

Processing generic masculine forms in typologically different languages

Jan Chromý (jan.chromy@ff.cuni.cz)¹, Mikuláš Preininger¹, Dario Paape², James Brand³

¹Charles University, ²University of Potsdam, ³Edge-Hill University

Background: In many gender-marking languages, masculine forms can refer generically to mixed-gender groups, yet a growing body of psycholinguistic evidence suggests that such forms trigger a male-biased interpretation [1, 2]. However, such a bias may be modulated by crosslinguistic differences in gender marking and by language-specific usage patterns. In particular, the extent to which gender is morphosyntactically pervasive in a given language may affect referent activation, and the availability of gender-inclusive alternatives may increase the male bias of “generic” masculine forms. We present preliminary results from a registered report that investigates the processing of plural role nouns in three languages. Czech and German both have three-gender systems but differ in gender pervasiveness, as Czech marks gender on more word classes than German. English does not have grammatical gender and is used as a control. We test two competing theories. A pragmatics-based account predicts stronger male bias for masculine forms in German due to the availability of gender-inclusive alternatives that would be preferred if an inclusive interpretation was intended. By contrast, an activation-based account predicts stronger male bias in Czech, where the more pervasive gender marking should strengthen the mapping between masculine forms and male referents.

Method: We are in the process of conducting a phrase-by-phrase self-paced reading (SPR) experiment with translation-equivalent materials in Czech (current $n = 164$), English (current $n = 94$), and German (current $n = 151$). Experimental sentences always contain plural masculine role nouns that are paired with either male-only or mixed-gender referents encoded via first names. The design manipulates referent gender (male-only vs mixed), stereotypicality of the role noun (neutral vs female vs male, *cyclists* vs *secretaries* vs *welders*), and linear order (role noun before vs after names). An example item in English is given below. In order to investigate whether frequency patterns, such as masculine forms being less frequent after mixed-gender referents, are predictive of reading times, we used EuroLLM [3] to compute surprisal values for the critical and spillover regions. For each language, a regression model was fitted to RTs across all sentences with surprisal as a predictor, and predicted RTs from the model were compared against empirical RTs.

Results: There was no evidence of any significant effects of the experimental manipulations on (log) RTs, nor on the EuroLLM surprisal values. Nevertheless, EuroLLM surprisal was predictive of human RTs at the critical and spillover regions across all three languages (all $p < 0.01$), though the proportion of variance explained was relatively modest (all $R^2 < 0.1$). At the critical region, EuroLLM surprisal tends to predict longer RTs compared to the actual human data (see Figure 1). A tendency was visible towards mixed names being read more slowly when they followed neutral role nouns in Czech and English, which however does not support either theory’s predictions.

Conclusion: Our self-paced reading study across three languages differing in the pervasiveness of gender marking (Czech, German, and English) revealed no evidence that referent gender or stereotypicality affects the online processing of sentences containing “generic” masculine role nouns. We are currently collecting naturalness ratings for the same sentences, which may reveal effects that come into play during offline processing.

References

1. Gygax, P. *et al.* Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes* **23**, 464–485 (2008).
2. Glim, S. *et al.* Early ERP indices of gender-biased processing elicited by generic masculine role nouns and the feminine–masculine pair form. *Brain and Language* **242**, 105290 (2023).
3. Martins, P. H. *et al.* Eurollm-9b: Technical report. *arXiv preprint arXiv:2506.04079* (2025).

Item Example

Linear order	Ref. gender	Sentence
role first	mixed	The most popular physicians in the hospital are surely <u>Martina and Samuel</u> , but they are not the most responsible.
role first	male	The most popular physicians in the hospital are surely <u>Edward and Samuel</u> , but they are not the most responsible.
names first	mixed	Martina and Samuel are surely the most popular physicians in the hospital, but they are not the most responsible.
names first	male	Edward and Samuel are surely the most popular physicians in the hospital, but they are not the most responsible.

Table 1: Example of an experimental item in English. The critical region is underlined. The bolded noun *physician* is grammatically masculine and but stereotypically neutral in Czech and German.

Figures and Tables

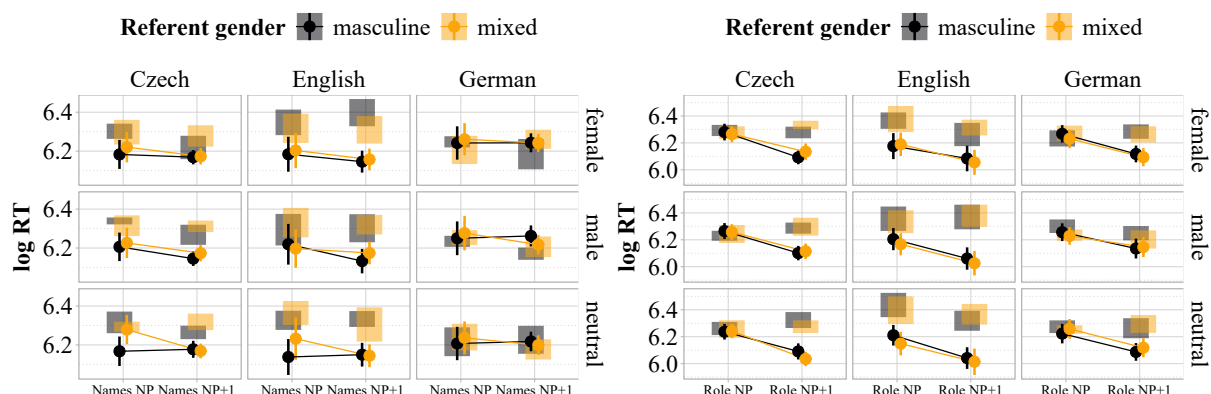


Figure 1: Mean reading times at the critical and spillover regions in role-first (left) and names-first (right) conditions. Error bars show 95% confidence intervals. Wide intervals show predictions based on EuroLLM surprisal.