

One form, two meanings? The semantics of generic and specific role nouns in German

Dominic Schmitz, Heinrich-Heine-Universität Düsseldorf

According to grammarians, generic masculines in German are gender-neutral in their meaning (cf. Doleschal, 2002). As an example, take the grammatically masculine role noun *Musiker* ‘musician’, which can be used as specific role noun, i.e. to refer to a male musician, or as generic role noun, i.e. to refer to a musician of any gender. Psycholinguistic research of the last decades, however, showed that generically used role nouns are not comprehended as gender-neutral but instead as biased towards male referents (e.g. Gygax et al., 2008; Schunack & Binanzer, 2022). Recently, Schmitz (2023) and Schmitz et al. (2023) added similar findings by way of computational methods. The aim of the present paper is to show that Schmitz et al.’s computational implementations come with noteworthy issues and to offer a computational alternative.

Schmitz (2023) and Schmitz et al. (2023) computed semantic vectors for German role nouns using naive discriminative learning (e.g. Baayen et al., 2011). While this approach is well-grounded in psychological research, the way Schmitz et al. implemented it led to a strong association of the semantics of ‘generic’ and the grammatical masculine, rendering a strong semantic connection between ‘generic’ and masculine forms as little surprising. Further, genericity was treated as a type of inflectional feature, even though it is not.

To circumvent these issues, the present paper proposes the use of instance vectors (Lapesa et al., 2018). Instance vectors are vector representations for individual instances of words rather than of lemmas. For their computation, a window of n preceding and following context words around a given target word is considered. The pertinent instance vector is the average of these n context words. Using instance vectors, no genericity vector is computed and, thus, genericity is neither correlated to other vectors nor treated as inflectional function.

Instance vectors were computed for 3,020 target word attestations based on 75 target words and their corpus attestations from Schmitz (2023). New attestations were sampled from the Leipzig Corpora Collection’s “news” sub-corpus (Goldhahn et al., 2012) where fewer than 10 attestations were contained in the corpus by Schmitz (2023). Instance vectors were computed with $n = 2$, $n = 5$, and $n = 8$ to see whether the amount of context included made a notable semantic difference. Finally, like in Schmitz (2023), the resulting semantic vectors were compared using cosine similarity, a measure regularly used to compare vector similarity. Cosine similarities were computed within a target word for the following comparisons: generic masculine vs. specific masculine; generic masculine vs. specific feminine; specific masculine vs. specific feminine.

Introducing beta regression in generalised additive mixed models using the *mgcv* package (Wood, 2017) in R (R Core Team, 2021), it was tested whether cosine similarity was significantly different for the three comparisons. Number, stereotypicality, word-form frequency, and overall frequency were included as control variables.

The results show that the generic masculine was semantically more similar to the specific masculine than to the specific feminine across all window sizes. The highest degree of similarity was found for the generic masculine and the specific masculine. Depending on window size, the least similar forms are either the generic masculine and the specific feminine ($n = 2$, $n = 5$) or the specific masculine and the specific feminine ($n = 8$).

The findings of the present study are in line with a large body of previous psycholinguistic research on the semantic nature of the generic masculine in German, and, regardless of the aforementioned issued, support the findings by Schmitz (2023) and Schmitz et al. (2023). The implications of the present study are twofold. First, the masculine bias in generic masculines in German is stable across a variety of linguistic methods. Second, computational methods seem to be a meaningful complement to psycholinguistic approaches in research on semantic genericity and gender-neutrality.

References

- Baayen, R. H., Milin, P., Đurđević, D. F., Hendrix, P., & Marelli, M. (2011). An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review*, *118*(3), 438–481. <https://doi.org/10.1037/a0023851>
- Doleschal, U. (2002). Das generische Maskulinum im Deutschen. Ein historischer Spaziergang durch die deutsche Grammatikschreibung von der Renaissance bis zur Postmoderne. *Linguistik Online*, *11*(2). <https://doi.org/10.13092/lo.11.915>
- Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*.
- Gygax, P., Gabriel, U., Sarrasin, O., Oakhill, J., & Garnham, A. (2008). Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes*, *23*(3), 464–485. <https://doi.org/10.1080/01690960701702035>
- Lapesa, G., Kawaletz, L., Plag, I., Andreou, M., Kisselew, M., & Padó, S. (2018). Disambiguation of newly derived nominalizations in context: A Distributional Semantics approach. *Word Structure*, *11*(3), 277–312. <https://doi.org/10.3366/word.2018.0131>
- R Core Team. (2021). *R: A language and environment for statistical computing* (4.0.4). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Schmitz, D. (2023). In German, all professors are male. In J. Pfeifer, S. Arndt-Lappe, H. Dorgeloh, G. Kunter, & C. Uffmann (Eds.), *INGO 6.0. The Proceedings. New empirical Insights about laNguage, presented on a Great day Out in September*. Preprint.
- 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).
- Schunack, S., & Binanzer, A. (2022). Revisiting gender-fair language and stereotypes - A comparison of word pairs, capital i forms and the asterisk. *Zeitschrift für Sprachwissenschaft*. <https://doi.org/10.1515/ZFS-2022-2008>
- Wood, S. N. (2017). *Generalized additive models: An introduction with R*. CRC Press. <https://doi.org/10.1201/9781315370279>