Context doesn't matter: The male bias of generic masculines in German remains stable across different context windows

During the last decades, evidence for a male bias in generically used masculines in German has been brought forward by numerous studies applying a multitude of methodological approaches (e.g. from Braun et al., 1998 to Glim et al., 2024). That is, generically intended masculines are understood not as gender-neutral but as mainly associated with male individuals. Only recently, computational studies entered this field of research (Schmitz et al., 2023), and only one computational study has considered that context may play a role in disambiguating generic from specific masculines (Schmitz, 2024). We aim at improving this approach with an LLM-based solution.

The goal of the present study was to investigate whether different amounts of context do indeed make a difference in disambiguating generic masculines. As data we used a corpus of German press texts, already annotated with a pre-selected set of personal nouns; specific masculines; generic masculines; or specific feminines (Müller-Spitzer et al., 2024). Targets were fed into the pre-trained LLM 'bert-base-german-cased' (Devlin et al., 2018) with differing amounts of preceding and following context: 5 and 10 words before and after the target; the entire sentence; and the sentence plus the sentence before and after.

To explore how semantically similar the targets were, we used beta regression to model the cosine similarity between their context-dependent semantic embeddings. Our results indicate that independent of context window, the generic and specific masculine are significantly more semantically similar than the generic masculine and the specific feminine.

Overall, our findings demonstrate that context does not matter when it comes to the male bias of generic masculine forms. The generic masculine remains semantically more similar to the specific masculine than to the specific feminine. We conclude that the male bias in generic masculine forms is retained even when potentially disambiguating context is taken into account.

References

Braun, F., Gottburgsen, A., Sczesny, S., & Stahlberg, D. (1998). Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift für Germanistische Linguistik*, *26*(3), 265–283. https://doi.org/10.1515/zfgl.1998.26.3.265

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *1*, 4171–4186.

Glim, S., Körner, A., & Rummer, R. (2024). Generic masculine role nouns interfere with the neural processing of female referents: Evidence from the P600. *Language, Cognition and Neuroscience*, 1–10. https://doi.org/10.1080/23273798.2024.2387230

Müller-Spitzer, C., Ochs, S., Koplenig, A., Rüdiger, J. O., & Wolfer, S. (2024). Less than one percent of words would be affected by gender-inclusive language in German press texts. *Humanities and Social Sciences Communications*, *11*(1), 1–13. https://doi.org/10.1057/s41599-024-03769-w

Schmitz, D. (2024). Instances of bias: The gendered semantics of generic masculines in German revealed by instance vectors. *Zeitschrift für Sprachwissenschaft*.

Schmitz, D., Schneider, V., & Esser, J. (2023). No genericity in sight: An exploration of the semantics of masculine generics in German. *Glossa Psycholinguistics*, *2*(1). https://doi.org/10.5070/G6011192