Finnish has a class of embedded clauses where the object systematically varies between nominative and genitive, as in (1) (Itkonen 1976, 1981).

(1) Pekka luul-i [mets-ää ole-va-n \\
Pekka.NOM think-PAST [forest-INE be-ACT.PRES-GEN \\
karhu ~ karhu-n]. \\
bear.?NOM ~ GEN]

‘Pekka thought there to be a bear in the forest’

(Anttila and Kim, 2017)

Itkonen’s (1976) production data from 126 participants showed that the variation reflects diachronic change: genitive is giving way to nominative in a manner predictable by embedded clause type (existential vs. predicative clause) and matrix voice (active vs. passive). Crossing these variables produces a predictable pattern of variation, where existentials and actives condition higher retention of the genitive than predicatives and passives.

To assess the current situation, we conducted a web-based replication of Itkonen’s original study. We found that the proportion of nominatives in all four environments is now significantly higher than in Itkonen’s (1976) experiment, suggesting the GEN > NOM change has progressed. Strikingly, our participants produced case proportions similar in their hierarchical distribution to those of Itkonen’s original 1976 participants (Fig. 1). A generalized logistic mixed effect model shows the expected significant effects of embedded clause type and matrix voice: passives ($\beta = 1.99, p < .001$) and predicatives ($\beta = 3.24, p < .001$) constructions favors nominative; participant and item were included as random effects. Of the social factors (age, education, region, gender), only age shows a significant effect: older participants disfavor the nominative, as expected.

Key problems requiring an explanation are the following: (i) How to predict the probabilistic variation itself? (ii) Why is the direction of the change GEN > NOM? (iii) Why have the quantitative hierarchical relationships remained constant through the change? (iv) Why is the change taking place in certain embedded clauses, but not in, e.g., simplex clauses? Minimalist accounts that don’t engage with probabilistic data struggle with (i); Anttila and Kim’s (2017) OT analysis explains (iii) and (iv), but fails to explain (ii); structural-ambiguity-driven analogy (Ylinärä & Frascarelli 2021) cannot explain (iii). Accounts that appeal to phonological leveling and markedness reduction (Timberlake 1977, Kulikov 2006) and the relative cognitive cost of different linguistic structures (Vogels et al., 2014, Goudbeek & Krahmer 2011) may shed light on (ii), and we suggest that such accounts are a promising avenue of inquiry to better understand how, as our results show, syntactic change in progress is constrained by language structure. Finally, our results demonstrate the importance of follow-up investigations in the realm of linguistic change.
Figure 1.

References


Kulikov, L. (2006). Case systems in a diachronic perspective. Case, valency and transitivity,
