligned and we do not have a proper word-level simplification this kind of metrics are not suitable.

3.4.1 The Flesch Readability Score

One of the metrics that does not compare the words in the texts directly is the Flesch readability score. The Flesch score was originally developed for English to score school reading material. It has been reported to correlate well with human-assigned scores (Gillen et al., 1977) and later it was adapted for German language by Amstad (1978). The formula calculates a score on the scale from 0 to 100 taking into account the average numbers of syllables per word and words per sentence:

$$RES = 180 - 58,5 * ASW - ASL,$$

where RES is the Reading Ease Score, ASW is the average number of syllables per word and ASL is the average sentence length. A decrease in the score corresponds to an increase in complexity.

3.4.2 Syntactic Distributions

As the goal of syntactic simplification was to bring Wikipedia's syntactic structure closer to that of Klexikon, this should be reflected in the syntactic distributions of the corpora. This metric is rather quantitative, so it cannot show directly how simple the text is for comprehension, but it can show how simple the text is structurally. The distributions of interest are those of parts of speech (POS), sentence length, length and proportion of appositions, relative, subordinated, coordinated, and passive clauses.

3.4.3 Human Evaluation

One German native speaker was invited to judge the quality of the transformations. The texts were evaluated against three criteria. Simplicity and grammaticality scores were assigned to individual sentences and coherence – to the whole text based on how logical and connected the ideas were presented. The scores were given on a Likert scale from 4 to 1. See the evaluation criteria in Appendix A.

3.5 Summarization

In order to compare our results with the previous findings, we used an off-the-shelf extractive summarizer.

Two freely available models were tested – statistical extractive Multilang Summarizer[2] and an abstractive transformer-based trained summarizer[3]. The latter was discarded because of the lack of adequacy in the produced summaries.

The Multilang Summarizer was set up to produce summaries of 436 words which is the average length of the Klexikon articles.

4 Experiment and Results

4.1 Data Set Reduction

Due to limited computational power, we had to reduce the data set by more than half to overcome stack overflow issues with the statistical summarizer. The resulting subset included 1,071 article pairs. All evaluation presented in the next section is performed on this subset.

4.2 Data Sets for Evaluation

The evaluation was performed on 5 different corpora. The Flesch metric and syntactic distributions were calculated for the entire subset corpora specified in subsection 4.1. The corpora include:

- Original Wiki;
- Summarized Wiki;
- Simplified Wiki;
- Simplified and summarized Wiki;
- Klexikon (target)

Due to the length of the articles from original and simplified corpora, human evaluation was done only on 5 documents from summarized, simplified and summarized, and target corpora.

One of these 5 documents was additionally evaluated in its original and simplified versions – sentences that were transformed by the simplification pipeline were extracted and evaluated without the unchanged rest of the article. This was necessary to gain more insight into the simplification performance on its own.

4.3 Evaluation

Figure 1 depicts part of speech distribution in the five corpora. Among the core POS, the adverbs, auxiliaries and verbs show a slight drop after simplification although in theory coordination splits

[2]Multilang Summarizer
[3]German Finetuned Summarizer

Table 2: Human evaluation scores for five German Wikipedia articles

Articles version	Simplicity score	Grammaticality score	Coherence
Summarized articles	1.4	1.3	3.0
Simplified and summarized articles	1.4	1.9	4.0
Klexikon articles	1.0	1.0	1.0

Notes. The scores are averages on the scale from 5 to 1 where 1 is the best.

Table 3: Human evaluation scores for one German Wikipedia article: transformed sentences only

Article version	Simplicity score	Grammaticality score	Coherence
Original article (78 sentences)	1.1	1.0	–
Simplified article (138 sentences)	1.8	2.0	2.3
Simplified article (30 passive sentences)	2.0	2.3	–
Simplified article (92 coordinated sentences)	1.8	1.9	2.6
Simplified article (6 AMS-containing sentences)	1.5	1.3	–

Notes. The scores are averages on the scale from 5 to 1 where 1 is the best.

should increase the number of auxiliaries. Summarization seems to increase their number, thus moving towards the target corpus. Adjectives, nouns and pronouns, on the contrary, move in the right direction after the simplification.

On the sentence level, the simplified corpus noticeably reduces the number of words per sentence, about one word more than the target requires (Figure 2). Summarizing drops this number further by 1-2 words.

As for the types of clauses shown in Figure 3, only subordinated clauses are more present in Klexikon than in the original Wikipedia and the rest should be reduced. In fact, simplification reduces the proportion of coordinated and passive clauses. Remarkably, this transformation also helps the summarizer produce less passive voice compared to its unsimplified output.

