

Abstract

The research reported in this thesis explores the degree of authenticity of the formulaic language used by NNSs and the extent to which a learner's L1 interferes in the production of different types of multi-word units, namely non-idiomatic recurrent three and four-word combinations (lexical bundles), phrasal and prepositional verbs and collocation. Drawing on Granger's Contrastive Interlanguage analysis (CIA 1996), the investigation is conducted on two different learner sample corpora and subsequently contrasted with a native sample corpus. The study aims to prove that multi-word units pose a challenge for learners for several reasons. In general terms, learners are assumed to operate predominantly on what Sinclair calls the open-choice principle, that is to say their production will be less idiomatic than that of native speakers'. This assumption is independently tested on different types of phraseological combinations. As regards non-idiomatic recurrent word combinations, learners are expected to be more repetitive in their three- and four-word combinations and use less creativity in their writing. Concerning the phrasal verbs, it is highly likely to observe a small number of phrasal verbs in the non-native writing whereas prepositional verbs are considered problematic for learners due to the choice of the appropriate preposition. In the chapter devoted to collocations, a weak sense of collocation salience is expected in the non-native samples. The study is performed in the light of corpus-based approach: different types of evidence (based on the BNC, the PIE, existing dictionaries and native speakers' introspection) were used in the evaluation of the findings.