Decoding gender assignment in an opaque system: Experimental evidence from Czech nonwords

Mikuláš Preininger[1], Martina Abbondanza[1], Lukáš Jiřička[1], Anna Chromá[1], Filip Smolík[1]

[1] *Institute of Psychology, Czech Academy of Sciences, Czech Republic*

Background: Czech exhibits a richly inflected and largely opaque grammatical gender system with four categories: masculine animate, masculine inanimate, feminine, and neuter. Phono-morphological cues to gender are often inconsistent (e.g., *chlopeň*, *srst*, *koleda* feminine; *stupeň*, *prst*, *předseda* masculine), with frequent cross-class syncretism [1]. Yet general ending-to-gender regularities exist, and children anticipate nouns utilizing gender features [2]. We examine formal cues Czech speakers use for gender categorization via nonwords, both in explicit judgments (Exp 1) and implicit activation during reading (Exp 2 and Exp 3).

Method: All three experiments described here were administered online and created using the jsPsych framework. **Experiment 1.** Participants were told they would listen to novel words that represent names of future professions or social roles. In each trial, they saw a pair of faces (feminine and masculine), listened to a nonword, and indicated by keys to which face the nonword fits better (Fig. 1). The nonwords consisted of 20 Czech-like, ortho- and phonotactically permissible pseudo-stems to which five typical nominative singular final segments were added: non-palatalized consonants (C, e.g., *dyr*), their palatalized variants (C', e.g., *dyř*), consonants + *−a* (e.g., *dyra*), consonants + *−ka* (e.g., *dyrka*) and consonants + *−ek* (e.g., *dyrek*). In total, 128 native speakers of Czech (107 women, M_{age}=22.8, SD_{age}=4.4) participated in the study. **Experiment 2.** To accompany the explicit probing in Exp 1, we selected nonwords ending with C (N=20) and *−a* (N=20) from the same list to test whether speakers try to automatically assign gender on encountering new words. We created sentences in which the nonwords functioned as subjects and were subsequently modified by masculine or feminine adjectives, with the context ensuring their inanimate interpretation (Ex. 1). Materials consisted of 80 experimental sentences divided into 4 lists (Latin-square) and 80 fillers also containing nonwords but with no gender marking, yielding 100 sentences per list. Participants (N=327; 263 women; M_{age}=22.6, SD_{age}=4.8) read sentences word-by-word in a moving-window self-paced reading task, with comprehension questions presented after every trial. **Experiment 3.** The experiment followed the same structure as Exp 2, utilizing nonwords ending with –ek (N=20) and –ka (N=20) instead. Responses from a total of 364 participants (296 women, M_{age}=22.61, SD_{age}=4.40) were collected.

Results: For Exp 1, we ran a generalized linear mixed-effects model predicting response (male/female face), with ending as a fixed effect and participant, stem and stem final consonant as random intercepts with varying slopes for different endings. The effect of ending was significant, showing that nonwords ending with C, *−ek* and C'tended to elicit more masculine responses (with the tendency being weaker in C') and nonwords ending with *−ka* and *−a* tended to elicit more feminine responses (Fig. 2) (with the tendency being weaker in the latter). For Exp 2, LMMs on adjective and spillover regions (fixed effects: ending × adjective gender, frequency as covariate; random structure: ending slopes [participants], intercepts [participants and items]) showed no effects on the adjective but a significant interaction on the spillover region (t(17350)=−4.79, p<.001): nonwords ending in *−a* showed faster RTs for feminine adjectives; those ending in C showed faster RTs when followed by masculine adjectives (Fig. 3). Repeating the described steps with Exp 3 again revealed no significant effects on adjectives and a significant effect of the nonword ending»agreement marking interaction (t(12540)=5.96, p<0.001), with potentially no effect for *−ka* (Fig. 4).

Discussion: These findings demonstrate that Czech speakers use word endings to assign grammatical gender in both explicit judgment (Exp 1) and implicit reading (Exp 2, 3) tasks, revealing a robust system of formal gender heuristics that operate independently of semantics. Nonwords ending in C/*−a* were interpreted as masculine/feminine, respectively, despite numerous counterexamples among Czech nouns. Future work will examine how these tendencies reflect actual distributional patterns using corpus anal-

yses.

