TS-ANNO: An Annotation Tool to Build, Annotate and Evaluate Text Simplification Corpora

Regina Stodden and **Laura Kallmeyer**
Heinrich Heine University
Düsseldorf, Germany
{stodden,kallmeyer}@phil.hhu.de

Abstract

We introduce TS-ANNO, an open-source web application for manual creation and for evaluation of parallel corpora for text simplification. TS-ANNO can be used for i) sentence–wise alignment, ii) rating alignment pairs (e.g., w.r.t. grammaticality, meaning preservation, ...), iii) annotating alignment pairs w.r.t. simplification transformations (e.g., lexical substitution, sentence splitting, ...), and iv) manual simplification of complex documents. For evaluation, TS-ANNO calculates inter-annotator agreement of alignments i) and annotations ii).

1 Introduction

A large number of texts are difficult to understand for many people, e.g., people with low literacy skills, non-native speakers, or people with cognitive disabilities (Alva-Manchego et al., 2020b). Text simplification (TS) aims to make complex texts more accessible by editing their wording and syntax, while preserving the original meaning (Alva-Manchego et al., 2020b).

In automatic TS, parallel corpora that align sentences from the original text with corresponding professionally simplified sentences are precious resources for training and evaluating TS systems. Currently, however, high-quality corpora of this type are rare and often of comparably small size (e.g., Zero Hora (Caseli et al., 2009) or Terence & Teacher (Brunato et al., 2015)). Therefore, often resources that were not designed for TS in the first place are used to train TS systems (e.g., Simple Wikipedia texts (Coster and Kauchak, 2011; Hwang et al., 2015)) (Štajner, 2021). As text simplification is often performed on sentence-level, two further problems of TS corpora arise: missing sentence-level alignment (e.g., see Newsela (Xu et al., 2015)) or error-prone automatic sentence alignment (e.g., see PWKP (Zhu et al., 2010)) (Štajner, 2021).

Furthermore, TS corpora are provided, if at all, with the alignment, e.g., WikiLarge (Zhang and Lapata, 2017) or Wiki-Auto (Jiang et al., 2020). Only a few corpora contain information about the actual types of simplification (simplification transformations, respective grammaticality, lexical complexity etc. of the aligned sentences, etc.) (e.g., see PorSimple Corpus Caseli et al. (2009), SimpleSEW corpus Amancio and Specia (2014) or Terence & Teacher corpus Brunato et al. (2015)). Collecting such data is difficult but could be useful to analyze the advantages and limitations of TS systems (Alva-Manchego et al., 2020b).

To facilitate the creation of manually annotated high-quality TS corpora with $n{:}m$ alignments of full documents, we developed TS-ANNO, an open-source, language-independent, all-in-one web application. The web application supports:

- web scraping of parallel websites and local files upload,

- $n{:}m$ sentence-wise manual alignment,

- sentence-wise rating of grammaticality, simplicity, coherence and ambiguity,

- pair-wise rating of meaning preservation, information gain, overall simplicity, structural simplicity and lexical simplicity,

- pair-wise annotation of (fine-grained) simplification transformations on word-level, phrase-level, sentence-level and paragraph-level, and

- evaluation of the data collected.

The main functionalities of the annotation tool are illustrated in Figure 1. A demonstration of the tool and a demo video can be found here `https://ts-anno.phil.hhu.de/`.[1] The source code is also available to create an own copy of the annotation tool.[2]

---

[1] Register yourself to annotate on a test basis or log in as *test*-User (password: *TS_anno22*) to test evaluations and downloads.

[2] https://github.com/rstodden/TS_annotation_tool

```
1) ⚒ Build & Upload    2) ≋ Annotation
1.1)    Upload         2.1)    Sentence Pairs
1.1.A)⚡ Web Crawling    2.1.A)⛴ Simplification
1.1.B)▯ Local Upload    2.1.B)⇄ Alignment

1.2)    Preprocessing  2.2)    Rating
1.2.1)▤ Sentence Splitting      ᴬᴮ Fluency, Adequacy, …
1.2.2)☰ Tokenization   2.3)    Identifying
                               Transformations
                               ⛏ Split, Replace, …

3) Evaluation      ⚖     4) Export            ⤓
3.1)    Inter-Annotator-  4.A)    Meta Data
        Agreement         4.B)    Alignment
3.1.A)👥 Cohen's Kappa     4.C)    Ratings
3.1.B)👥 Fleiss' Kappa     4.D)    Transformations
                          4.E)    Data Sheet
```

Figure 1: Visualization of TS-ANNO's main functionalities: i) build & upload, ii) annotation, iii) evaluation, and iv) export. The numbered lists describe consecutive steps, whereas the lettered lists describe alternatives.

In the remainder of the paper, we elaborate technical details and functionalities of TS-ANNO and exemplify its usage.

2 TS-ANNO: Preparing the data

This section introduces implementation details, user administration, data uploading and pre-processing options of TS-ANNO.

2.1 System Architecture

TS-ANNO is an open-source web-based application implemented in Python (Version 3.8)[3] using the Django web framework (Version 4)[4] and PostgreSQL[5] for the underlying database structure. NGINX[6] is used to configure the server of TS-ANNO. The responsive interface of the application is designed with Bootstrap (Version 4.5)[7]. Currently, all interface instructions are in English, but they can also be translated to other languages if required. For each annotation step, the time is measured to identify more or less difficult corpora or domains. To further develop the tool, contributions by the community are welcome by participation on GitHub or by using the changelog function of the annotation tool itself.