The average clauses length (Figure 4) is also successfully reduced by the simplification for the coordinated clauses. The passive clause length, on the other hand, rises. This must be connected with the fact that passive-to-active transformation usually requires 1-2 more words.

In terms of syntactic closeness, the simplification does show positive trends. At least quantitatively, the simplified corpus's syntactic structure makes an improvement over the original articles decreasing the sentence length, noun and adjective rate, coordination and passive proportions.

In Table 4, Flesh scores for all versions of the articles are provided. Surprisingly, the readability score is very high even for the original Wikipedia and that of the Klexikon exceeds the maximum of 100. These extremes might mean that the formula should be adapted not only to a language but also to a specific domain of discourse. However, the general trend in these numbers indicates that the simplified corpus is simpler than the original.

Table 4: Flesh score statistics for different versions of the German Wikipedia articles

Articles version	Flesh score
Original articles	91.03
Simplified articles	94.82
Summarized articles	99.93
Simplified and summarized articles	106.15
Klexikon articles	110.34

As for the quality of the texts, human evaluation scores for 5 articles are summarized in Table 2. Compared to simplified summaries, the summaries of the original articles are noticeably better in terms of grammar and coherence and almost the same in terms of sentence simplicity. This is in line with the earlier work (Vale et al., 2020) which also found no improvement of summarization if prior simplification is used.

It is worth noting though that the summarizer itself proved to perform very poorly with the coherence score of 3 that corresponds to "very hard to understand".

Table 3 shows results for one article from the original and simplified corpora. Here, the evaluated sentences are only those that were transformed. Since in the original article the sentences were taken from different places of the text, the coherence score is not provided. However, in the trans-

Figure 1: Part of Speech Distributions

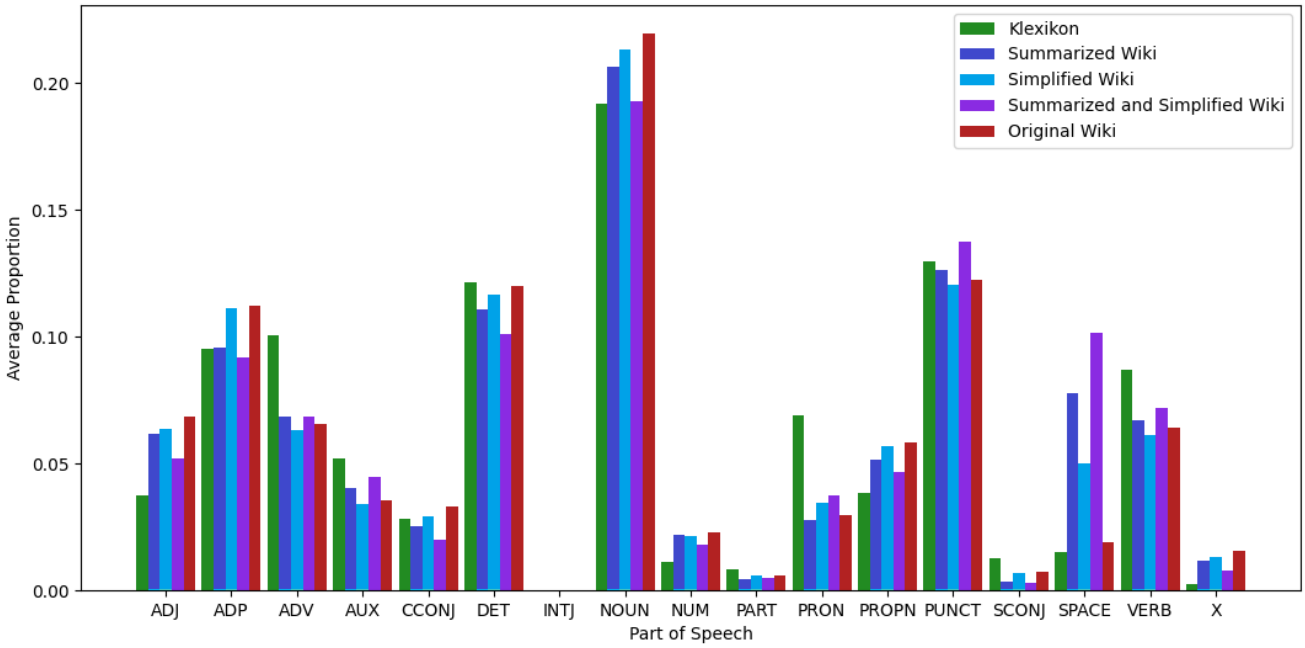


Figure 2: Average Sentence Length (in words)

