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

Dominic Schmitz
Heinrich-Heine-Universität Düsseldorf
dominic.schmitz@uni-duesseldorf.de

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 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 comes with a twofold issue and to introduce a potential solution to this issue.

First, their way of computing semantic vector representations for role nouns involved a vector for generic usage which is strongly correlated with the grammatical masculine, rendering all generic role nouns potentially overly correlated with the grammatical masculine. Second, the authors treated genericity 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 computed 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, 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 et al. (2023) showed that instance vectors deliver similar results. Across all window sizes (n = 2, 5, 8), 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: • 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. • 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* 41(2).