2.2 Administration + User Management

Django comes by default with an administrator interface. In our case, it is helpful for the control of corpora, users and annotations. In addition to the basic user information of Django, upon registration, users are asked for some optional demographic characteristics, such as native language and language level of the language to annotate. This information can help to better understand the users' ratings, especially regarding simplicity, which has been shown to be subjective (Štajner, 2018).

2.3 Uploading and Metadata

TS-ANNO supports different approaches for data insertion. Either local data can be uploaded or online data can be automatically crawled and read to the database.

Local Upload. TS-ANNO permits the upload of parallel documents either as plain texts, paragraph segmented texts, or pre-aligned texts.[8] If no simple version of a document exists, the "to_simplify" option can be ticked to add the data for the manual simplification part of the tool. Furthermore, before uploading, additional metadata regarding copyright, domain and language levels of the documents are requested to incorporate practices for responsible data (re-)use (Rogers et al., 2021) from the start.

The functionality of uploading pre-aligned data is illustrated in the online demo with the manual simplifications of the ASSET corpus (Alva-Manchego et al., 2020a). This corpus is selected because a further analysis regarding the rewriting transformations applied in combination with fine-grained manual ratings of the simplification seems to be a relevant supplement to ASSET.

Web Crawling. More and more manually and professionally simplified texts are available on websites. Among others, it might be due to the recommendation of the European standard for digital accessibility (European Telecommunications Standards Institute, 2021) to provide easy-to-read texts on (at least public authority) websites. Mostly these texts are aligned with parallel versions in standard language.

---

[8]TS-ANNO does not support automatic sentence alignment, but it can handle already aligned sentences. Before uploading, the texts can be aligned manually or with any automatic alignment algorithm (e.g., MASSalign (Paetzold et al., 2017), CATS (Štajner et al., 2018), or a neural CRF model (Jiang et al., 2020)).

Battisti et al. (2020) have shown that TS can benefit from these websites by developing a web crawler to download their texts, images and typography. In order to also access the texts of the parallel web pages, TS-ANNO integrates a web crawler (build with the Python library Beautiful Soup[9]) that enables the extraction and alignment of these valuable documents. The web crawler automatically aligns complex documents with parallel, simple documents and recognizes paragraph endings.[10]

Currently, the system contains example web crawlers for the websites of "Inclusion Europe"[11] and "Alumniportal Deutschland"[12]. On the website of Alumniportal Deutschland, openly licensed, parallel German documents (original: CEFR level B1-B2, simple A1-A2) are published, which exemplify the annotation of everyday documents in TS-ANNO. The website of Inclusion Europe includes parallel complex-simple documents in four languages (German, English, Spanish and French), which make possible to create a multi-lingual simplification corpus with TS-ANNO.

2.4 Pre-processing

During upload, no matter whether local or online, the data will be pre-processed with a language-specific NLP pipeline of SpaCy (Version 3) (Honnibal et al., 2020).[13]

Possible pre-processing problems, e.g., segmentation errors or HTML left-over of crawling, can be reported per sentence in the interface.

3 TS-ANNO: Annotation

The main functionality of TS-ANNO is the annotation, which comprises alignment, transformation annotation, rating and manual simplification. Overall, the annotation is structured by the corpora to annotate. Each corpus contains at least one document pair, which in turn consists of a complex document and a parallel simplified document (except for manual simplification). For detailed instructions on how to annotate and how to use the annotation tool, we refer to the annotation guidelines.[14]

---

[9]https://www.crummy.com/software/BeautifulSoup/
[10]If the web crawler is used, please pay attention to the copyright of the texts of the websites.
[11]https://www.inclusion-europe.eu/
[12]https://www.alumniportal-deutschland.org/
[13]SpaCy currently supports 18 languages, 44 are planned and a multi-language model exists for other languages.
[14]https://github.com/rstodden/TS_annotation_tool/tree/master/annotation_schema

The online demo contains pre-aligned sentence pairs of ASSET (Alva-Manchego et al., 2020a) and a few annotated document pairs of Alumniportal to illustrate the functionalities of TS-ANNO.

3.1 Aligning Sentences

Alignment is the process of finding and grouping text elements, such as documents, paragraphs or sentences, of at least two parallel or comparable texts with a quite similar meaning. In TS-ANNO, in any case, complex and simple documents are aligned when uploading the data (see subsection 2.3). Depending on the input data, sentence pairs can also be already aligned during the upload. The annotation tool further supports the manual alignment (or alignment correction) of paragraphs and sentences. Each complex and simple sentence co-occurs with a button that highlights the most similar sentence(s) in the corresponding text (based on SpaCy's word embeddings) to facilitate sentence alignment.

In TS, sentence alignment pairs mostly contain only one sentence *of each* document (1:1, e.g., copying or rephrasing the text). However they can also contain only one sentence *of one* document (1:0 or 0:1, e.g., sentence omitting for removing unimportant text or sentence insertion for explanations), several sentences in *one* document (1:n or n:1, e.g., splitting a sentence or merging sentences) or several sentences in *both* documents (n:m, e.g., sentence fusion) (Alva-Manchego et al., 2020b). As illustrated in Figure 2, any number of sentences of both documents can be selected, therefore, TS-ANNO supports all of the named (n:m) alignments.

