

Abs-178

Antonio Pinna and David Brett

Variability patterns in formulaic language in spoken academic English

Biber (2009) investigated variability within multi-word units using two corpora: a 4.5 M word corpus of American English conversation; and a 5.3 M word corpus of academic prose. Initially, the corpora were searched for 4-grams, discarding sequences with a frequency of less than 10 occurrences per million words. Each corpus was then searched for a series of sequences composed of three of the components of each 4-gram, allowing variability in the fourth slot, e.g. * 2 3 4, 1 * 3 4 etc. If the token in a given slot in each 4-gram composed less than 50% of the results for that slot, it was deemed to be variable, and marked with an asterisk. This procedure permitted the identification of typical patterns of variability in the formulaic sequences across the two corpora. For example, internal variability in one slot (1 * 3 4/1 2 * 4) was seen to be relatively common in Academic Prose, whereas initial and final variability (* 2 3 *) was more frequent in the conversation data.

This paper will illustrate the results of a study that adopted a similar corpus-driven procedure to investigate variability within fixed sequences in the sum of two corpora, which, albeit composed of spoken language, could be reasonably expected to display some features of written language: BASE and MICASE.

The results obtained display some aspects which bear remarkable similarities to those obtained by Biber. For example, the * * * * sequence, in which variability is high in all slots, is the least frequent both in our work and in Biber's spoken and written categories. Rather surprisingly, however, in terms of variability in formulaic sequences, spoken academic language appears to bear little resemblance to academic prose. To the contrary, our results display trends that closely mirror those of Biber's conversation data, in that the predominant patterns consist of variability in the external slots (1 2 3 *, * 2 3 4 and * 2 3 *). Patterns of variability in the second or third slot, which were seen to be particularly characteristic of academic prose, are not a consistent feature of the spoken variety.

Differences in the distribution of the patterns across the BrE and AmE dimension, as well as that of academic macrofield, will also be discussed.

Reference

Biber, D. (2009) "A corpus-driven approach to formulaic language in English" *International Journal of Corpus Linguistics*, 14:3, 275-311.