Pronoun omission in Estonian dialects: dialectal differences, language contacts, and functional motivations

Liina Lindström, Maarja-Liisa Pilvik, Helen Plado, Kristel Uiboaed (University of Tartu)

The subject pronoun drop is a feature of many languages, including Estonian. From the historical perspective, the pronoun usage has been secondary in Uralic languages; person endings, which are developed from pronouns (Janhunen 1982), have been the primary means of person reference. Thus, using both person ending and a pronoun is, historically, a doubling of the same element.

In contemporary spoken Estonian, pronouns occur with 82% of sg1 verb forms (Duvallon & Chalvin 2004). A previous study on a subset of dialects uncovered remarkable dialectal variation in the expression of 1sg pronouns (Lindström et al. 2009). In Estonian and Finnish, there are also many other factors influencing the choice between overt expression or omission of the pronoun, such as referential distance or type of the verb (Lindström et al. 2009; Helasvuo 2014; Väänänen 2015; Väänänen 2016).

In our study, we look at the choice between expressing or omitting the pronoun with 1sg verb forms (e.g. (ma) maga-n ‘I sleep’), based on spoken dialect data obtained from the Corpus of Estonian Dialects (CED). There are shy of 11 000 occurrences of 1st person affirmative verb forms in the final dataset.

Based on simple frequency data, the differences between Estonian dialects are substantial (overt pronoun expression is between 24-71%), which can be explained partly by language contacts. We employ nonparametric regression and ensemble methods for obtaining the individual and complex effects of several explanatory variables: referential distance (Givón 1983), verb type (verbs of communication, verbs of cognition, other), form of the previous reference to the same referent, tense, occurrence of the sg1 ending -n, dialect, and subdialect.

Results show that the primary factors influencing the pronoun omission are dialect area and referential distance: the closer the last mention of the same referent, the more likely the pronoun is omitted from the clause. We also confirmed that dialectal differences are remarkable and are at least to some degree in correlation with Germanic influence on dialects. However, as long-term language contacts are hard to detect, the contact-induced influence can be estimated only roughly, on a very general level. The effect of the occurrence of the person ending -n on the pronoun omission, however, remained unsubstantial.