All sentences which are identical in the complex and simplified document are automatically aligned and disabled in the front-end to speed-up manual alignment. Furthermore, after the manual alignment of a document, all not aligned simple and complex sentences are automatically aligned as insertion or omitting.

Currently, for most existing corpora, the alignments are automatically generated. However, TS-ANNO does not include alignment algorithms due to their questionable quality. The automatic alignment models can often only identify pairs of (nearly) identical sentences but cannot correctly extract "strong paraphrases" where the structure and semantic were highly changed (Štajner, 2021). Therefore, at least for test sets, manually alignment or alignment post-editing are highly recommended.

Text Simplification Annotation Tool

Complex Document:

abhängig von der Region in Deutschland – verschiedene Namen.

☑ Von „Fasching" spricht man in Teilen Bayerns und auch in Norddeutschland, von „Fastnacht" unter anderem in Hessen, Baden und Württemberg.

☑ Der „Karneval" hat seine Hochburgen am Rhein wie in Köln oder Düsseldorf.

☐ Es gibt verschiedene Theorien über die Wurzeln des Karnevals.

Simple Document:

nach Region in Deutschland – verschiedene Namen.

☑ Von „Fasching" spricht man in Teilen Bayerns und auch in Norddeutschland, von „Fastnacht" unter anderem in Hessen, Baden und Württemberg, der „Karneval" ist besonders populär am Rhein in Städten wie Köln oder Düsseldorf.

☐ Traditionell beginnt Karneval/Fastnacht/Fasching am Dreikönigstag, dem 6. Januar.

☐ An vielen Orten beginnt die „fünfte Jahreszeit" aber schon

Text Source Complex Document: https://www.alumniportal-deutschland.org/digitales-lernen/deutsche-sprache/lesetexte/b1-b2/online-deutsch-lernen-uebungen-karneval-1-b/ (Last accessed: March 29, 2021)
Text Source Simple Document: https://www.alumniportal-deutschland.org/digitales-lernen/deutsche-sprache/lesetexte/lesetexte-sprachniveau-a1-a2/online-deutsch-lernen-uebungen-karneval-1-a/ (Last accessed: March 29, 2021)

List of aligned sentence pairs:

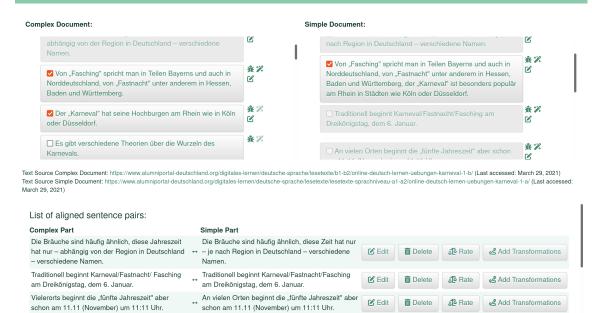| Complex Part		Simple Part				
Die Bräuche sind häufig ähnlich, diese Jahreszeit hat nur – abhängig von der Region in Deutschland – verschiedene Namen.	↔	Die Bräuche sind häufig ähnlich, diese Zeit hat nur – je nach Region in Deutschland – verschiedene Namen.	✏ Edit	🗑 Delete	⚖ Rate	⚒ Add Transformations
Traditionell beginnt Karneval/Fastnacht/ Fasching am Dreikönigstag, dem 6. Januar.	↔	Traditionell beginnt Karneval/Fastnacht/Fasching am Dreikönigstag, dem 6. Januar.	✏ Edit	🗑 Delete	⚖ Rate	⚒ Add Transformations
Vielerorts beginnt die „fünfte Jahreszeit" aber schon am 11.11 (November) um 11:11 Uhr.	↔	An vielen Orten beginnt die „fünfte Jahreszeit" aber schon am 11.11 (November) um 11:11 Uhr.	✏ Edit	🗑 Delete	⚖ Rate	⚒ Add Transformations
Sie beginnt mit Weiberfastnacht, auch		Sie beginnt mit Weiberfastnacht, auch				

↻ Reset   💾 Save

Figure 2: Screenshot of TS-ANNO during the sentence alignment of two complex and one simple sentences of a German document pair of Alumniportal Deutschland. The upper part contains the complex document on the left and the simple document on the right. Per each document, all sentences are listed, including paragraph markers and three buttons: i) a bug button to report content issues, ii) a button to focus most similar sentences in the counterpart, and iii) an edit button to change the alignment of the sentence. In the lower part, all already aligned sentences are shown accompanied with buttons i) to edit or ii) to delete the alignment, iii) to rate the aligned sentence pair or iv) to add the rewriting transformations of the pair.

3.2 Transformation Annotation of Alignment Pairs

In the simplification process, various different rewriting strategies can be applied to the complex texts and, hence, it can result in several different simplified sentences. For the creation of some TS corpora, annotators were asked to simplify the text following a given list of rewriting transformations, e.g., Barancikova and Bojar (2020); Alva-Manchego et al. (2020a) or were asked to add the transformations during (or after) alignment, e.g., Caseli et al. (2009); Bott and Saggion (2011b); Amancio and Specia (2014); Brunato et al. (2015).

