

Corpus-based dialect syntax: variation and frequency of partitive subjects in Estonian dialects

LIINA LINDSTRÖM
University of Tartu

The paper tackles the use of a syntactic phenomenon characteristic to Estonian and other Balto-Finnic languages – subjects of intransitive clauses case-marked with the partitive – in Estonian dialects (ex. 1).

(1) *nohh ol'' säärtsit vannu meehi* (Võru dialect)
PART be-PST.3SG such.PL.PRT old. PL.PRT man. PL.PRT
'there were such old men'

The use of partitive subject (PS) is restricted mainly to existential and possessive clauses (Erelt & Metslang 2006) where PSs alternate with nominative subjects. However, this alternation is less systematic than in the closely related Finnish. The constraints regarding the choice between nominative and partitive subject marking in Estonian are not clear, although some efforts have been made to explain it in Standard Estonian (Metslang 2012).

The paper focuses on the distribution of PSs in Estonian dialects. Estonian dialects are traditionally classified into North-Estonian and South-Estonian dialect groups. However, when syntactic features are taken into account, the East-West distinction seems to be more relevant than the traditional North-South distinction. The main research questions concern whether there are differences between dialects in the use of PS; whether these differences follow the East-West distribution as expected; whether the differences appear in frequency data (as many syntactic features do, see Szmrecsanyi 2013), and whether there are any substantial differences in conditions and constraints on using PS in dialects.

The data is obtained from the Corpus of Estonian Dialects, which consists of spoken dialect interviews recorded mainly in the 1960s-1970s. Morphologically annotated texts have undergone automatic syntactic parsing. All occurrences of PSs have been automatically selected and checked manually during the coding of different morphosyntactic and semantic factors. On the basis of the data, statistical analysis has been conducted. As a result, we show that differences between dialects are easier to distinguish when we include various linguistic factors, rather than simple frequency data.