Patterns of Lexis in Learner Language: Lithuanian Learners vs. Native Speakers

Keywords: learner corpus, vocabulary acquisition, lexical patterns

Corpus linguistics approach to the study of learner language continues to give many new insights into different aspects of vocabulary acquisition. This study, which is a part of my PhD research into lexical differences between L1 and L2 written English, has also been inspired by corpus studies. More specifically, the presentation will discuss those lexical differences which could be established by analysing multi-word units, i.e. lexical bundles, understood as "most frequent recurring sequences of words" (Biber 2006: 133). This research draws on corpus studies of English (Altenberg 1998, Biber et al. 1999, Biber et al. 2004, Biber 2006, Hyland 2008) and learner corpus research (Granger 1998b, De Cock 2004). The present study is expected to contribute to the field in two aspects. Firstly, to the best of my knowledge corpus-driven approach has been hardly ever used to the study of vocabulary acquisition of Lithuanian EFL learners which only became possible with the compilation of LICLE, a Lithuanian component of the ICLE project, launched by S. Granger (Grigaliūnienė et al. 2008). Secondly, an attempt is made here to analyze learner language in terms of Pattern Grammar (Hunston and Francis 2000) which views lexis and grammar as inseparable aspects of language welding into one unit different aspects of language competence. This study has been designed as Contrastive Interlanguage Analysis (Granger 1998a) to compare lexical patterns established in written English produced by native and non-native speakers of the language. The data has been retrieved from three corpora of student essays, namely, corpora AFK1 and LICLE both of which represent written English of Lithuanian university students at two different levels of proficiency and LOCNESS, a corpus of native speaker English compiled at the University of Louvain (Granger 1998a). The method involves both automated and manual analysis. The WordSmith Tools program has been used to retrieve lexical bundles of three, four, five and six words. The lists of the bundles have been manually edited to eliminate sequences of non-authentic language, e.g. quotations, fiction titles etc., and to conflate identical bundles of varying lengths, e.g. end of the and the end of the. The shortened lists have been used to describe the prevailing patterns of lexis following in as much as possible patterns presented in Hunston and Francis (2000). Preliminary findings show that in terms of different pattern types, the analysed corpora bear similarities, which is perhaps the outcome of similarity in language type, namely written argumentative essays. For example, noun patterns, e.g. a N of as in a problem of, a matter of are the most numerous in all the corpora. Yet the language of native speakers is much more lexically varied than NNS language which is evidenced by different realizations of individual patterns. Another early observation is related to the completeness of patterns in the bundle lists. The NS corpus yields a much larger proportion of complete patterns whereas the NNS corpora contain significantly fewer complete-pattern bundles which possibly indicates a transitional stage on the continuum of vocabulary acquisition.

References