Enriching aligned corpora with transformation annotations could help to improve TS systems by adding a preceding sequence labeling step for transformation identification (e.g., Dong et al. (2019); Kumar et al. (2020); Omelianchuk et al. (2021)). TS systems with other approaches could benefit from the transformation annotations by splitting and mixing the data splits based on the annota-

tions or gaining more insights into the generated simplifications (Alva-Manchego et al., 2020b). Furthermore, the number of different transformations can be used to quantify the "simplicity gain" (Xu et al., 2016).

TS-ANNO permits the annotator to choose affected tokens in the alignment pair and to assign transformation labels to it; multiple labels are possible. The tokens can optionally be color-coded, e.g., red for delete, orange for replaced, and blue for added, to emphasize the changes in the sentences. The label can correspond to the general transformation level (paragraph, sentence, phrase or word), or a transformation class name can be specified per level. In addition, the labels can be chosen even more fine-grained as for some transformation classes sub-transformation labels are provided.

The transformation levels, classes and sub-transformations can be dynamically changed to consider language-wise differences and preferred annotation schemes. As default, TS-ANNO contains transformations that were used (with similar

terms) in existing TS annotation schemes (Bott and Saggion, 2014; Brunato et al., 2015; Gonzalez-Dios et al., 2018; Koptient et al., 2019): 1. *delete*, 2. *insert*, 3. *merge*, 4. *reorder*, 5. *split*, 6. *lexical simplification*. In addition, we add *verbal changes* as transformation because, in German text simplification, the verb's voice or mood are often changed. A list of the default labels of TS-ANNO is provided in Appendix A.2.

3.3 Rating of Alignment Pairs

Following Alva-Manchego et al. (2020b), human assessment of system predictions is for now the most reliable evaluation method of text simplification systems. Furthermore, rating of the simplification pairs helps to reveal the unclear initial state of a (web-) corpus, e.g., i) are the original and simplified sentences grammatically correct, ii) is the (simplified) sentence really simple, iii) to which extent are the simplified sentences simpler than the original sentences or iv) to which extent are the pairs lexically or syntactically simplified. An imbalance in the data could for instance lead to training a TS system to only correct grammar issues, to only produce syntactic simplifications or only weak simplifications.

Therefore, TS-ANNO supports relative rating of the aligned sentence pairs, i.e., change between original and simplified sentence (see example iii) and iv)), and absolute ratings of the original and simplified sentence (see example i) and ii)) on a Likert-scale. As many different rating schemes for TS exist (for a summary see Alva-Manchego et al. (2020b); Štajner (2021)), the rating aspects and the rating scale size can be dynamically changed in TS-ANNO.

All aspects are accompanied by a statement for which the annotators are asked to agree or disagree, following Alva-Manchego et al. (2020a); Maddela et al. (2021). An overview of all default aspects, including all statements, is provided in Appendix A.1. The chosen default scale is a 5-point Likert-scale, normally ranging from 1 to 5. However, rating of simplicity is a subjective task (Štajner, 2018), hence, different ratings are expected. To ensure that the ratings are due to subjective perspectives on simplicity and not due to different understandings of the scale,[15] the scale endpoints of some

aspects can be changed to -2 to $+2$ to emphasize the meaning of the lowest (-2, reverse change) and the middle point (0, no change). Furthermore, annotation guidelines with the annotation scheme chosen should be handed out to all annotators.

3.4 Manual Simplification

As an additional feature, TS-ANNO supports sentence-wise manual simplification. Most of the available simplification corpora focus on Wikipedia data or news texts (Štajner, 2021). However, in 2014, Pellow and Eskenazi already justified the need to simplify everyday documents. Therefore, we encourages the simplification of texts of other domains, such as illustrated with how-to-articles of wikiHow[16] (Koupaee and Wang, 2018) in the online demo. Furthermore, the manual simplification option can be used to generate alternative simplifications of existing simplification pairs for a better evaluation (Alva-Manchego et al., 2020a).

After the complex data upload, an annotator can select at least one sentence of a complex document on the left and add a simpler version in the text box on the right. To ease the simplification, the guideline of Inclusion Europe is linked as simplification instructions for easy-to-read language (CEFR level A1);[17] as soon as the ISO standard of plain language is published these instructions will also be linked for plain language (CEFR level A2 to B1).[18]

Furthermore, TS-ANNO exemplary integrates the multi-lingual TS system MUSS (Martin et al., 2020) to provide suggestions on how to simplify the marked complex sentence(s); it can be easily exchanged with other TS systems.

4 TS-ANNO: Evaluation and Export

4.1 Evaluation / Inter-Annotator Agreement

The annotation tool also provides some evaluation approaches. The inter-annotator agreement (IAA) of the alignment and the rating is calculated per all corpora, each corpus, each domain and each document.[19]

However, a low level of IAA does not always indicate a bad quality of the annotations. On the

---

[15]Stodden (2021) shows that annotators of TS corpora have different understandings of the lowest scale point in simplicity rating, i.e., either same and higher complexity or only higher complexity.

