Journal of English for Academic Purposes 31 (2018) 44–57

Contents lists available at ScienceDirect

Journal of English for Academic Purposes

journal homepage: www.elsevier.com/locate/jeap

Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement

Ahmed Masrai [a, *], James Milton [b]

[a] Department of Languages, King Abdulaziz Military Academy, Riyadh, Saudi Arabia
[b] Department of English Language and Literature, Swansea University, Swansea, UK

ARTICLE INFO

Article history:
Received 9 August 2016
Received in revised form 20 December 2017
Accepted 22 December 2017

Keywords:
Academic vocabulary
Receptive knowledge
Vocabulary size
Rasch model
Test validity

ABSTRACT

The Academic Word List (AWL) (Coxhead, 2000) is widely used in preparing non-native speakers for academic courses, and it is thought that the words in this list are essential for the understanding of English academic texts (Cobb & Horst, 2004). It is also thought that the AWL is a list of infrequent and specialised words inaccessible from general language. These preconceptions are challenged in the current study. With reference to BNC/COCA word lists, the study demonstrates that the majority of the words from the AWL fall within the 3000 most frequent words, a grouping that Schmitt and Schmitt (2014) describe as highly frequent. Using a specifically created test of the AWL and a test of overall vocabulary size (XK-Lex; Masrai & Milton, 2012), the study demonstrates that the learning of the AWL appears to be strongly influenced by the frequency of these words in general corpora and that the AWL test very strongly resembles a test of overall vocabulary size. Knowledge of the AWL also adds marginally to the power of overall vocabulary size in explaining variance in grade point average (GPA) scores. This conclusion matches that of Townsend, Filippini, Collins, and Biancarosa (2012), although the tests in the current study appear to have greater explanatory power.

1. Background

1.1. The academic word list and academic success

Horst and Cobb (2006), in considering the impact that recent vocabulary acquisition research has had on the English as a foreign language (EFL) world, suggest that most research on the lexicon is considered peripheral to the design and content of EFL programmes. They go on to point out, however, that Coxhead's (2000) AWL is an exception to this. The AWL is a list of the words that are important to communicating the concepts taught in school, university or in English for academic purposes (EAP) programmes. There is support in the academic literature for this claim. Knowing these words has been shown to be essential for L2 learners' comprehension of academic written text (Cobb & Horst, 2004; Dang & Webb, 2014). Gardner and Davies (2014) ascribe academic vocabulary a central role in school success for both native and non-native speakers. Not surprisingly, therefore, the AWL has become central to the teaching of EAP.

* Corresponding author. King Abdulaziz Military Academy, Riyadh 11583, Saudi Arabia.
E-mail address: a.masrai@hotmail.com (A. Masrai).

https://doi.org/10.1016/j.jeap.2017.12.006
1475-1585/

Coxhead's AWL comprises 570 headwords selected from a corpus of 3.5 million running words of academic texts drawn from a wide range of academic genres. The criteria for the creation of the list are that these words should be:

a) *specialised*, so "[t]he word families had to be outside the first 2000 most frequently occurring words of English as represented by West's (1953) GSL" (Coxhead, 2000, p. 221).
b) *generic to academic discourse*, rather than specialist vocabulary items restricted in use to only a limited number of subjects, and a *Range* criterion was used to ensure this was occurred.
c) *frequent*, so items on the list had to occur at least 100 times in the academic corpus.

In principle, therefore, the AWL represents a selection of vocabulary that is not drawn from the most frequent words (Kremmel & Schmitt, forthcoming) but appears to be thought of as something beyond, perhaps well beyond, basic levels of vocabulary knowledge (Coxhead, 2000) and which is unlikely to be accessed through general language exposure (Townsend et al., 2012).

The rationale for the importance of the AWL comes primarily from evidence of the contribution to coverage the list provides. The AWL is generally thought to provide approximately 10% coverage of academic written texts (e.g., Chen & Ge, 2007; Cobb & Horst, 2004; Coxhead, 2000). The list, with the knowledge of the words in West's (1953) General Service List (GSL), gives about a 90% coverage of academic written texts (Nation, 2004). Just how important the AWL is may vary from one subject domain to another. Coxhead's (2000, p. 222) figures suggest that the AWL is more useful in her Commerce sub-corpus than in the Science sub-corpus, for example. In Commerce, the AWL by itself contributed 12% to coverage and, combined with the GSL, comprised 88.8% of the sub-corpus. In Science, the AWL contributed 9.1% to coverage and, with the GSL, comprised 79.8% of the sub-corpus. Slightly different results were obtained by Dang and Webb (2014) who investigated the coverage of the AWL in academic spoken English by analysing the vocabulary in 130 lectures and 39 seminars from four sub-corpora of the British Academic Spoken English (BASE). Their findings suggest that the AWL accounts for 4.41% coverage of academic spoken English, and that its coverage in each sub-corpus varied from 3.82% to 5.21%. They conclude that, with the aid of the AWL and knowledge of proper nouns and marginal words, learners will need a vocabulary size of 3000 and 8000 word families to attain 95% and 98% coverage of academic spoken English, respectively.

Nonetheless, studies appear to support the importance of the AWL in academic texts across a variety of academic fields, such as engineering (Mudraya, 2006), medical research (Chen & Ge, 2007), and applied linguistics (Chung & Nation, 2003). The use of the AWL sub-lists in setting goals for learning is thought to be additionally useful in promoting a significant improvement in learners' overall vocabulary knowledge (Snow, Lawrence, & White, 2009).

1.2. The AWL and general vocabulary

One criticism that is repeatedly levelled at the AWL is the difficulty that exists in distinguishing the words it contains from frequent general vocabulary (Gardner & Davies, 2014). Coxhead's (2000) claim that the words in the AWL are specialised and distinct from general frequent vocabulary falls down when the words in the list are compared with word frequency lists drawn from general language sources. A comparison of the AWL with Kilgarriff's (2006) lemmatised frequency lists and BNC/COCA word family lists (Nation, 2012) is made in Table 1, which summarises the frequency distribution of words in the AWL.

Kilgarriff's (2006) lists are lemmatised and constructed slightly differently from the AWL which, like the BNC/COCA lists is based on word families. Base words in the AWL include morphemic derivations, which in Kilgarriff's lists, may be counted as two or more words. The BNC and BNC/COCA lists are also slightly different in their distribution of words across the 1000 word levels. Despite these differences, it is clear from both just how heavily weighted to the most frequent words the AWL is. In Kilgarriff's lists, 369 AWL words (65% of the total) fall within the first 3000 words. In the BNC/COCA lists, 484 AWL words (84% of the total) fall within the most frequent 3000 words, a grouping that Schmitt and Schmitt (2014) call high frequency. However, there is also a spread of frequency in the AWL, with a small number of items appearing highly infrequently and occurring at or beyond the 8000-word level.

Viewed from this perspective, and taking the AWL as simply a list of words, it is not clear how specialised the majority of the words in the AWL really are. Estimates which link lexical size to hierarchies of performance, such as the Common

Table 1
Frequency distribution of words in the AWL.

Frequency band	BNC (Kilgarriff, 2006)	BNC/COCA (Nation, 2012)
1 k	94	19
2 k	151	136
3 k	124	329
4 k	81	62
5 k	51	15
6 k	35	6
7 k	9	1
Off list	25	2