

# The Application of Corpus Tools in the Teaching of Discipline-Specific Academic Vocabulary: A Case Study for Information Engineering Undergraduates

Min Zhang, Faculty of Foreign Studies, Beijing Information Science and Technology University, Beijing, China

---

## ABSTRACT

*This article concerns a corpus-based lexical study that is aimed at teaching academic vocabulary to address the specific needs of Chinese undergraduates majoring in information engineering. A 1,024,882-word corpus of Information Engineering English Corpus (IEEC) was built on the basis of university-level textbook materials drawn from ten compulsory courses for information engineering students. A quantitative analysis was carried out to seek for an optimal frequency threshold for extracting frequently occurring academic vocabulary specific to the discipline of information engineering based on Coxhead's Academic Word List (2000). An innovative 2-dimensional categorization of AWL for EAP students was adopted in the study. As a result, a 100-word highly frequent core AWL and a 147-word frequent AWL were compiled for IEEC under two parameters of lexical frequency and specificity to individual sub-corpora. The present study further explored the application of corpus tools to highlight and effectively teach the discipline-specific academic vocabulary and collocations to promote learners' autonomy and enhance their lexical competence in the study of specialized courses.*

*Keywords:* Academic Vocabulary, Corpus, Information Engineering, Information Engineering English Corpus (IEEC), Vocabulary Teaching

---

## INTRODUCTION

The rapid development of information and communication technology (ICT) in the past two decades has greatly transformed not only the way people communicate but also the way people learn to communicate, especially in foreign languages. For learners of any foreign language to develop any communicative

competency, vocabulary learning is a crucial component and can hardly occur naturally like native speakers. "Not surprisingly, the development of various kinds of word lists has been a recognized contribution to the facilitation of vocabulary learning" (Li & Qian, 2010, p. 402). Without exception, the development of data-driven corpus compilation and statistical tools has also revolutionized the way word lists are generated. Coxhead's Academic Word List (2000) is generally regarded as the first suc-

DOI: 10.4018/ijcallt.2013100104

successful attempt to adopt computerized corpus philosophy and methodology in word list compilation based on the 2000 most frequent word families contained in the General Service List (GSL) (West, 1953) and is designed particularly to serve the need of teachers of EAP classes to “set goals for their students’ vocabulary learning” (Coxhead, 2011, p. 357).

For many students at tertiary level who major in science, technology, engineering and mathematics (STEM) around the world, there has been a growing need for EAP courses and instructions in the past decade. As a major, if not universal scientific language, English has become part of their academic study as an important source of specialized knowledge or a useful tool to get their research findings published, and the awareness is even stronger when English is not their native language (Hyland, 2006; Lam, 2001; Mudraya, 2006; Martínez, Beck, & Panza, 2009; Ward, 2009). Among the efforts EAP teachers have made to address the specific needs of university students, much importance has been attached to learners’ lexical development on the grounds that effective and efficient comprehension would only occur with adequate lexical coverage and the availability of sight vocabulary of a reasonable size (Laufer & Ravenhorst-Kalovski, 2010).

## THE AWL AND RELATED STUDIES ON ACADEMIC VOCABULARY

The Academic Word List (AWL) is a list of 570 word families (head words) which covers 3107 types (individual word forms including the headwords and their inflectional and derivative family members) and is grouped into 10 sublists that are developed by using a 3.5 million-word written academic corpus covering four discipline areas of arts, commerce, law and science (Coxhead, 2011, p. 355). Data from Brown Corpus show that a reader needs to know another 3000 words to raise the coverage of texts from 79.7% (the first 2000 words from GSL) to 88.6% (Nation, 2001, p. 15). The AWL

is ground-breaking in that upon mastering the GSL words, university students can expect to reach comparable text coverage and know about 90% of the running words they will meet in any academic text if they embed another 570 academic words into their lexical repertoire. A number of studies have shown that the coverage of AWL words in academic texts is consistently around 10% in both multi-disciplinary corpus and discipline-specific corpus (Coxhead, 2011). They make it possible for EAP learners to enhance vocabulary recognition and reading abilities notably “for a relatively modest learning investment” (Cobb, n.d.).