[16]https://www.wikihow.com/Main-Page

[17]Guidelines in many languages: https://www.inclusion-europe.eu/easy-to-read-standards-guidelines/

[18]The ISO/WD 24495-1 is currently under development see: https://www.iso.org/standard/78907.html.

[19]Depending on the number of annotators, either Cohen's Kappa (2 annotators) or Fleiss' Kappa ($>$ 2 annotators) is calculated.

one hand, it can be due to annotation errors, but on the other hand, it can also be due to different subjective perspectives on the task (see Reidsma and op den Akker (2008)). Rating of simplicity is such a subjective task. Hence, we plan to implement a disagreement metric, similar to the polarization index of (Akhtar et al., 2019), that relates annotation choices to the demographic characteristics of the annotators.

4.2 Data Export

Alignment Export. TS-ANNO supports three formats of alignment export: i) parallel files with a simplification instance per line as most common practice is TS research (Xu et al., 2016; Alva–Manchego et al., 2020a), ii) crossed sentence pairs of full documents with a label specifying whether aligned or not (e.g., see Jiang et al. (2020)), or iii) parallel files with a continuous document text per line (e.g., see Sun et al. (2021)). The first format encourages sentence-level simplification, the second training of a automatic sentence alignment model and the third document-level simplification.

Annotation and Rating Export. Furthermore, the data can also be exported in a CSV file containing all annotated information per aligned sentence pair per user, e.g., one column per evaluation aspect and one column per transformation. The output can be filtered per corpus and per annotator.

Following the recommendations of Prabhakaran et al. (2021) on transparency and increased utility of datasets for downstream use cases, in the export of the annotation tool, all annotations of all raters are included without any aggregation to keep all possibly subjective annotations and to facilitate evaluation with disagreements.

Metadata Export. TS-ANNO provides automatic support for completing data sheets based on Huggingface Data Cards[20] as demanded for producing responsible NLP (Rogers et al., 2021). The system makes proposals for the data sheets questions based on the given metadata, e.g., demographics of the annotators or domains of the corpora.

5 Use Case

TS-ANNO was already tested by aligning German parallel web texts, e.g., the openly licensed, parallel documents of the website "Alumniportal

Deutschland". The documents have been uploaded with the internal web crawler and were automatically pre-processed. Two annotators have manually aligned the sentence pairs and annotated them with rewriting transformations and rating aspects.[21] The demonstration version of TS-ANNO contains a few annotated documents of this corpus, including different alignment types, ratings and rewriting transformations to exemplify the evaluation and download options.

Overall, the test usage of TS-ANNO led to some minor improvements, which are already included in this system description.

6 Related Work

Several aligned TS corpora exist, some with annotated rewriting transformations and some test sets with rated evaluation aspects (see (Alva-Manchego et al., 2020b) as an overview). However, most authors have not provided a reusing option for their annotation interfaces or have only focused on one of the TS annotation tasks, e.g., aligning, simplification, rating, or transformation annotation. Therefore, we present the related work per TS task.

Alignment. The most comparable manual alignment annotation frameworks to TS-ANNO are Tiedemann (2006); Bott and Saggion (2011a); Paetzold et al. (2017), and Jiang et al. (2020), which show also both documents (the original and simplified), in parallel and highlight paragraph blocks. However, in contrast to TS-ANNO, ISA (Tiedemann, 2006) supports only 1:1 alignment, and, in the annotation interface of Jiang et al. (2020), sentence borders can be changed as they are highlighted and not tagged as in our tool.

Simplification. For writing simplifications, several commercial and non-commercial tools, or computer-aided translation software exist to facilitate writing easy to understand texts. Some of them offer more writing support than the others: e.g., LanguageTool for German,[22] FriendlyReader for

---

[21]Both annotators are German native speakers, trained in linguistics as well as simple languages, have at least a graduation diploma, and were paid for their work with at least the minimum wage of their country of residence. The annotators were provided with the following annotation guidelines and instructions on how to use the annotation tool: https://github.com/rstodden/TS_annotation_tool/blob/master/annotation_schema/Annotationsrichtlinien_TS_anno-DE.pdf.

[22]https://languagetool.org/de/leichte-sprache/

Swedish,[23] or Hero-App for English and Dansk,[24] offer most of it by highlighting complex passages and showing rewriting suggestions. The simplification annotation editor of Caseli et al. (2009) also offers rewriting suggestions, which the annotator can accept or decline.

TS-ANNO does not support rewriting suggestions yet and is more similar to the simplification procedure described in Alva-Manchego et al. (2020a), in which crowd workers were provided only with annotation guidelines and should rewrite the sentences without any assistance. However, they added their simplifications within a crowd-sourcing platform and not in an annotation tool.

Rating. Even if manual rating of TS output is performed on many TS system outputs, no shared annotation tool for rating exists. Often crowdsourcing platforms, e.g., Amazon Mechanical Turk or Figure Eight, are used to ask the rating questions, such as for the ASSET Corpus (Alva-Manchego et al., 2020a) and the Simplicity-DA Data Set (Alva-Manchego et al., 2021).

Transformations. The annotation of rewriting transformations is a general sequence labeling problem, therefore, popular sequence labeling tools such as BRAT (Stenetorp et al., 2012) could be used. Gonzalez-Dios et al. (2018) adapt BRAT for TS rewriting annotation by pointing an affected sequence in the original sentence and labeling it with the transformation. In contrast, Koptient et al. (2019) annotated the transformations on the word-level at the parallel text using a modification of YAWAT (Germann, 2008). TS-ANNO is more similar to YAWAT as it facilitates the annotation within a parallel setting, but extends it via the annotation of also phrase-, sentence-, and paragraph-level transformations.

