

Automatic German Text Simplification

Text Simplification

Aim

- Adaption of a given text to improve the text comprehension for a specific target group, e.g.:
 - children,
 - non-native (German) speakers,
 - people with reading problems.

How to Simplify

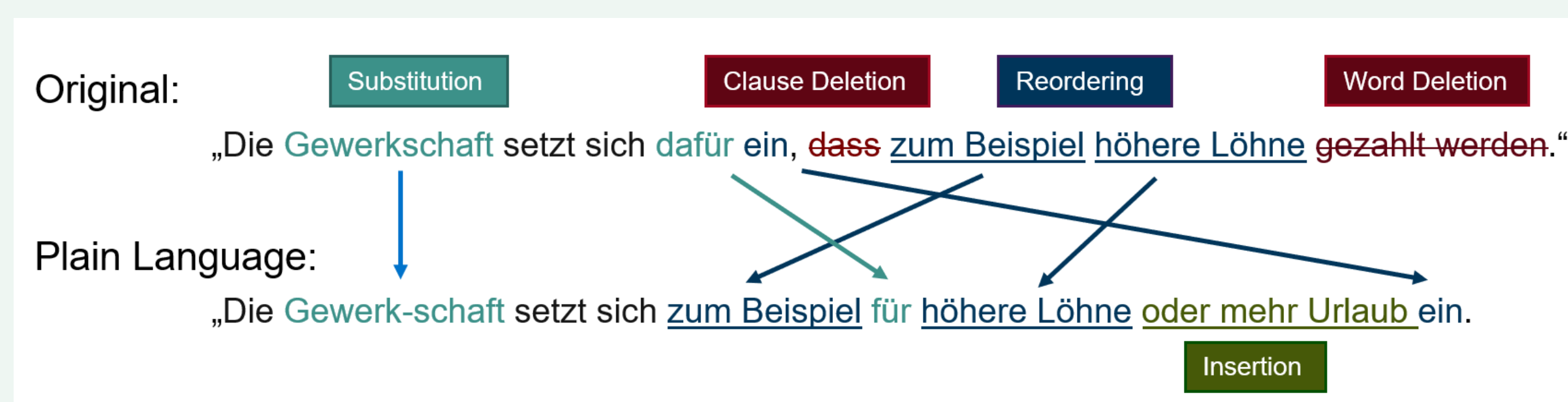
- sentence split,
- complex word substitution,
- rephrasing,
- compound segmentation

Evaluation

- Human Evaluation:
 - grammaticality,
 - simplicity,
 - meaning preservation

Data

Example



Data Set

- original-simple sentence pairs of different text domains, e.g.,
 - medical data (Apotheken Umschau),
 - news texts (Austrian Press Agency),
 - literature (Kafka, Brothers Grimm), etc.
- pairs are annotated with rewriting transformations (see right)
- each sentence rated with evaluation aspects, e.g.,
 - grammaticality,
 - simplicity,
 - coherence, etc.
- pairs rated with evaluation aspects, e.g.,
 - meaning preservation,
 - simplification,
 - information gain, etc.

Annotation

Text Simplification Annotation Tool

Choose all Transformations of the Alignment Pair #3704

Color Coding of Changes: ☒ Green (Normal): Token occurs in both sentences or is already annotated.
☐ Red: Token does not occur in the simple but in the complex sentence(s).
☐ Blue: Token does not occur in the complex but in the simple sentence(s).
☐ Orange: Token might be replaced.

Complex Sentence: 2018 erhielt sie den Hauptpreis des Deutschen Schulpreises .

Simple Sentence: 2018 hat die Schule den „ Deutschen Schulpreis „ bekommen .

Text Source Complex Document: <https://www.inclusion-europe.eu/de/inklusion-hat-keine-grenzen-die-grenzen-gibt-es-nur-in-unseren-koepfen/> (Last accessed: June 3, 2022)
Text Source Simple Document: <https://www.inclusion-europe.eu/de/inklusion-kennt-keine-grenzen-die-grenzen-sind-nur-in-unseren-koepfen-etr/> (Last accessed: June 3, 2022)

Select the level of the transformation and afterwards the type of transformation. If the subtype of the transformation is obvious, please select it as well. Please also mark all tokens which are affected by the current transformation. If no token can be clearly marked, because for example the whole sentence was tran

☒ Word

☐ deletion

☒ lexical_substitution

☐ insert

☐ Phrase

☐ Sentence

☐ Paragraph

Method

Sequence to Sequence

- Insert a complex sentence, get a simple sentence
- Using large language models and transformers
- The "black box" of the neural net, learns on the given examples how to simplify

Sequence Labeling + Downstream Tasks

- Insert a complex sentence → label the tokens with rewriting transformations → apply the transformation to the sentence → get a simplified sentence
- Using large language models and transformers
- The "black box" of the neural net, learns on the given examples how to tag the tokens of the sentence
- Another algorithm, for example, predicts the best candidate to substitute a difficult word

Ethics

Risks and Harms of Text Simplification (Gooding, 2022)

- "The Homogeneity Effect"
 - Does everyone need the same simplification?
 - Do children and non-native speaker both need syntactical simplifications?
- Meaning Distortion
 - Can readers rely on the automatic simplification? (see above example)
 - Can we use automatic simplification for all text domains, medical data or laws?
- Paternalism
 - Who should decide what to simplify? The reader or the translator?
- Misuse Potential
 - Could someone train the system in the opposite direction to build a system to make a text more complex?