In the past decade since Coxhead’s AWL was published, there has been a wide application of the word list in EAP teaching, materials compilation and vocabulary testing (Schmitt & Schmitt, 2005). It has also inspired researchers to develop academic word lists from corpora of academic texts of specific disciplines to address EFL learners’ diversified needs ranging from reading English-written textbooks of a particular field to writing English research articles. For instance, corpus-based studies on the frequency and coverage of AWL have been carried out recently in discipline-specific or genre-specific corpus consisting of data from applied linguistics research papers (Vongpumivitch, Huang & Chang, 2009) and financial texts (Li & Qian, 2010) with AWL coverage of 11.17% and 10.46% respectively. Besides, researchers also explored the same topic in academic corpus of science and engineering disciplines, such as medicine (Chen & Ge, 2007), agriculture (Martínez et al., 2009) and engineering (Ward, 2009) especially since Coxhead’s Academic Corpus does not include an engineering section. The results of these studies rested on the same line with AWL coverage of 10.07% and 9.06% in medical and agricultural research articles and 11.3% in university-level engineering textbooks. In addition to the evidence of the reliable coverage of AWL words in academic texts, these discipline-specific corpora also show certain preferences for word selections and collocation patterns.

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/the-application-of-corpus-tools-in-the-teaching-of-discipline-specific-academic-vocabulary/100371](http://www.igi-global.com/article/the-application-of-corpus-tools-in-the-teaching-of-discipline-specific-academic-vocabulary/100371)

## Related Content

---

### International Students and Their Technology Proficiency

Jacob Manuand Emmanuel Mensah (2015). *Promoting Global Literacy Skills through Technology-Infused Teaching and Learning* (pp. 135-146).

[www.irma-international.org/chapter/international-students-and-their-technology-proficiency/115702](http://www.irma-international.org/chapter/international-students-and-their-technology-proficiency/115702)

### Developing Enjoyable Second Language Learning Software Tools: A Computer Game Paradigm

Chee S. Angand Panayiotis Zaphiris (2006). *User-Centered Computer Aided Language Learning* (pp. 1-21).

[www.irma-international.org/chapter/developing-enjoyable-second-language-learning/30640](http://www.irma-international.org/chapter/developing-enjoyable-second-language-learning/30640)

### Examining EFL Teachers' TPACK Perceptions, Web 2.0 Tools Usage, Workload, and Technostress Levels

adman Hunutluand Sevda Küçük (2022). *International Journal of Computer-Assisted Language Learning and Teaching* (pp. 1-19).

[www.irma-international.org/article/examining-efl-teachers-tpack-perceptions-web-20-tools-usage-workload-and-technostress-levels/315306](http://www.irma-international.org/article/examining-efl-teachers-tpack-perceptions-web-20-tools-usage-workload-and-technostress-levels/315306)

### An Exploratory Study of Using Personal Blogs for L2 Writing in Fully Online Language Courses

Lina Lee (2020). *Recent Developments in Technology-Enhanced and Computer-Assisted Language Learning* (pp. 145-163).

[www.irma-international.org/chapter/an-exploratory-study-of-using-personal-blogs-for-l2-writing-in-fully-online-language-courses/242427](http://www.irma-international.org/chapter/an-exploratory-study-of-using-personal-blogs-for-l2-writing-in-fully-online-language-courses/242427)

### Vocabulary Learning and Consolidation with Mobile Application

Dongshuo Wang, Bin Zouand Minjie Xing (2014). *International Journal of Computer-Assisted Language Learning and Teaching* (pp. 101-112).

[www.irma-international.org/article/vocabulary-learning-and-consolidation-with-mobile-application/109833](http://www.irma-international.org/article/vocabulary-learning-and-consolidation-with-mobile-application/109833)