English for Specific Purposes 39 (2015) 1–11

Contents lists available at ScienceDirect

English for Specific Purposes

journal homepage: http://ees.elsevier.com/esp/default.asp

# A corpus-based environmental academic word list building and its validity test

Jia Liu [1], Lina Han*

*Department of Foreign Languages, Nankai University, Tianjin, China*

ARTICLE INFO

*Article history:*
Available online

*Keywords:*
AWL
Word selection criteria
Validity test

ABSTRACT

This study discusses the necessity of creating field-specific academic word lists and seeks to establish the compilation criteria and the validity test criteria for field-specific academic word lists. These three aspects are explored with reference to an environmental science corpus.

The results indicate that the Academic Word List (AWL) (Coxhead, 2000) is not entirely useful for environmental science learners because of the narrow coverage of some word families and the shortage of frequently used environmental academic words. Therefore, it is necessary to establish a field-specific word list that better reflects specialized features. Referring to Juilland's 'usage' (Juilland & Chang-Rodríguez, 1964), this study proposes 'optimized usage' to determine whether words should be included in an academic word list. It also establishes the first academic word list for environmental science, the Environmental Academic Word List (EAWL). Moreover, the list's validity test criteria are established by way of coverage comparison and hypothesis testing. The EAWL established in this study aims to be maximally useful for environmental science learners.



## 1. Introduction

Academic vocabulary is defined as "formal, context-independent words with a high frequency and/or wide range of occurrence across scientific disciplines, not usually found in basic general English courses" (Farrell, 1990, p. 11). These words are often used to clarify academic viewpoints and to write academic articles. The acquisition of academic vocabulary is an important measure of learners' academic competence (Kuehn, 1996). Therefore, a study examining academic vocabulary is of practical significance.

Over the past decades, researchers focusing on academic vocabulary have produced a number of useful academic word lists, which can be roughly categorized into two types: general academic word lists and field-specific academic word lists.

General academic word lists are concerned with common academic words among different disciplines. Several academic word lists have been compiled over the years (e.g., Campion & Elley, 1971; Gardner & Davies, 2014; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972; Xue & Nation, 1984), and these have played a positive role in teaching and researching academic vocabulary. Today, the most widely used list is Coxhead's Academic Word List (Coxhead, 2000). Coxhead established the Academic Corpus, from which she identified 570 word families to compile the AWL; she found that the AWL coverage in the

* Corresponding author. Room 537 Main Building, Nankai University, 94th Weijin Road, Tianjin 300071, China. Tel.: +86 13920233115 (mobile).
*E-mail address:* hi_linahan@163.com (L. Han).
[1] Tel.: +86 13752319412 (mobile).

http://dx.doi.org/10.1016/j.esp.2015.03.001
0889-4906/

Academic Corpus could reach around 10%. The important role that the list has played in guiding the acquisition of academic English vocabulary has been confirmed by many teachers and researchers (e.g., Huntley, 2005; Li & Qian, 2010).

Field-specific academic word lists (Martínez, Beck, & Panza, 2009) or discipline-based lexical repertoires (Hyland & Tse, 2007) focus on the academic vocabulary closely related to disciplines. Mudraya (2006) reduced the nearly 2,000,000 words in her SEEC (Students Engineering English Corpus) to the 1200 word families encountered in engineering textbooks that are compulsory for all engineering students regardless of their fields of specialization. Wang, Liang, and Ge (2008) compiled a Medical Academic Word List totaling 623 word families; it serves as a word list of the most frequently used medical academic words in medical research papers. Vongpumivitch, Huang, and Chang (2009) and Martínez et al. (2009) separately established an Applied Linguistics Academic Word List containing 603 word families and an Agricultural Academic Word List including only 92 word families; the former was compiled from a corpus obtained from research articles published in five distinguished applied linguistic journals, and the latter from a corpus consisting of 218 research articles taken from the *Science Citation Index* database.

To prove that an academic word list can be effectively applied in academic learning and research, it is necessary to test its validity. For example, Coxhead compiled a second corpus of academic texts in English, using the same criteria and sources to select texts and dividing them into the same four disciplines, and examined AWL's coverage in this corpus. The result shows that the AWL's coverage of the second corpus is 8.5%. Gardner and Davies (2014) have also used a similar technique to validate their new academic vocabulary list (AVL), using the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA). The results indicate that AVL is "a truly 'academic' list" (Gardner & Davies, 2014, p. 18). However, it seems that most researchers focusing on field-specific academic word lists have built no validating corpora; they merely examined the coverage of academic word lists in the original corpora from which the lists are compiled. As a result, the validity of their lists can be questioned, and the lack of validity testing makes it difficult to prove that field-specific academic word lists are more viable to certain fields of specialization than the AWL. Therefore, establishing validity test criteria for academic word lists is of great significance for the further study of academic word lists.

Environmental science is an interdisciplinary field, including both natural and social sciences. Because the vocabulary involves a wide range of fields, acquiring a good command of specialized English is likely to prove challenging for learners. Therefore, an environmental science academic word list will help learners in this endeavor. However, to date no such list has been developed. The present study, taking environmental science as an example, explores word list-establishing and validity test criteria of field-specific academic word lists and establishes the first environmental academic word list.

The following research aspects are to be addressed in this paper:

(1) This study explores the necessity of establishing the Environmental Academic Word List;
(2) It proposes word selection criteria for field-specific academic word lists and compiles the first environmental academic word list; and
(3) It establishes validity test criteria for academic word lists.

## 2. Methodology

### 2.1. Corpus compilation

For the study, we built the environmental science corpus following the criteria of establishing a specialized English corpus (Liu & Han, 2013; Sinclair, 1991, 2005), considering the specialty of the texts, their sizes, representativeness, and dates of publication. The corpus consists of texts from a single genre – research articles – because reading and writing research articles are the fundamental concerns for most students majoring in environmental science. These students are certain to encounter academic vocabulary when writing or reading research articles.

#### 2.1.1. Data collection

All of the research papers in the environmental science corpus were downloaded from the *Science Citation Index* (SCI) database. First, a professor from the environmental science department from our university contributed by classifying the discipline into ten subject areas and separately recommending one highly regarded SCI journal for each area. Next, 20 research articles written in the internationally conventionalized IMRD (Introduction–Method–Result–Discussion) format (Swales, 1990) were randomly chosen from each academic journal. The articles included in the corpus were published between 2010 and 2013, and were kept at their original lengths.

#### 2.1.2. Data processing

The research articles were standardized before being stored in the corpus. To eliminate factors unrelated to the lexical analysis, charts, diagrams, numbers, appendices, bibliographies, equations, and other textual components that could not be processed by computer software programs or were not to be included in the lexical analysis were removed. Finally, these texts were classified by subject area.

The environmental science corpus contains 200 texts, 862,242 tokens, and 26,612 types. Information relating to its ten subject areas is illustrated in Table 1.