Methodological issues regarding a corpus-based analysis for the development of the CEFR Grammar Profiles for Estonian

Jelena Kallas

Institute of the Estonian Language

In the talk, we will introduce the Estonian Profile (the Estonian Vocabulary Profile and the Estonian Grammar Profile), which is designed to support the CEFR illustrative descriptors scales of linguistic competence with language-specific description. We use corpus research techniques as our principal method for investigating how learners acquire grammar. Different types of corpora were used for the development and validation of the Estonian Profile, including the L2 Estonian coursebook corpora Estonian as a Second Language Coursebook Content Corpus 2017, Estonian as a Second Language School Coursebook Content Corpus 2021, and the L2 Estonian learner language corpora Estonian learners' language corpus EMMA and the TLU interlanguage corpus. For the analysis, we used the Corpus Query Systems KORP, EMMA and Sketch Engine. We will demonstrate the main functionalities of CQSs used for grammar profile development and propose possible future research to facilitate a corpus-based analysis for the development of the CEFR Grammar Profiles.

Keywords

learners' corpora; textbook corpora; CEFR proficiency levels; lexicographic resources, Estonian