To the best of the authors' knowledge, no annotation tool exists yet, combining all needs of building text simplification corpora, i.e., manual simplification, pair rating, transformation annotation and sentence alignment.

7 Limitations

So far, TS-ANNO has been tested only for English, Spanish, French, Farsi, and German. SpaCy, which is used for pre-processing, should support many languages, though maybe not always be with the

same quality. For languages that SpaCy does not support, corpora have to be sentence split and tokenized before uploading them. However, it is also possible to exchange SpaCy with another NLP framework.

The annotation tool does not support active learning yet; its current focus is on high-quality manual alignment, rating and annotation of parallel data.

Štajner (2021); Alva-Manchego et al. (2020b) state that rating is not enough to evaluate TS texts, comprehension tests, measuring reading time and eye fixations are also relevant. Unfortunately, TS-ANNO only supports rating as it is yet the most dominant approach.

8 Conclusion & Further Work

We presented TS-ANNO, an open-source, web-based application for the purpose of facilitating the time-consuming process of building high-quality corpora for text simplification or of evaluating quality of existing corpora. The annotation tool combines relevant functionalities for building TS corpora, e.g., crawling parallel web documents, sentence-wise alignment of parallel documents, human assessment of alignment pairs, annotation of transformations applied to get the simplified sentence of the pair and evaluating the data, e.g., via inter-annotator agreement. The human assessment of simplification pairs could help to identify and filter out pairs of sentences with an increase of complexity (rather than a decrease) from the complex to the simple document. Furthermore, transformation labels allow characterizing the performed changes. Both information could be used to evaluate TS systems and give insights in their "black boxes".

In future work, we plan to extend the simplification option by highlighting complex phrases. In addition, it would be interesting to integrate active learning or a neural alignment system, which could suggest possible sentence alignments. It is also planned to add more websites to the crawler option of TS-ANNO to facilitate creating corpora with texts of other languages or other domains.

Acknowledgements

---

[23]http://www.friendlyreader.se/
[24]https://heroapp.ai/en/

9 Ethics/Impact Statement

Intended Use. In general, simplifying texts makes sense in order to give more people access to information. In addition to analog information dissemination via newspapers or books, digital information dissemination via websites is becoming increasingly relevant. In addition to technical barriers, such as the readability and user guidance of a website, the complexity or language level of texts on websites also plays an important role. The simpler the text, the more people can understand it. The more people understand the text, the more people have the chance to get involved, e.g., in the form of discussions or dissemination of information.

The here proposed annotation tool is designed to create parallel datasets with aligned sentence pairs, one complex and one simplified for TS. Additionally, the sentence pairs can be rated regarding their complexity, grammaticality, meaning preservation and added with transformations applied during the simplification process. The output, a dataset for TS, could be used to train an automatic TS system which helps people with reading problems to understand more texts.

Failure modes & Misuse Potential. The usage of the annotation tool highly depends on what the administrator and user do with it. Undesired texts could be inserted or annotators could write harmful comments or simplifications. The resulting corpus could be misused to train a system to make texts even more difficult to read than, as intended, more simpler. Also, the annotations (ratings, alignments or transformation annotations) could be intentionally manipulated or unintentionally wrongly preprocessed by the underlying systems, e.g., sentence splitting or tokenization.

However, the dataset's quality is essential for the use case of TS, as people will rely on it. Hence, TS-ANNO tries to support quality checks of the produced data, e.g., by calculating the inter-annotator agreement, reporting errors in the texts due to preprocessing, and by rating the sentence pairs before publishing the dataset. Users identified as not reliable or even harmful could be banned from the annotation platform by the administrator. Additionally, some unexpected behavior of the annotation tool cannot be precluded entirely. However, users can report issues either on GitHub to the developers or via the changelog feature to the administrator.

Biases. ML systems can get biased based on the data they are trained on, therefore we show texts from different domains in our example. Administrators of TS-ANNO should be aware of this bias and carefully select the texts to annotate. Furthermore, the ratings regarding simplicity are highly subjective, hence, the selection of annotators, e.g., only people with high literacy and language level, can bias the rating. The ratings might be completely different if the target group of the simplified texts would rate the sentence pairs. Administrators should keep it in mind and always describe the annotator group. Therefore, the metadata export of TS-ANNO includes metadata of the texts and the annotators to support the administrators.

Annotators. During registration, annotators are asked to voluntarily add some relevant demographic characteristics, such as native language and language level of the language to annotate. Also, the username is freely selectable, hence, if not desired, no personal information must be shared. All information is optional but highly preferable as they are helpful to understand the users' annotations.

Furthermore, the annotators of the example dataset were paid for their annotations following with at least the minimum wage of Germany, the country of residence.

Computing Time. The annotation tool is a web application, hence, a server is required to run the annotation tool. Administrators can run it either locally or on an external server. No additional computing power or hardware is required. However, to ensure access to the annotation tool, the server needs to be permanently run or the access time should be restricted.