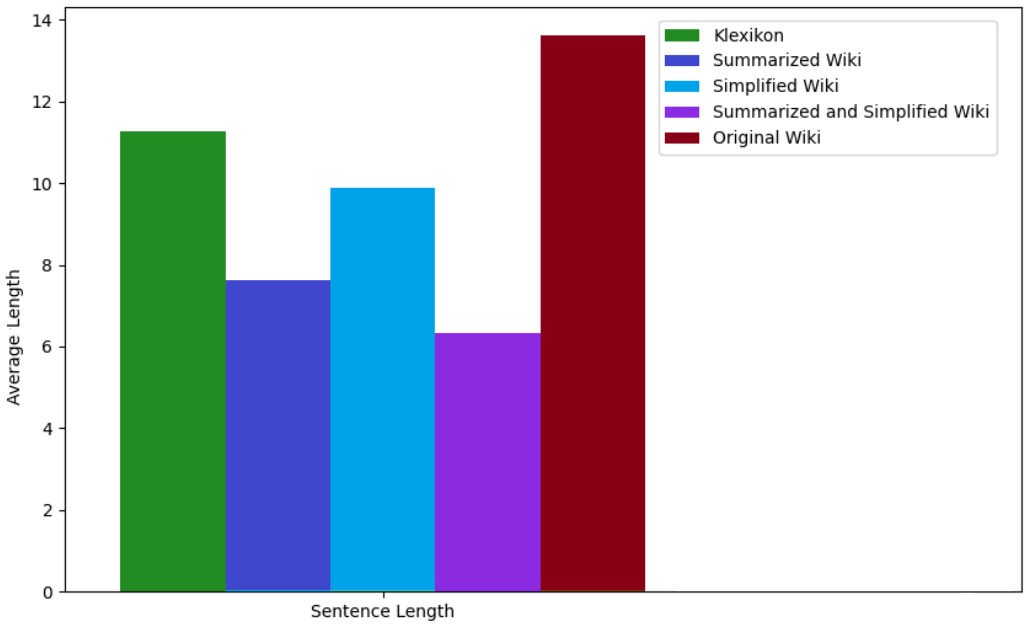


Figure 3: Clause Distributions

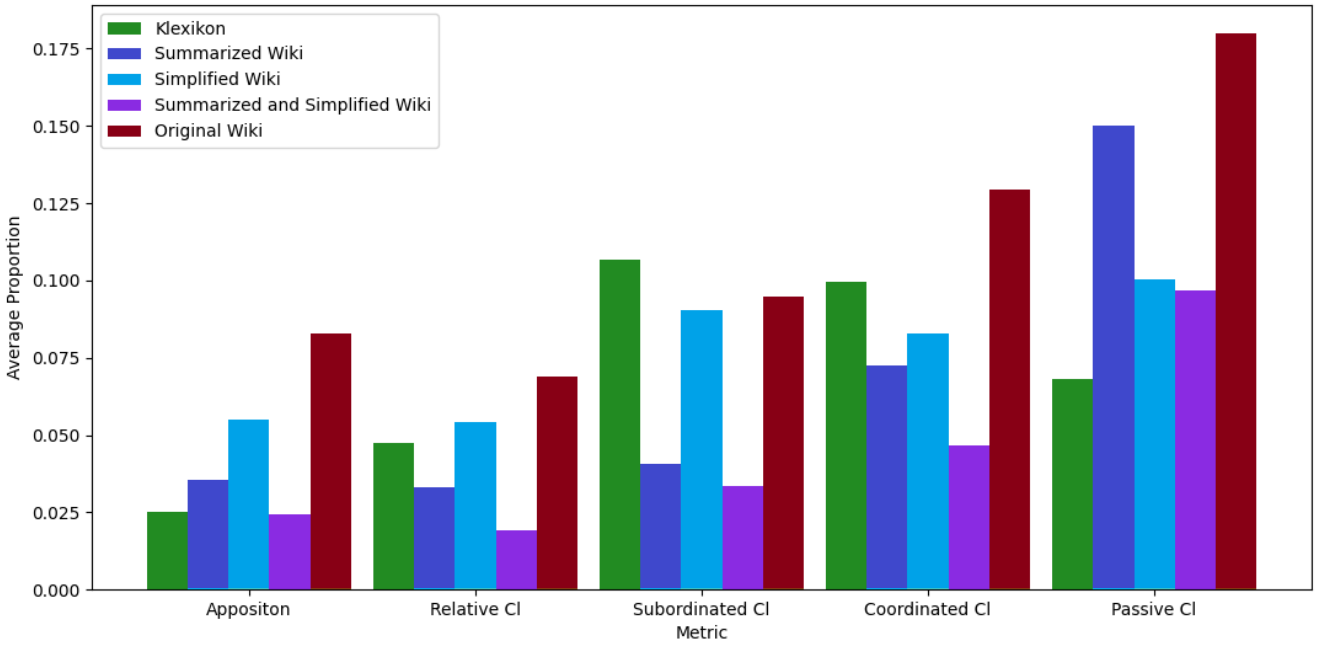
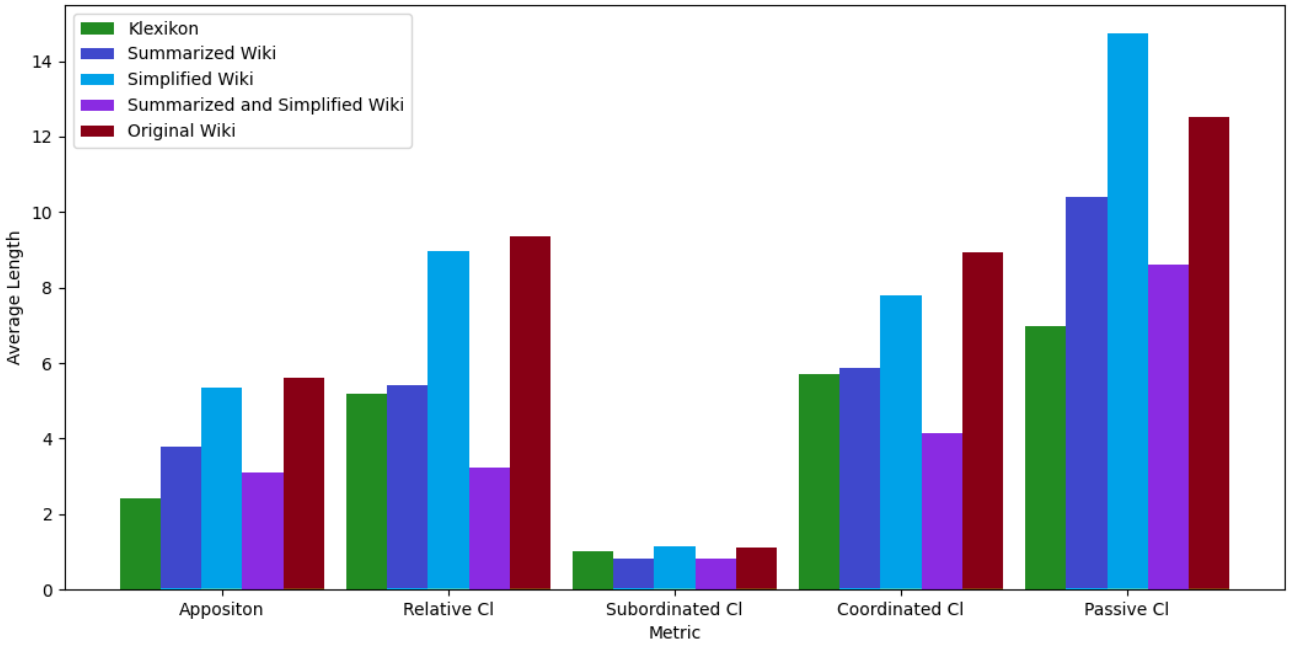


Figure 4: Length of Clause Types (in words)



formed version the sentences often form a chunk of text (e.g. when coordinated sentences are detached) that should follow each other. For that reason we provide the coherence score for this version.

The upper part of the table shows overall scores for all types of transformations. The lower part shows scores for individual types of changes. Overall, the scores are within the range of "not completely grammatical" and "not so easy" but do not reach the 3 which would be "very ungrammatical" and "hard", taking more than 2 times to read. Individually, the scores reflect the complexity of the transformations, with the word-level AMS being the simplest of the transformations to passive changes being the most complex to perform.

5 Discussion

Even thought we have decided to compare the results to the previous work using summarizers, those that are available have produce a very poor text quality. It suffers from a lot of repeated text pieces and unreasonable sentences even when applied to the original articles. Hence, we cannot argue that simplification in this experiment failed to improve the summarizing as neither of the tasks were performed at a satisfactory level. On the contrary, when taking the quantitative metrics only, we observe an improved performance if summarization is preceded by the simplification rather than not.

