Automatic German Text Simplification

Heinrich Heine University Düsseldorf

Regina Stodden[1]

[1]Dept. of Computational Linguistics, Faculty of Arts and Humanities

NRW-FORSCHUNGSKOLLEG ONLINE-PARTIZIPATION

regina.stodden@hhu.de

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

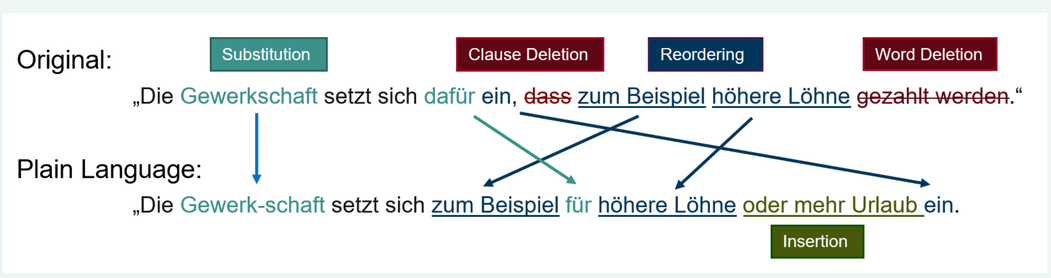
- sentence split,
- complex word substitution,
- rephrasing,
- compound segmentation

Evaluation

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

Data

Example



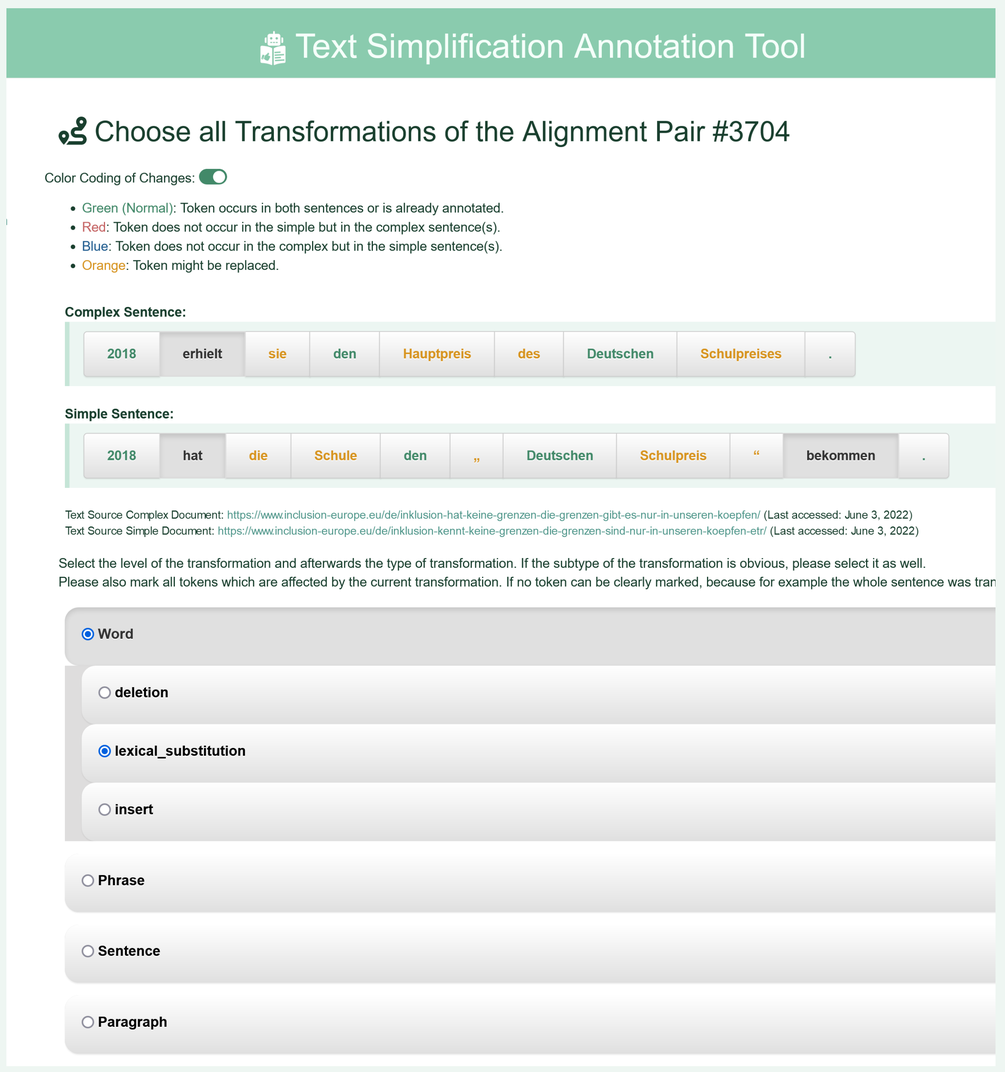
Data Set

- 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



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 $\rightarrow$ label the tokens with rewriting transformations $\rightarrow$ apply the transformation to the sentence $\rightarrow$ 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?