
Empirical evidence for pejoration of ‘woman’-terms: Challenges of estimating inter-annotator agreement with complex annotation

Carla Sökefeld, Heike Zinsmeister

Universität Hamburg

carla.sokefeld@uni-hamburg.de, heike.zinsmeister@uni-hamburg.de

This poster presents a small dataset of manually semantically annotated historical dictionary entries and discusses the challenges of estimating inter-annotator agreement given a complex and – to some extent – subjective annotation task. The annotation study is part of a PhD thesis on semantic change: Many German terms denoting women have historically undergone a pejoration, that is, their meaning has changed to have a more negative bias (e.g., *Frau* has changed from ‘woman with high social status’ to just ‘woman’). To track the semantic change of 31 different German nouns denoting either women or men, a collection of 160 entries in eight historical dictionaries (16th-19th century) were annotated by two annotators (148,000 tokens; entries vary greatly in size from a minimum of two tokens to more than 21,000 tokens). Previous research has shown that women tend to be portrayed stereotypically and negatively while men are shown in a more positive light (Warnke 1993). The tagset crucially differentiates between two levels of semantic information: The *meaning components* given for the lexeme and the portrayal of men and women in *language examples*. This distinction allows for a nuanced analysis: We can trace the semantic change as reflected in the dictionaries and further gain insights into gender stereotypes conveyed in usage examples (see also Nübling 2009). The dataset will be made available on Zenodo together with detailed annotation guidelines.

We will discuss the challenges of estimating a chance-corrected inter-annotator agreement score on this data set due to the variable extensions of the annotation units as well as potential ambiguity or subjectivity (Barteld et al. 2014) and the option of multiple valid annotations (Marchal et al. 2022).

References: • Barteld, F.; Ilden, S.; Schröder, I.; Zinsmeister, H. (2014): Annotating descriptively incomplete language phenomena. In *Proceedings of LAW*, 99–104. • Marchal, M.; Scholman, M.; Yung, F.; Demberg, V. (2022): Establishing Annotation Quality in Multilabel Annotations. In *Proceedings of COLING*, 3659–3668. • Nübling, D. (2009): Zur lexikografischen Inszenierung von Geschlecht – ein Streifzug durch die Einträge von *Frau* und *Mann* in neueren Wörterbüchern. *ZGL* 37/3, 593–633. • Warnke, Ingo (1993): Zur Belegung von ‘Frau’ und ‘Weib’ in historischen deutschen Wörterbüchern des 16. und 17. Jahrhunderts. In B. Hufeisen (ed)., “*Das Weib soll schweigen. . .*” (1. Kor.14,34): *Beiträge zur linguistischen Frauenforschung*. Frankfurt a.M.: P. Lang, 127–152.