

The development of probabilistic grammars in spoken EFL: The case of the genitive alternation

Tanguy Dubois
(UCLouvain and KU Leuven)

Whereas most previous research on alternation phenomena in learner language focuses on the relation between learners' L1 and the choice of variant, this study investigates how proficiency level modulates the probabilistic grammar of learners of English as a Foreign Language (EFL) regarding the use of the genitive variants (e.g., *the dog's tail* vs. *the tail of the dog*). First, I collected genitive observations from the Trinity Lancaster Corpus (Gablasova, Brezina & McEnery 2019), which consists of recordings from an official language exam between a native speaker of British English and low-intermediate (B1) to advanced (C2) learners of English from several L1 backgrounds. Non-interchangeable genitive structures such as partitive genitives (e.g., *the cup of wine*) and classifying genitives (e.g., *a children's book*) were excluded. The remaining 2,302 genitive observations were annotated for various probabilistic constraints, such as the length, animacy and definiteness of the constituents (see Rosenbach 2014). The data was then analyzed via mixed-effects logistic regression, where the constraints were allowed to interact with the speaker's proficiency level so as to reveal how proficiency modulates the constraints' influence on the choice of variant. Other sources of variation, such as learners' mother tongue, were controlled for through random effects. Results show that although native speakers and learners are similar, low-proficiency learners are less sensitive to possessor definiteness and possessor animacy. This corpus-based analysis was complemented with a rating experiment, where 25 native speakers and 101 learners from various L1 backgrounds rated the naturalness of either variant in corpus excerpts (see Engel et al. 2022). The ratings were found to correlate with the predictions from the corpus-based model, indicating that learners' intuitions reflect their language production in the corpus. This correlation is slightly weaker for low-proficiency learners, which can partially be explained by low-proficiency learners again being less sensitive to the animacy constraint, which is a very strong predictor of genitive choice in the corpus model. I argue that learners struggle mainly with possessor animacy because it is encoded by the semantics of the noun phrase in a given context rather than by any formal cues, which makes it more difficult to learn its statistical association with the choice of genitive variant.

Keywords: genitive alternation; proficiency level; SLA; probabilistic grammar; rating task experiment

References

- Engel, A., Grafmiller, J., Rosseel, L., & Szmrecsanyi, B. (2022). Assessing the complexity of lectal competence: The register-specificity of the dative alternation after *give*. *Cognitive Linguistics*, 33(4), 727–766.
- Gablasova, D., Brezina, V., & McEnery, T. (2019). The Trinity Lancaster Corpus: Development, description and application. *International Journal of Learner Corpus Research*, 5(2), 126–158.
- Rosenbach, A. (2014). English genitive variation—The state of the art. *English Language and Linguistics*, 18(2), 215–262.