Determining the Lexical-grammatical Specificity of Text, Genres and Specialised Corpora through the Systematic Analysis of ‘Concgrams’ (B-Q11N)

Martin Warren (PI), Winnie Cheng and Chris Greaves

31 December 2008 - 30 December 2011

Background

Underpinning this research project examining phraseology is the highly innovative ConcGram software, written by senior research fellow Chris Greaves (2009), which is able to fully and automatically uncover all of the phraseology in a text or a corpus of texts, irrespective of variation.

What is a concgram?

ConcGram’s outputs, termed ‘concgrams’, best illustrate the benefits to be gained from studying the variation that characterises much of the phraseology of the English language. Figure 1 illustrates how the two-word concgram ‘expenditure/government’ consists of both constituency and positional variation, but retains its core meaning.

Examples of ‘constituency variation’ are when other words fall between the core words in the phrase such as ‘government’s own expenditure’ and ‘government has made an effort to sustain expenditure’. ‘Positional variation’ is when the sequencing of the words in the phrase changes as in examples such as ‘government expenditure’ and ‘expenditure patterns of the government’.

Types of phraseology uncovered by concgramming

The research team is exploring three main kinds of phraseology (Cheng 2009; Cheng, Greaves and Warren, 2006; Cheng, Greaves, Sinclair and Warren, 2009; Greaves and Warren, 2007) which are exemplified on the right.

1 Meaning shift units

These consist of lexically-rich phraseologies.

2 Collocational frameworks

These are grammatical words which are co-selected to frame meaning shift units.

3 Organisational frameworks

These are co-selections of words which function to organise interconnected sections of the discourse.

References


All the examples of concgrams presented here are from a five-million word sample of the British National Corpus (three-million written and two-million spoken).