

## Frequencies of verbs and verbal constructions across Estonian dialects and standard language

Frequency effects on different levels of language have recently become the focus in many different studies (Divjak & Gries 2012; Gries & Divjak 2012; Bybee 2007; Bybee & Hopper 2001 among many others). Frequency has been stated as an important factor when observing the variation of language (Bod *et al.* 2003, p.3) and some authors suggest that frequency is a better indicator of variational patterns than traditionally used dialect atlases (Szmrecsanyi to appear).

Considering the small size of the area where Estonian is spoken, the differences between traditional dialects are remarkable. The present study focuses on the frequencies of verbs and two different kinds of verbal constructions in Estonian dialects, standard spoken and written Estonian. Previous studies indicate substantial variation of verbal constructions across dialects and these variational patterns do not follow the traditional dialect classification patterns (Author *et al.* 2013; Author 2010). Some previous studies suggest that certain dialects prefer verb + verb complement constructions (ex. 1) opposed to others that show preferential tendencies to the same verb and action nominal constructions (ex. 2) in the similar functions (Neetar 1986).

- |     |   |           |                |
|-----|---|-----------|----------------|
| (1) | Põrand  | taha-b    | pest-a.        |
|     | floor   | want-3.SG | wash-INF       |
|     | ‘Floor needs to be washed ( <i>lit.</i> wants to be washed).’ |           |                |
| (2) | Põrand  | taha-b    | pesemis-t.     |
|     | floor   | want-3.SG | washing-SG.PRT |
|     | ‘Floor needs to be washed ( <i>lit.</i> wants washing).’      |           |                |

In our presentation we first observe the overall frequencies of verbs in Estonian dialects and we hypothesize that these frequencies vary considerably across dialects. We expect the eastern dialects to use less verbs than western dialects and we expect these results to correlate with previous observations (Author *et al.* 2013). We also include verb and verbal complement/action nominal constructions into the analysis and compare these results in all the dialects. According to the previous observations clear variational patterns should emerge here. We compare all the dialectal findings with standard spoken and written Estonian and expect the standard language to be more similar to Mid and/or Western dialects, as standard language is based on the Mid dialect. Potential differences might indicate, on the one hand, different language influences (Germanic influence in West and Russian and Latvian influence in East and South), on the other hand, they reveal auxiliatation of certain verbs in verb+verb constructions. More generally, dialectal differences in verb frequencies may show the possible regional sources of changes that have taken place primarily in some dialect areas and later become a part of modern standard Estonian.

The data is obtained from the Corpus of Estonian Dialects (CED) which contains the morphologically annotated data from all the ten dialects of Estonian language. The standard spoken data is obtained from the Corpus of Spoken Estonian (Hennoste 2000) and from the Corpus of Written Estonian (Kaalep *et al.* 2000).

## References

- Author. *et al.* 2013. Variation of verbal constructions in Estonian dialects. *Literary and Linguistic Computing*. Available at: <http://llc.oxfordjournals.org/content/early/2013/01/10/llc.fqs053.abstract>.
- Author. 2010. Ühendverbid eesti murretes. *Keel ja Kirjandus*, 1, pp.17–36.
- Bod, R., Hay, J. & Jannedy, S. 2003. Introduction. In R. Bod, J. Hay, & S. Jannedy, eds. *Probabilistic Linguistics*. MIT Press, pp. 1–10.
- Bybee, J. & Hopper, P. 2001. *Frequency and the emergence of linguistic structure*. Amsterdam; Philadelphia: Benjamins.
- Bybee, J.L. 2007. *Frequency of use and the organization of language*. Oxford: Oxford University Press.
- CED. Corpus of Estonian Dialects. Available at: <http://www.meertens.knaw.nl/edisyn/searchengine/>.
- Divjak, D. & Gries, S.T. 2012. *Frequency effects in language representation*. Berlin: De Gruyter Mouton.
- Gries, S.T. & Divjak, D. 2012. *Frequency effects in language learning and processing*. Berlin: De Gruyter Mouton.
- Hennoste, T. 2000. Suulise eesti keele uurimine: Transkriptsioon, taust ja korpus. *Keel ja Kirjandus*, 43(2), pp.91–106.
- Habicht, K., Kaalep, H.-J., Muischnek, K., Müürisep, K., Rääbis, A. 2000. Kas tegelik tekst allub eesti keele morfoloogilistele kirjeldustele?. *Keel ja Kirjandus*, 43(9), pp.623–633.
- Neetar, H., 1986. Mõnest teonimetarindist eesti murretes. *Emakeele Seltsi Aastaraamat*, 32, pp. 36–45.
- Szmrecsanyi, B. to appear. Forests, trees, corpora, and dialect grammars. In B. Szmrecsanyi & B. Wälchli, eds. *Aggregating dialectology, typology, and register analysis: Linguistic variation in text and speech*. Berlin: Mouton De Gruyter.