

## Instance vectors as a window to (non-)generic role noun semantics

Dominic Schmitz

*Heinrich-Heine-Universität Düsseldorf*

Grammarians have long assumed that generic masculines in German are gender-neutral (cf. Doleschal, 2002). A word like *Koch* ‘cook’ can be used in a generic way, i.e. to talk about a cook of any gender, and in a specific way, i.e. to talk about a male cook. However, mostly psycholinguistic research has shown that generic masculines are not comprehended as gender-neutral but biased towards maleness (e.g. Schunack & Binanzer, 2022). Recently, Schmitz (2023) and Schmitz et al. (2023) conducted the first studies on generic masculines’ semantics using computational methods. The aim of the present paper is to show that Schmitz et al.’s approach is not without issue and to introduce a potential solution to this issue.

Schmitz (2023) and Schmitz et al. (2023) compute semantic vectors not for whole words, e.g. generic masculines, but for their compositional parts using naive discriminative learning (Baayen et al., 2011). That is, they assume that a word like *Koch* is the sum of its base meaning, case, number, grammatical gender, and genericity (i.e. ‘generic’ or ‘specific’). By doing so, two issues arise. First, the feature ‘generic’ always co-occurs with the grammatical masculine, as only grammatically masculine forms were tagged as generic, leading to a strong correlation of the vectors for ‘generic’ and the grammatical masculine. Second, the authors treat genericity as a type of inflectional feature. However, genericity is not an inflectional feature such as case or number are. Thus, their approach also comes with a theoretical weakness.

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

Replicating the cosine similarity analysis of Schmitz (2023) using the same corpus attestations, beta regression in generalised additive mixed models showed that instance vectors deliver similar results. Across all window sizes ( $n = 2, 5, 8$ ), it was found that the generic masculine was semantically more similar to the specific masculine than to the specific feminine. Overall, the highest degree of similarity was found for the two masculine forms.

The results of the present study demonstrate that instance vectors are a feasible approach to the computational investigation of (non-)generic word semantics. While their results are in line with those of previous computational implementations, they circumvent the issues of the latter.

### 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. doi: 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). doi: 10.13092/lo.11.915
- 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. doi: 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. doi: 10.3366/word.2018.0131
- 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). doi: 10.5070/G6011192
- 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*. doi: 10.1515/ZFS-2022-2008