
Evaluation of neural coreference annotation of simplified German

Sarah Jablotschkin, Nele Benz & Heike Zinsmeister

Universität Hamburg

Griet.v.Koeln@severinstor-koeln.de, J.a.Rhein@eigelstein-koeln.de,
hennes@zoo-koeln.de

Simplified variants of German (i.e. *Leichte Sprache* and *einfache Sprache*; cf. Bredel/Maaß 2016, Baumert 2019) focus on lexical and syntactic simplification in order to make written information accessible to low-literacy readers. Simplification efforts include using simple sentences and avoiding personal pronouns. However, this also affects coherence phenomena: Relations between nominal form, information status and anaphoric potential are different than we would expect them to be in Standard German (e.g. we would expect a pronoun as anaphor instead of the definite NP in example (1)).

- (1) [Die meisten Besucher] sind verkleidet. [Die Besucher] haben zum Beispiel bunte Kleidung. (LeiKo 1.5 core corpus)
'Most visitors are dressed up. Fore example, the visitors have colourful clothes.'

In our poster, we evaluate the effect of these differences on the performance of state-of-the-art coreference resolution developed on Standard German news texts. Our evaluation is based on the manually corrected core corpus of LeiKo 1.5 (doi: 10.5281/zenodo.6362739) (1,071 sentences / 1,401 coreference relations). The raw coreference resolution was created by the neural c2f coreference resolver (Schröder et al. 2021). We evaluate the performance quantitatively (v.8.01 of the CoNLL-2011/2012 reference scorer, Pradhan et al. 2014) and provide a detailed error analysis. C2f performs robustly on simplified German texts with respect to definite and demonstrative coreference but misses out, e.g., on example (1), which contains the indefinite expression *die meisten* ('most'). Other error sources include speaker reference in reported speech and ambiguous reference. We will also discuss consequences for coreference annotation of a larger parallel corpus of simplified and Standard German currently under construction.

References: • Baumert, A. (2019). Mit einfacher Sprache Wissenschaft kommunizieren. essentials. Wiesbaden: Springer Fachmedien Wiesbaden GmbH. • Bredel, U. & C. Maaß (2016). *Leichte Sprache. Theoretische Grundlagen, Orientierung für die Praxis*. Berlin: Dudenverlag. • Pradhan, S., X. Luo, M. Recasens, E. Hovy, V. Ng & M. Strube (2014). Scoring Coreference Partitions of Predicted Mentions: A Reference Implementation. In Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 30–35. • Schröder, F., H. O. Hatzel & C. Biemann (2021). Neural End-to-end Coreference Resolution for German in Different Domains. In Proceedings of KONVENS 2021, 170–181.