

Breaking the ice: Studying formulaic language in high and mid-level frequency bands.

Elsa Morgan-Pientka, Swansea University

A major issue in vocabulary acquisition has risen with Nation's (2006) work on mid-range vocabulary (2000-9000 words). Traditional frequency levels defined low frequency as '2,000+ academic vocabulary' in Nation's programme and high frequency as typically the most frequent 2,000 making up approximately 99% of average texts (Schonell, Meddleton and Shaw, 1956). Martinez and Schmitt (2012: 302) as well as Nation and Waring (1997: 18) have determined that frequency is the best indicator of usefulness of formulaic sequences. Moreover, Boers and Lindstromberg (2009) argue that the most frequently occurring multi-word expressions are more likely to be assimilated and so 'mid-frequency' formulaic sequences could be particularly relevant in L2 pedagogy. However, there is no research looking into updating the General Service List to fully include mid-level vocabulary and formulaic sequences despite the recent emergence of two amended versions of the General Service List- the New General Service List or NGSL by Browne et al (2013) and the New-General Service List or New-GSL by Brezina and Gablasova (2013). Nevertheless, Martinez and Schmitt (2012: 302) have concluded that current wordlists possess a "key deficiency" in being restricted to individual words despite the general consensus that formulaic language is ubiquitous in language use. In point of fact, the creation of any list that excludes formulaic language has a wider impact than is accepted.

As it has been noted (Nation, 1997) that the increase of technological power provides the ultimate helping hand to any linguist. Corpus analysis is now much easier to study word frequency and range as well as the prevalence of formulaic sequences. In light of this many linguists either do not or cannot fully utilise the available power by relying on existing programs instead of taking up the reins so to speak. This means that a large amount of data gathered has been altered or interpreted by the designer of the program. For the purposes of this presentation, a program was designed or "scratch-built" to construct a totally comprehensive look at unadulterated data of the most frequently occurring formulaic sequences in the British National Corpus (BNC).

While the PHRASE list developed by Martinez and Schmitt has been calculated to the top 5000 words from the BNC. However the PHRASE list has been selected to show only the most transparent meaning of the formulaic expressions included, whereas the expressions gathered with the program specifically designed for the task of identifying multiword units shows the top 10,000 most frequent expressions without any removal or interpretation of data allowing a more comprehensive view of formulaic expressions in the BNC. The information gathered is presented in Type [phrases that occur], Occurrences [number of occurrences] and the Rank [most occurrences] of each formulaic sequence in the corpus.

Boers, F. and S. Lindstromberg. 2009. *Optimizing a Lexical Approach to Instructed Second Language Acquisition*. Basingstoke: Palgrave Macmillan.

Brezina, V. & Gablasova, D. (2013). Is There a Core General Vocabulary? Introducing the New General Service List. Oxford University Press: *Applied Linguistics* pp. 1-23.

Browne, C. et al. (2013). A New General Service List: the most important words for second language learners of English. Available from: <http://www.newgeneralservicelist.org/> [Accessed on:03/02/2014].

Martinez, R. & Schmitt, N. (2012). A Phrasal Expressions List, *Applied Linguistics* 33/3: pp. 299-320.

Nation, I.S.P. (1997), "Vocabulary size, text coverage, and word lists", in Schmitt; McCarthy, *Vocabulary: Description, Acquisition and Pedagogy*, Cambridge: Cambridge University Press, pp. 6–19.

Nation, P. (2006). How large a vocabulary is needed for reading and listening? In M. Horst & T. Cobb (Eds.), *Second special vocabulary edition of Canadian Modern Language Review*, 63(1), pp. 1-12 .

Nation, I. S. P. and R. Waring. (1997). 'Vocabulary size, text coverage and word lists' in N. Schmitt and M. McCarthy (eds): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge University Press, pp. 6–19.

Schonell, F.J., I.G. Meddleton and B.A. Shaw. 1956. A Study of the Oral Vocabulary of Adults. Brisbane: University of Queensland Press.