DETECT: Determining Ease and Textual Clarity of German Text Simplifications

Maria Korobeynikova, Alessia Battisti, Lukas Fischer, Yingqiang Gao

Abstract:

Current evaluation of German automatic text simplification (ATS) relies on general-purpose metrics such as SARI, BLEU, and BERTScore, which insufficiently capture simplification quality in terms of simplicity, meaning preservation, and fluency. While specialized metrics like LENS have been developed for English, corresponding efforts for German have lagged behind due to the absence of human-annotated corpora. To close this gap, we introduce DETECT, the first German-specific metric that holistically evaluates ATS quality across all three dimensions of simplicity, meaning preservation, and fluency, and is trained entirely on synthetic large language model (LLM) responses. Our approach adapts the LENS framework to German and extends it with (i) a pipeline for generating synthetic quality scores via LLMs, enabling dataset creation without human annotation, and (ii) an LLM-based refinement step for aligning grading criteria with simplification requirements. To the best of our knowledge, we also construct the largest German human evaluation dataset for text simplification to validate our metric directly. Experimental results show that DETECT achieves substantially higher correlations with human judgments than widely used ATS metrics, with particularly strong gains in meaning preservation and fluency. Beyond ATS, our findings highlight both the potential and the limitations of LLMs for automatic evaluation and provide transferable guidelines for general language accessibility tasks.