Abstract

The thesis analyses lexical complexity in the written production of L2 proficient learners of English (the highest C2 level). Additionally, it compares L2 lexical complexity with L1 lexical complexity of English native speakers. This lexical complexity is investigated in two key parameters: lexical diversity and lexical sophistication. A quantitative analysis is made by the means of single indicators and is followed by an analysis where composite indicators VOCD-D and MTLD are employed to measure lexical diversity. Lexical sophistication is explored through the English Vocabulary Profile tool (EVP), which categorises words in a text according to predefined frequency word lists (A1-C2 types). The main hypothesis presumed that the lexical complexity of L2 English speakers is inferior to L1 English speakers, despite the fact that their L2 English language competence is at the highest level possible (C2 level), often compared to "native-speaker competence". It was expected that the results in respective groups (L2 and L1 speakers) would be similar. Another working hypothesis is that low-frequency words at the C2 level will be smaller for L2 English speakers than that for L1 English speakers.

The data comprises 20 comparable texts of L2 proficient English speakers in the dimension of their length (the mean is 781 words), genre (economy), aim (to write a commentary based on a published article) and structure (recommended structure of the commentary), and 20 texts of L1 English speakers, again with very similar parameters. Firstly, the data was analysed independently in their respective groups (L2 and L1) and afterwards, between these two distinctive groups. For these analyses, an automated website software Text Inspector was applied. Its detailed description can be found in the thesis.

The results showed that lexical diversity inside each group (L2 and L1) is similar. On the other hand, if compared, lexical diversity was substantially higher for L1 English speakers measured by VOCD-D and MTLD indicators. Another interesting result is that L2 English speakers tend to repeat the same words more than L1 English speakers, thus not using synonymy to a greater extent. In the case of lexical sophistication, the hypothesis has not been proven as the results were statistically insignificant. The word distribution into different types (A1-C2) reached similar results with the decreasing number of words from A1 to C2 in both groups. A large number of unlisted types was the result of the fact that the EVP database does not include specific economic lexis and that economic texts incorporate more numerals, proper names and abbreviations (mainly acronyms) than other general texts. It would be advisable to

increase the data sample in future research to achieve better reliability of results. It could be achieved via the greater number of texts or the longer text length. The research showed that there is a genre sensitivity and for this reason, it would be recommended to keep the same genre in follow-up research. Another strong influence on the results is a so-called “priming effect” (the tendency of students to repeat words from the chosen article in their commentaries), which automated tools like the Text Inspector can not measure.

Keywords

Lexical complexity, L2 language, written language, lexical frequency wordlist, English Vocabulary Profile (EVP), Text Inspector