The Emergence of Formulaic Language in EFL Writing – A Cross-linguistic and Cross-sectional Perspective

Agnieszka LEŃKO-SZYMAŃSKA
Institute of Applied Linguistics
Warsaw University
a.lenko@uw.edu.pl

New approaches to second language acquisition underline the importance of formulaic sequences in L2 learning. It is frequently maintained that phraseology is one of the hardest aspects of foreign language learning and poses a problem even to advanced L2 users. For the most part, collocations are problematic in language production rather than language comprehension since a large number of formulaic sequences are semantically transparent. It is even claimed that apart from a faulty pronunciation, incorrect use of collocations is the most salient marker of non-nativeness in advanced learners' output.

Most of the studies on the use of phraseology by second language learners concentrate on advanced L2 users (e.g. Nesselhauf, 2003; Leńko-Szymańska, 2008). Researchers attempt to tease out to what extent learners' phraseology is different from the native one. There are almost no accounts of formulaic language emerging at the early stages of learning, particularly in foreign language settings, when learners are not faced with the demands of real communication in L2. The existing studies also neglect the whole range of other learner variables such as age and L1. This gap was already observed by Wray (2002: 144):

> "Here, research has often tended to assume that the 'learner' label overrides all others, so that individuals who would easily be acknowledged as different in aspects of their L1 behaviour and, indeed, different in all other respects of their L2 learning, suddenly become a homogeneous group when it comes to formulaicity."

The research reported in this paper attempts to bridge this gap. It is exploratory in nature and investigates the emergence and use of formulaic language by a range of students learning English in the classroom setting. The data is analyzed with reference to learners' ages, stages of proficiency, and L1 backgrounds.

The data analyzed in the study was drawn from the ICCI corpus. The International Corpus of Crosslinguistic Interlanguage is a collection of essays written in L2 English by primary and secondary school students from grade 4 to grade 12. The analyzed essays were collected in Austria, Poland, Spain, Israel and Taiwan. All the five subcorpora were included in the analysis, yet a decision was made to focus on the data at intervals larger than one grade, ideally essays collected from students at important turning points in their education. However, since the educational system in each of these countries is different, an arbitrary decision was made to analyze essays from grades 6, 9 and 12.

The compositions were processed with the help of Collocate (Barlow 2004), a software tool extracting phraseological items from texts. The retrieved items were compared against the list of frequent English n-grams. Native speakers of each language were also consulted to establish if the equivalents of the retrieved phrases exist in learners' L1s.

The analyses demonstrated that the number of formulas retrieved from the five subcorpora increased with the years of learning. This observation is hardly surprising considering the fact that the length of produced essays increased with age and proficiency level. At each stage the list of retrieved sequences contained both native-like and non-native like units. They were analyzed in terms of their structure and function. An attempt was also

made to establish as to what extent they could have had their source in learners' first language.

References

Barlow, M. (2004). *Collocate 1.0: Locating collocations and terminology*. Houston, TX: Athelstan.

Leńko-Szymańska A. (2008). 'Formulaic sequences in apprentice writing – Does more mean better?' In *Proceedings of the 8th Teaching and Language Corpora*. Lisbon: AEIC-ISLA. 207- 212.

Nesselhauf, N. (2003). 'The use of collocations by advanced learners of English and some implications for teaching.' *Applied Linguistics*: 223-242.

Wray, A. (2001). *Formulaic Language and the Lexicon.* Cambridge: Cambridge University Press.