Diachronic corpus-based study of Estonian semantic cases: methodological basis and possible methods

People tend to prefer one theoretical approach to other, but more productive trend at least in corpus linguistics should be to combination of quantitative and qualitative approaches to get results that are multidimensional. Douglas Biber (2004) goes deeper and says that multi-dimensional analysis comprises different statistical methods: factor analysis and cluster analysis.

Diachronic approach explains the patterns of language-usage. Theoretical background and optimal methods which form the methodological plan depend on the question and the aim of the research. It is brought out different methods, for example linguistic cluster analysis (see Trainis & Allkiri 2014, also Ots 2012), morphoanalysis, to make a usage-based analysis of Estonian semantic cases. Next to the description and explanation of methods it is shown, how these methods make up a methodological basis of the study. Among other points it is enlarged on the problems and possibility of mixing methods of different fields.

Methodological background is illustrated with the pilot analysis of the corpuses of Estonian belletristic language from the second half of the 19th (TÜ1) and 20th (TÜ2) century, which shows how native speakers prefer to express their thoughts in writing. In the usage of (semantic) cases it can be seen difference between language usage in second half of 19th century and in second half 20th century.

Constructions are taken from usage and may be analysed as a trend in the language usage: the usage-based analysis of language has a practical output.

Significant topics connected to the study are homonymy, syncretism and reduction (Grünthal 2003), frequency and preferences of the language users (Eslon 2009; 2013), grammar dependence of the vocabulary, idioms in the language (Muischnek, 2005). Which of the factors named above have the decisive role in the usage and structure of Estonian semantic cases? Diachronic corpus-based analysis should give to answer to the question.

References

Biber, D. (2004). Conversation text types: A multi-dimensional analysis. *7es Journées internationales d'Analyse statistique des Données Textuelles*, 15-34. http://lexicometrica.univ-paris3.fr/jadt/jadt2004/pdf/JADT_000.pdf (13.01.2014).

Eslon, P. (2009). Eestikeelses tekstiloomes eelistatud konstruktsioonid ja käändevormid. Korpusuuringute metodoloogia ja märgendamise probleemid. *Tallinna Ülikooli eesti keele ja kultuuri instituudi toimetised 11*. Tallinn, 30–53.

Eslon, P. (2013). Kahe keelekasutusvariandi võrdlus: morfoloogilised klassid ja klastrid. *Lähivõrdlusi/Lähivertailuja 23*, 13-38.

Grünthal, R. (2003). *Finnic adpositions and cases in change*. Helsinki: Suomalais-Ugrilaisen Seuran Toimituksia 244.

Muischnek, K. (2005). Eesti keele verbikesksed püsiühendid tekstikorpuses. *Emakeele Seltsi aastaraamat 51*, 80-106.

Ots, S. (2012). *Statistikapõhise tarkvara loomine morfoloogiliste kollokatsioonide eraldamiseks eesti keele tekstidest*. Juhendaja Erika Matsak. Bakalaureusetöö. Tallinn: Tallinna Ülikooli informaatika instituut.

Trainis, J. & Allkivi, K. (2014). Ilukirjanduskeelest uue pilguga. *Estonian Papers in Applied Linguistics 10*, 283-306. doi:10.5128/ERYa10.18.

TÜ1 = Belletristic language corpus of 2nd half of 19th century. http://www.cl.ut.ee/korpused/baaskorpus/1890/ (13.01.2014).

TÜ2 = Belletristic language corpus of 2nd half of 20th century. http://lepo.it.da.ut.ee/~heli_u/SA/ (13.01.2014).