Abstract Submission for *Evaluation in Academic Discourse*

**Name:** David Oakey

**Affiliation:** University of Birmingham, UK

**Address:**
EISU (Westmere)  
The University of Birmingham  
Birmingham B15 2TT  
United Kingdom

**Tel:** +44 (0)121 414 5697

**Fax:** +44 (0)121 414 3600

**Email:** D.J.Oakey@bham.ac.uk

**Title:** Academic Vocabulary in Academic Discourse: the phraseological behaviour of EVALUATION in the discourse of Economics.

**Keywords:** Written Academic Discourse; Phraseology; Vocabulary; Specialised Corpora

**Abstract (991 words)**

Recent corpus-based approaches to vocabulary in English for Specific Purposes (Coxhead 1998; 2000; Coxhead and Nation 2001) have been based on distinctions between “general English” vocabulary, “academic” vocabulary, and subject-specific “terminology”. Academic, or sub-technical, vocabulary is not specific enough in meaning to belong to the terminology of a specific discipline, but at the same time is rather more formal than “general” English. The Academic Word List (AWL) (Coxhead 1998) is an attempt to specify the most frequent, non-discipline specific academic vocabulary for learners of English. It consists of 570 “families” of single orthographic words, principally chosen on the basis of the “coverage” of academic texts which these words provide. If a learner who knows the top 2000 general English words also learns the members of the 570 word families from the AWL, the “coverage” of academic text increases from 78.1%, where roughly one word in five is unknown to the learner, to 86.6%, where merely one word in ten is unknown. (Nation 2001: 17). The AWL has become a benchmark for designers of materials for teaching academic English (e.g. Cobb 2002; Campbell and Thompson 2003; Haywood 2003; O’Regan 2003; Nelson 2003).

However, since this notion of “coverage” applies only to single orthographic words, it seems worthwhile to question how this general academic vocabulary behaves in academic discourse. Corpus informed analyses of language have long questioned the links between meaning, phraseology, and the single orthographic word form. Sinclair, for example, contends that it is now possible to compare the usage patterns of, for example, all the forms of a verb, and from this to conclude that they are often very different one from another ... there is a good case for arguing that each distinct form is potentially a unique lexical unit, and that forms should only be conflated into lemmas when their environments show a certain amount and type of similarity.  (1991: 8)

Applying this to the AWL, it is not clear that the collocational and colligational environments of members of the same word family need be similar across each inflected and derived word form.
Thus the members of the word family for EVALUATION, as presented in the sublist 2 of the AWL (p17),

- evaluate
- evaluated
- evaluates
- evaluating
- evaluation
- evaluations
- evaluative
- re-evaluate
- re-evaluated
- re-evaluates
- re-evaluating
- re-evaluation

although presented as members of the same word family, may combine in widely dissimilar ways, and a learner who attempts to learn these words in isolation may find that they have little in common other than their spelling.

The open choice and idiom principle (Sinclair 1991) similarly highlights the idea that a text in part consists of “a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.” (ibid. 1991: 110). Some of these segments, presumably, could be single orthographic words, and thus the AWL may break up more stable word combinations. In another corpus-based study of academic discourse, Biber et al. (1999: 990) describe “lexical bundles”. These are frequently recurring strings of three or more words which “commonly go together in natural discourse” (ibid.) and which make up just over 20% of the corpus. Finally Danielsson (2000) has drawn attention to a little known consequence of the work of Zipf. The relationship between words in a frequency list has a mathematically definable structure

\[ ab^2 = k \]

i.e., the number of words which occur with a particular frequency (a) multiplied by the square of the frequency (b) is a constant (k), regardless of the point in the frequency list. According to Zipf, this relationship holds for approximately 98% of the words in an English frequency list, but fails with the top 2%. Zipf tried using the k constant to predict the frequency of occurrence of the top word in any wordlist, the, with the result that the should only occur 0.000025 times. To Zipf this was “a very absurd statement no matter how a word is defined.” (Zipf 1965 [1935]:43). However Danielsson suggests that “these irregularities at the top end of the frequency list can be explained by the fact that single word units are inadequate as units of analysis in language.” (Danielsson 2000: 55).

Of course, the creators of the AWL never said that learning individual words - orthographic words bounded by spaces - was sufficient for academic proficiency. Learners are intended to add layers of meaning to the meaning they originally associated with this word when they first learned it, according to the context in which they find the word in subsequent encounters with it. But in presenting the benefits of the AWL, a primary distinction is made between the number of known words versus the number of unknown words. Words in the AWL “make the difference between 80% coverage (one unknown word in every five running words) and 90% coverage (one unknown word in every ten running words)” (Coxhead and Nation 2001: 252) The learner’s improved
recognition of individual words, which is assisted by the “coverage” provided by the list, is implicitly equated with improved understanding of meaning.

This paper thus goes on to test a word family from the Academic Word List - in this case in deference to the title of the conference, EVALUATION - against the discourse of Economics. It is apparent that the complexity of the subject means that such single orthographic words do not always carry precise enough meaning “on their own”. The word evaluations in one particular Economics journal article, for example, is never used “on its own”; it is always part of a compound noun, very often in a binomial pair, and usually as part of a postmodifying prepositional phrase. The paper will conclude with a presentation of the phraseological environments of the different members of the EVALUATION word family in a million-word corpus of Economics research articles, and identify their most frequent recurring collocational and colligational patterns.

The conclusion will be that there are many forms of computer analysis available to help us to know the texts we try to teach, both at text level and word level. As a result, the single orthographic word will necessarily remain the primary unit of analysis, but it will not necessarily always be the primary unit of meaning. Therefore, in conclusion, it should not necessarily be the primary unit of vocabulary teaching.

References