

## Is there a Formula for Formulaic Language?

**Richard S. Forsyth (16 Halliday Avenue, Leeds);  
Łukasz Grabowski (Opole University, Poland).**

### **Abstract:**

Ideally, anyone who claims that a text exhibits "formulaic language" (Wray & Perkins, 2000) should be able to give the formula, i.e. to specify a grammar that generates that text. Such an ideal is far from realization. Nevertheless, just as it is possible to decide that a language belongs to the Indo-European family without producing a complete grammar and lexicon of proto-Indo-European, it is possible to detect traces of formulaic language without full specification of the presumed "formula" that generated it. This paper focuses on ways of detecting such traces at the phraseological level (Granger & Paquot, 2008). In particular, we test a number of formulae for quantifying the degree to which a text type incorporates inflexible sequences. We argue that such indices should be relatively insensitive to the raw frequencies of the sequences in question and assess a number of candidate indices against that criterion, using a pharmaceutical corpus of over 2 million words divided into four text types.

For this purpose we adopt the concept of "phrase-frame" defined by Fletcher (2007) as a set of variants of an n-gram identical except for one word. Phrase-frames offer a means of comparing pattern variability across different text types (Roemer 2009). So far, however, there has been a dearth of studies explicitly addressing the problem of measuring pattern variability of phrase-frames, with the exception of Roemer (2010), who proposed the variant-to-phrase-frame (VPR) ratio. We examine VPR along with several other indices, including the Herfindahl-Hirschman index, Simpson's diversity index and Shannon entropy (Upton & Cook, 2006), and report which of these indices of micro-productivity (in individual phrase-frames) best meet our criterion. We also show how such indices may be combined to give an indication of macro-productivity (in text registers), i.e. an inverse measure of a register's reliance on formulaic subsequences.

### **References**

- Fletcher, W. 2007. KfNgram. Annapolis, MD: USNA.
- Granger, S. & Paquot, M. 2008. „Disentangling the phraseological web”. In S. Granger & F. Meunier (Eds.), *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins, 27-50.
- Roemer, U. 2009. English in Academia: Does Nativeness Matter? *Anglistik: International Journal of English Studies*, 20 (2), 89-100.
- Roemer, U. 2010. Establishing the phraseological profile of a text type. The construction of meaning in academic book reviews. *English Text Construction*, 3 (1), 95-119.
- Upton, G. & Cook, I. (2006). *Oxford Dictionary of Statistics*. Oxford: Oxford University Press.
- Wray, A. & Perkins, M.R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20, 1-28.