References

Sohail Akhtar, Valerio Basile, and Viviana Patti. 2019. A new Measure of Polarization in the Annotation of Hate Speech. In *Proceedings of the XVIIIth International Conference of the Italian Association for Artificial Intelligence (AI*IA 2019)*, pages 588–603. Springer International Publishing.

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020a. ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679, Online. Association for Computational Linguistics.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020b. Data-Driven Sentence Simplification: Survey and Benchmark. *Computational Linguistics*, 46(1):135–187.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2021. The (Un)Suitability of Automatic Evaluation Metrics for Text Simplification. *Computational Linguistics*, pages 1–29.

Marcelo Amancio and Lucia Specia. 2014. An Analysis of Crowdsourced Text Simplifications. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 123–130, Gothenburg, Sweden. Association for Computational Linguistics.

Petra Barancikova and Ondřej Bojar. 2020. COSTRA 1.0: A Dataset of Complex Sentence Transformations. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3535–3541, Marseille, France. European Language Resources Association.

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.

Stefan Bott and Horacio Saggion. 2011a. An Unsupervised Alignment Algorithm for Text Simplification Corpus Construction. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 20–26, Portland, Oregon. Association for Computational Linguistics.

Stefan Bott and Horacio Saggion. 2011b. Spanish Text Simplification: An Exploratory Study. *PROCESAMIENTO DEL LENGUAJE NATURAL*, 47(0):87–95.

Stefan Bott and Horacio Saggion. 2014. Text Simplification Resources for Spanish. *Language Resources and Evaluation*, 48(1):93–120.

Dominique Brunato, Felice Dell'Orletta, Giulia Venturi, and Simonetta Montemagni. 2015. Design and Annotation of the First Italian Corpus for Text Simplification. In *Proceedings of The 9th Linguistic Annotation Workshop*, pages 31–41, Denver, Colorado, USA. Association for Computational Linguistics.

Helena Caseli, Tiago F. Pereira, Lucia Specia, Thiago Alexandre Salgueiro Pardo, Caroline Gasperin, and Sandra Maria Aluísio. 2009. Building a Brazilian Portuguese Parallel Corpus of Original and Simplified Texts. In *In: 10th Conference on Intelligent Text Processing and Computational Linguistics, Mexico City*, pages 59–70.

William Coster and David Kauchak. 2011. Simple English Wikipedia: A New Text Simplification Task. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 665–669, Portland, Oregon, USA. Association for Computational Linguistics.

Yue Dong, Zichao Li, Mehdi Rezagholizadeh, and Jackie Chi Kit Cheung. 2019. EditNTS: An Neural Programmer-Interpreter Model for Sentence Simplification through Explicit Editing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3393–3402, Florence, Italy. Association for Computational Linguistics.

European Telecommunications Standards Institute. 2021. Accessibility requirements for ICT products and services - EN 301 549 (V3.2.1). https://www.etsi.org/deliver/etsi_en/301500_301599/301549/03.02.01_60/en_301549v030201p.pdf.

Ulrich Germann. 2008. Yawat: Yet Another Word Alignment Tool. In *Proceedings of the ACL-08: HLT Demo Session*, pages 20–23, Columbus, Ohio. Association for Computational Linguistics.

Itziar Gonzalez-Dios, María Jesús Aranzabe, and Arantza Díaz de Ilarraza. 2018. The Corpus of Basque Simplified Texts (CBST). *Language Resources and Evaluation*, 52(1):217–247.

Matthew Honnibal, Ines Montani, Sofie Van Landeghem, and Adriane Boyd. 2020. spaCy: Industrial-strength Natural Language Processing in Python.

William Hwang, Hannaneh Hajishirzi, Mari Ostendorf, and Wei Wu. 2015. Aligning Sentences from Standard Wikipedia to Simple Wikipedia. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 211–217, Denver, Colorado. Association for Computational Linguistics.

Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Model for Sentence Alignment in Text Simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7943–7960, Online. Association for Computational Linguistics.

Anaïs Koptient, Rémi Cardon, and Natalia Grabar. 2019. Simplification-induced Transformations: Typology and some Characteristics. In *Proceedings of the 18th BioNLP Workshop and Shared Task*, pages 309–318, Florence, Italy. Association for Computational Linguistics.

Mahnaz Koupaee and William Yang Wang. 2018. WikiHow: A Large Scale Text Summarization Dataset. *CoRR*, abs/1810.09305.

Dhruv Kumar, Lili Mou, Lukasz Golab, and Olga Vechtomova. 2020. Iterative Edit-Based Unsupervised Sentence Simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7918–7928, Online. Association for Computational Linguistics.

Mounica Maddela, Fernando Alva-Manchego, and Wei Xu. 2021. Controllable Text Simplification with Explicit Paraphrasing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3536–3553, Online. Association for Computational Linguistics.

Louis Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes, and Benoît Sagot. 2020. Multilingual Unsupervised Sentence Simplification. *CoRR*, abs/2005.00352.

Kostiantyn Omelianchuk, Vipul Raheja, and Oleksandr Skurzhanskyi. 2021. Text Simplification by Tagging. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 11–25, Online. Association for Computational Linguistics.

