In our poster, we studied errors at the different proficiency levels assessed according to the Common European Framework of Reference for Languages (CEFR) (2001). Although error analysis has long traditions in learner language and language development studies and is a debated method (see e.g. Ellis & Barkhuizen 2005), error identification, classification and explaining are still essential in the learner corpus research (LCR). Recent learner corpus studies on English have proven that error analysis can reveal accurate information on the development of language proficiency (see Thewissen 2015). In the present study, proficiency levels are assessed using CEFR definitions, which have rarely been used in LCR, and the focus is on morphologically rich language which brings challenges to error analysis.

In the analysis, we concentrate mainly on the differences and developmental patterns at CEFR levels A2 – B2. Primary comparison focuses on the main error categories defined in Brunni, Lehto, Jantunen & Airaksinen (2015), but also some significant subcategories are discussed if they are frequent and essential in describing the differences between levels.

The data comes from the International Corpus of Learner Finnish (ICLFI), the size of which is million. At the moment, the error-annotated sub-corpus of ICLFI contains 1.182 texts, with a total of 184,000 tokens. The error classification system of ICLFI is hierarchical, and covers different components of language. The main error categories are orthographic, phonological, morphophonological, morphological, morphosyntactic, syntactic, lexical, phraseological and unexplainable classes, some containing several subcategories. (Brunni, et al. 2015.)


Ellis, Rod & Gary Barkhuizen 2005. Analysing learner language.


Keywords: error annotation, learner corpus research, accuracy, proficiency levels.