Abstract

The diploma thesis is concerned with the analysis of recurrent word-combinations in the speech of advanced Czech speakers of English and native speakers of English. The data used for the analysis is extracted from two corpora, learner corpus LINDSEI and native speaker corpus LOCNEC. The aim of the thesis is to compare the two groups of speakers, determine differences in their use of recurrent word-combinations and compare the findings to previous studies involving speakers of different languages. The quantitative analysis is performed on a sample of 50 speakers from each corpus and the frequency data is used to compare the two groups as to the number of types of word-combinations they use and how frequently they do so. The qualitative analysis is performed on a sample of 15 speakers from each corpus to determine functional differences. Four categories of word-combinations are determined in the analysis. In the conclusion, the quantitative and qualitative findings are compared to previous research involving speakers of different languages.

Keywords: spoken language, learner language, n-grams, n-gram analysis, recurrent word-combinations, lexical bundles, learner corpus