Gustavo Paetzold, Fernando Alva-Manchego, and Lucia Specia. 2017. MASSAlign: Alignment and annotation of comparable documents. In *Proceedings of the IJCNLP 2017, System Demonstrations*, pages 1–4, Tapei, Taiwan. Association for Computational Linguistics.

David Pellow and Maxine Eskenazi. 2014. An Open Corpus of Everyday Documents for Simplification Tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 84–93, Gothenburg, Sweden. Association for Computational Linguistics.

Vinodkumar Prabhakaran, Aida Mostafazadeh Davani, and Mark Diaz. 2021. On Releasing Annotator-Level Labels and Information in Datasets. In *Proceedings of The Joint 15th Linguistic Annotation Workshop (LAW) and 3rd Designing Meaning Representations (DMR) Workshop*, pages 133–138, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Dennis Reidsma and Rieks op den Akker. 2008. Exploiting 'Subjective' Annotations. In *Coling 2008: Proceedings of the workshop on Human Judgements in Computational Linguistics*, pages 8–16, Manchester, UK. Coling 2008 Organizing Committee.

Anna Rogers, Timothy Baldwin, and Kobi Leins. 2021. 'Just What do You Think You're Doing, Dave?' A Checklist for Responsible Data Use in NLP. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4821–4833, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Sanja Štajner. 2018. How to Make Troubleshooting Simpler? Assessing Differences in Perceived Sentence Simplicity by Native and Non-native Speakers. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Paris, France. European Language Resources Association (ELRA).

Sanja Štajner. 2021. Automatic Text Simplification for Social Good: Progress and Challenges. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 2637–2652, Online. Association for Computational Linguistics.

Sanja Štajner, Marc Franco-Salvador, Paolo Rosso, and Simone Paolo Ponzetto. 2018. CATS: A Tool for Customized Alignment of Text Simplification Corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107, Avignon, France. Association for Computational Linguistics.

Regina Stodden. 2021. When the Scale is Unclear – Analysis of the Interpretation of Rating Scales in Human Evaluation of Text Simplification. In *Proceedings of the First Workshop on Current Trends in Text Simplification (CTTS 2021)*, pages 84–95. CEUR-WS.

Renliang Sun, Hanqi Jin, and Xiaojun Wan. 2021. Document-Level Text Simplification: Dataset, Criteria and Baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7997–8013, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Jörg Tiedemann. 2006. ISA & ICA - Two Web Interfaces for Interactive Alignment of Bitexts alignment of parallel texts. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in Current Text Simplification Research: New Data Can Help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing Statistical Machine Translation for Text Simplification. *Transactions of the Association for Computational Linguistics*, 4:401–415.

Xingxing Zhang and Mirella Lapata. 2017. Sentence Simplification with Deep Reinforcement Learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark. Association for Computational Linguistics.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A Monolingual Tree-based Translation Model for Sentence Simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361, Beijing, China. Coling 2010 Organizing Committee.

A Appendix

A.1 Rating Aspects

Aspect	Statement
Grammaticality	The simplified sentence is fluent, there are no grammatical errors.
Grammaticality (original)	The original sentence is fluent, there are no grammatical errors.
Meaning Preservation	The simplified sentence adequately expresses the meaning of the original sentence, perhaps omitting the least important information.
Information Gain	In the simplified sentence, information is added or gets more explicit than in the original sentence.
Overall Simplicity	The simplified sentence is easier to understand than the original sentence.
Structural Simplicity	The structure of the simplified sentence is easier to understand than the structure of the original sentence.
Lexical Simplicity	The words of the simplified sentence are easier to understand than the words of the original sentence.
Simplicity (simple)	The simplified sentence is easy to understand.
Simplicity (original)	The original sentence is easy to understand.
Coherence (simple)	The simplified sentence is understandable without reading the whole paragraph.
Coherence (original)	The original sentence is understandable without reading the whole paragraph.
Ambiguity (simple)	The simplified sentence is ambiguous. It can be read in different ways.
Ambiguity (original)	The original sentence is ambiguous. It can be read in different ways.

Table 1: Default rating aspects of TS-ANNO.

A.2 Rewriting Transformation Label Scheme

Level	Class Name	Sub Transformation
Word	Deletion	Discourse Marker
		Abbreviation
		Filler Words
		Other
	Lexical Substitution	Compound Segmentation
		More Frequent Word
		Abbreviation
		Anaphora
		Shorter Word
		Synonym
		Hyponym
		Hypernym
		Nominalization
		Methaphor
		Number
		Date
		Other
	Inflection	
	Insert	Ellipsis Filled
		Other
Phrase	Reorder	Discontinuity Resolution
		Other
	Deletion	Phrase
		Clause
		Replace
		Less Adjunct Phrase
		Other
	Rephrase	
Sentence	Split	Coordinate Clause
		Subordinate Clause
		Appositive Phrase
		Adverbial Phrase
		Relative Clause
		Other
	Verbal Changes	Voice of Verb
		Verb Tense
		Verb Mood
	Lexical Substitution	Verbalization
		Other
	Reorder	Subject-Verb Reorder
		Genetive to Dative
		Negative to Positive
		Other
	Rephrase	
	No Operation	
Paragraph	Reorder	Sentence-Order Changed
		Other
	Merge	
	Deletion	
	Insert	Explanation
		Exemplification
		Other

Table 2: Default rewriting transformation labels of TS-ANNO.