Pronoun omission in Estonian dialects: contact-induced changes vs. functional motivations

Liina Lindström
University of Tartu
Maarja-Liisa Pilvik
University of Tartu
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 (developed from pronouns (Janhunen 1982)) have been 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 82% of sg1 verb forms (Duvallon & Chalvin 2004). A previous study on subset of dialects discovered remarkable dialectal variation in the expression of 1sg pronouns (Lindström et al. 2009).   In Estonian and Finnish, there also many other factors influencing the choice between overt expression or omission of the pronoun (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 the spoken dialect data obtained from the CED.

By simple frequency data, differences between Estonian dialects are substantial (overt pronoun expression is between 24-71%), which partly can be explained 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 (cognitive verb vs. other), tense, occurrence of the sg1 ending *-n*, polarity, dialect, and subdialect.

Results show that referential distance is the primary factor influencing the pronoun omission: the closer the last mention of the same referent, the more likely the pronoun is dropped from the clause. The occurrence or omission of the sg1 ending *–n* has not much effect. In other words, the speakers of Estonian dialects in general do not compensate the omission of the sg1 ending with a corresponding pronoun.   Thus, the person reference in Estonian dialects depends strongly on the context.

CED = Corpus of Estonian dialects. *http://www.murre.ut.ee/estonian-dialect-corpus/*