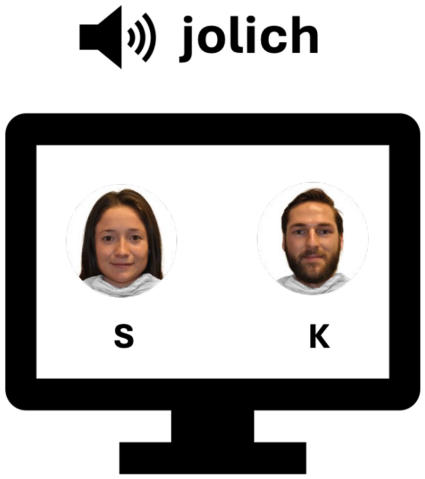

Figure 1: Example of a trial from Experiment 1.

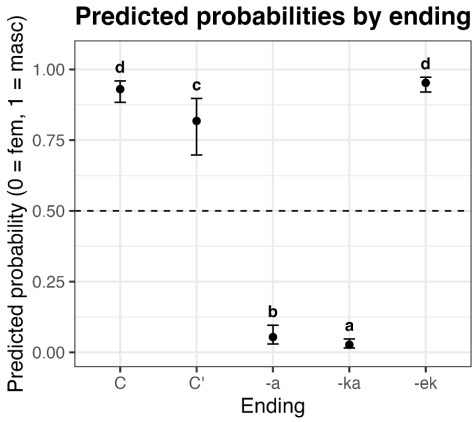

Figure 2: Predicted probabilities of feminine/masculine responses in Experiment 1 by ending. The conditions that significantly differ from each other are labeled with different letters.

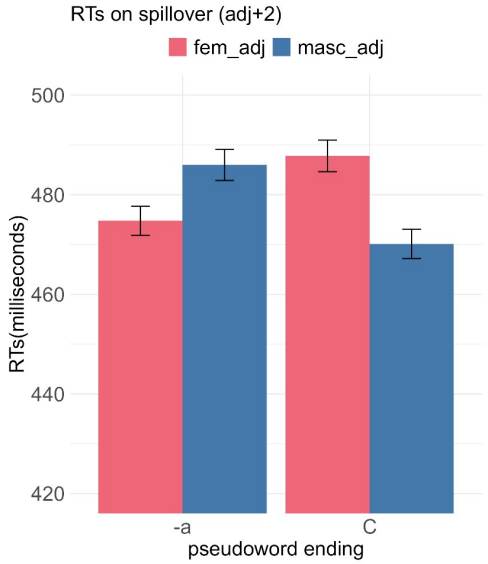

Figure 3: Average reading times on the spillover region from Experiment 2, with respect to gender agreement marking on the adjective.

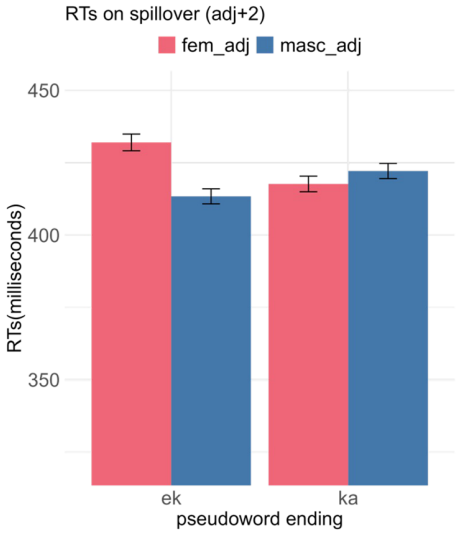

Figure 4: Average reading times on the spillover region from Experiment 3, with respect to gender agreement marking on the adjective.

(1) *Zaznamenali jsme, že dyr/dyr**a** je slepen-ý/slepen-á a ne-jde*
notice$_{:PST:1PL}$ that dyr/dyr**a** be$_{:PRS.3SG}$ stick.together$_{:ADJ-MASC/FEM}$ and NEG-can.be
rozlepit.
separate$_{:INF}$
'We noticed that the dyr/dyra is stuck together$_{[FEM/MASC]}$ and cannot be separated.'

References

[1] Naughton, J., & Kunes, K. V. (2020). *Czech: An Essential Grammar*. Routledge. https://doi.org/10.4324/9781003017530
[2] Smolík, F., & Bláhová, V. (2019). Czech 23-month-olds use gender agreement to anticipate upcoming nouns. *Journal of Experimental Child Psychology*, *178*, 251–265. https://doi.org/10.1016/j.jecp.2018.10.004