Statistical Models of Genitive Acquisition in Child Language

Linh Hoang
Friedrich Schiller University of Jena, Germany

Genitive variation is among the best extensively researched syntactic alternations in English linguistics (Rosenbach, 2014). They are two near-synonymous constructions containing two noun phrases linked by an *'s* inflection (i.e. *s*-genitive) or a preposition *of* (i.e. *of*-genitive) as exemplified in Example 1. Both constructions consist of a possessor and a possessum placed in different positions. While possessor precedes possessum in the *s*-genitive, it follows possessum in the *of*-genitive.

(1) a. *s*-genitive: the girl's name [possessor's possessum]
b. *of*-genitive: the name of the girl [possessum of possessor]

Findings from genitive alternation studies have shown that speakers count on multiple factors such as animacy, definiteness, syntactic weight, final sibilancy, thematicity and semantic relation when deciding between two constructions (Gries & Stefanowitsch, 2004; Hinrichs & Szmrecsanyi, 2007). These phonological, semantic and pragmatic variables interact simultaneously to create a context that motivates a corresponding genitive realization. Even though genitive alternation is exhaustively researched in adult language, it has not received much attention in language acquisition. Therefore, this thesis aims to investigate the production of genitive variation in child language. The research questions are whether children are influenced by the same factors as adults when choosing between genitive variants and whether these factors operate simultaneously. We compare the language of the children (i.e. the output) with that of their caretakers (i.e. the input) as a way to explore the effect of linguistic experience on language acquisition.

To address these questions, we collect data samples of genitive variants produced by children and their caretakers from the Child Language Data Exchange System corpus (MacWhinney, 2010). We first conduct a lexical survey on the data using a frequency list and a collostruction analysis to detect the specific kind of noun phrases frequently used in and statistically attracted to each genitive construction in child and adult language. Next, we perform a logistic regression model with mixed-effects that tests the effects of all the factors simultaneously on the genitive production in child and child-directed speech. A conjoined model with mixed data from both groups of speakers is also fitted to compare the predictor effects on their language and explore whether their genitive production varies significantly. Last, we employ methods of conditional inference trees and random forests to verify the results from the regression models and to measure and rank the relative importance of the predictors on genitive choices of children and their caretakers.

The findings show that children consider the same lexical, phonological, semantic and pragmatic factors as adults when choosing a genitive construction. These predictors interact and operate at the same time in conditioning their genitive choice. Among five examined factors (possessor animacy, definiteness, sibilancy, semantic relation, and length difference between possessor and possessum), children are affected by possessor animacy, definiteness and length in the same way as their caretakers. The difference in child and child-directed speech lies in their sensitivity to the effects of possessor sibilancy and semantic relation. Children show a lower level of impact because their vocabulary knowledge is not at a ready stage to facilitate these factors. Concerning their limited topics of conversation and lexical resources, there is a shortage of noun phrases that contain a final sibilant or trigger a non-prototypical semantic relation. Evidently, children can only mirror adult speech to the best of their lexical ability. However, the results support the usage-based perspective that linguistic experience plays an important role in children's language acquisition. Children use the available frequency information from adult speech to recognize and acquire probabilistic patterns of the collexemes that are statistically attracted to a construction. This confirms the effect of frequency from linguistic exposure on children's acquisition of syntactic variation, or more generally, a probabilistic grammar (Diessel, 2007; Bresnan, 2007; De Marneffe, Grimm, Arnon, Kirby, & Bresnan, 2012).

References

Bresnan, J. (2007). Is syntactic knowledge probabilistic? experiments with the english dative alternation. *Roots: Linguistics in search of its evidential base*, *96*, 77–96.

De Marneffe, M.-C., Grimm, S., Arnon, I., Kirby, S., & Bresnan, J. (2012). A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, *27*(1), 25–61.

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New ideas in psychology*, *25*(2), 108–127.

Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based perspective on alternations. *International Journal of Corpus Linguistics*, *9*(1), 97–129.

Hinrichs, L., & Szmrecsanyi, B. (2007). Recent changes in the function and frequency of standard english genitive constructions: A multivariate analysis of tagged corpora. *English Language and Linguistics*, *11*(3), 437.

MacWhinney, B. (2010). The childes project: Tools for analyzing talk (Third ed.) [Computer software manual]. Mahwah, NJ: Lawrence Erlbaum Associates.

Rosenbach, A. (2014). English genitive variation - the state of the art. *English Language and Linguistics*, *18*(2), 215-262. doi: 10.1017/S1360674314000021