The reason for this poor performance at least for the simplification system, we believe is, rooted in the lack of the NLP tools for analysis and basic language transformations. SpaCy capacity is limited to detect the tense, mood and voice with high precision.[4] Nor are there free productive libraries for German such as declination, conjugation and other verb-related tools. The correctness of these fundamental operations are especially crucial for clause-level transformations, in our case for passive and coordination rules. For example in the passive transformation, some verbs require a specific prepositions in the active voice which is not expressed in the passive form and hence cannot be recovered from the context.

Another limitation connected with the rule-based approach is linguistic ambiguity. This ambiguity is almost ubiquitous so that every rule had to handle it. In the AMS rule, some abbreviations are context dependent such as *min* (*Minute* or *mindestens*). In the passive module, the so called Zustandspassive was left out because of its identical form with the Perfekt. It appears that these phenomena could only be disambiguated by neural models that might be used in combination with the rules to guide the analysis of the text.

The syntactic analysis introduced in subsection 3.2 motivated more rules than are presented here. The originally planned rule set contained participial modifiers and appositive phrase transformations. Since these kinds of changes also require a precise tackling of verb and noun forms and because of the limited time resources, those rules had to be abandoned.

Another potential extension for the system could be a noun-to-pronoun change when the coordination split takes place. Originally, this change was intentionally avoided because it could present coreference issues if it is blindly applied every time. If done carefully, this change might refine the final text coherence. Raising the coherence score might be a drastic improvement for the whole system as simplicity and grammaticality have not been shown to impede comprehension on the sentence level.

6 Conclusion

In this project we have developed a simplification system for German texts. This system aims to adapt the syntactic structure of the original text to be simpler for children at the age of 8-12. The simplification is rule-based and all the rules were driven by the syntactic analysis of the target corpus.

The quantitative evaluation of the syntax of the produced texts with the Flesch readability score and syntactic distributions showed the desired difference between the original corpus and its simplified version. However, the human evaluation points out that there is still room for improvement in terms of the quality of the simplification output. Coherence suffers the most from the pipeline's unpredicted behavior and this is the most crucial criterion for the discourse meant for children.

All in all, the result of this work has revealed an acute need for more fundamental NLP tools like tense recognition or a conjugation library for German before sophisticated and complex systems such as simplification can be developed. This might serve as a good basis for software development not only for German but also for other non-English languages.

[4] One library available for German was no longer maintained.

References

Toni Amstad. 1978. *Wie verständlich sind unsere Zeitungen?* Universität Zürich.

Dennis Aumiller and Michael Gertz. 2022. Klexikon: A German Dataset for Joint Summarization and Simplification. *arXiv e-prints*.

Barry Gillen, Philip C. Kendall, and A.J. Finch Jr. 1977. Reading ease and human interest scores: A comparison of flesch scores with subjective ratings. *Teaching of Psychology*, 4(1):39–41.

Christiane Maaß. 2015. *Leichte Sprache: Das Regelbuch*. LIT-Verlag, Berlin, Germany.

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 *Third Workshop on New Frontiers in Summarization*, pages 152–161.

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

Julia Suter, Sarah Ebling, and Martin Volk. 2016. Rule-based automatic text simplification for german. In *13th Conference on Natural Language Processing (KONVENS 2016)*, pages 279–287.

Rafaella Vale, Rafael Lins, and Rafael Ferreira. 2020. An assessment of sentence simplification methods in extractive text summarization. In *DocEng'20: ACM Symposium on Document Engineering 2020*, pages 1–9.

A Appendix

These are the human evaluation criteria.

SIMPLICITY: sentence-based

How simple/easy is the sentence to understand?

1 – Easy, took me one time to read

2 – Not easy, thought about some references (e.g. pronoun, relationships between objects/concepts) for a bit

3 – Hard, confusing, took me quite some time to make sense of the sentence

4 – Very hard, the sentence doesn't make sense to me

GRAMMATICALITY: sentence-based

How grammatically correct is the sentence? Are there any violations to the grammar?

1 – Grammatically correct, haven't seen any violations

2 – Not completely correct, but I could recover the correct form with little effort (takes 1 time to read and fix the form and the sentence meaning is still clear)

3 – Incorrect, took me some time (2-3 times to read) to understand the meaning of the sentence

4 – Completely incorrect, I could not make sense of the sentence because of the severe grammar violations

COHERENCE: article-based

How logical and connected are the ideas in the article?

1 – The article makes perfect sense to me

2 – The article is logical overall, but there are some occasional unclear sentences/references that don't fit the context

3 – The article is very hard to understand, not very logical, I could only get a general idea and some specifics

4 – The article overall makes no sense at all