Coxhead’s AWL across ESP and Asian EFL Journal Research Articles (RAs): A Corpus-Based Lexical Study

Mohammad Bagher Shabania, a*, Khalil Tazikb

aImam Khomeini International University, Qazvin, 34149-16818, Iran
bIlam University, Ilam, 6939177111, Iran

Abstract

In the past few years, several cross-disciplinary corpus-based studies have been carried out on the frequency and coverage of 570 word families from Coxhead's (2000) academic word list (AWL). Some reported high coverage of this word list in their corpora while some others questioned its generality and stated that this word list is far from complete. Hence, along with these studies, the present study attempted to examine the word frequency and text coverage of AWL on 80 research articles (RAs) written in English with 320310 running words across two Asian EFL and ESP journals. Using frequency and range as the criteria for word form selection, this study identified 438 words as the academic words and 144 new added academic words to the list which was called revised academic word list (RAWL). Applying both AWL and RAWL into the entire corpus, results support high coverage and importance of academic words in both ESP and EFL RAs. However, when two word lists were examined into ESP corpus (ESPC) and EFL corpus (EFLC) separately, academic words had higher coverage in ESPC than EFLC. From these findings it is concluded that (1) academic words play an important role in academic texts; therefore, acquisition of them seems to be essential for language learners and users, (2) because of the nature of ESP articles, ESPC holds higher coverage of academic words, (3) some of the words included in the AWL are field-specific and (4) direct attention to these words from behalf of the material and syllabus designers and teachers can lead to a better understanding of these words; hence, students' development in their writing and reading.

© 2014 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).
Selection and peer-review under responsibility of Urmia University, Iran.

Keywords: Coxhead's AWL, corpus-based lexical study, revised academic word list (RAWL), ESP/EFL academic words, reading/writing learners

* Corresponding author.
E-mail address: shabani_m_b@yahoo.com
1. Introduction

Vocabulary as a crucial component of language learning has received great attention by many scholars (see for example, Alderson, 1984; Shaw, Perfetti, & Lauer, 1991; Coany, Magoto, Hubbard, Graney, & Mokhtari, 1993; Nation, 2001). Those researchers who are investigating the students' problems in reading and writing in EFL/ESL contexts perceived that vocabulary is the predominant language problem for both native and non-native students (see for example, Shaw, 1991). Some others such as Read (1998, cited in Wang et al. 2008) and Nagy (1988) claimed that students' vocabulary knowledge directly influences their writing accuracy and reading comprehension. However, selection of those vocabularies worth teaching is not a clear-cut and simple matter. Nation's (2001) division of vocabularies into four levels—high frequency words, academic vocabulary, technical vocabulary, and low frequency words—indicates that some vocabularies need more attention in different phases of language or for different purposes. According to Nation (2001), High-frequency words constitute the majority of running words in all types of writing in different genres and fields of study. That's why these words are the most accessible words for language learners. In contrast, technical words do not occur in all the subjects of study but they may be used in specialized fields. Low-frequency words are the terms which are rarely used. Academic words have plenty of chance for occurring in academic texts from different genres and fields. The difference that academic vocabularies may have with high frequency words is that these words have less occurring frequency in colloquial conversations and speeches. Because of their high frequency in academic texts acquiring these words is essential for language learners who are preparing for English for specific purposes (ESP), English for academic purposes (EAP), English for medical purposes (EMP)… and ENP in general (N stands for any field).

Objectives and significance of the study

This study investigated the frequency of AWL in two ESP and Asian EFL journal research articles (RAs). Therefore, it sought to find answers for the following question:

What AWL word forms occur with high frequency in ESP and Asian EFL journal RAs?

Since the reliability of Coxhead's AWL is under the question and, on the other hand, both teachers and students of English in different levels of study, BA, MA, or PhD as those in other fields of study need a comprehensive list of academic words, results of this study, if they turn out to support the Coxhead's AWL, can give the teachers more confidence in focusing on these words while teaching the AWL. The results of this study can enhance undergraduate students' awareness of academic words and this familiarization help them to write effective essays and supply themselves with a schemata image of frequent academic words in different disciplines and sub-disciplines before involving in such contexts. More importantly, findings of this study can identify the differences, if any, between ESP and non-ESP articles in terms of distribution and frequency of AWL. In the case of observing such differences, ESP teachers should provide a revised AWL for teaching to their students. However, if the results of this study fail to support the coverage of Coxhead's AWL in the fields of applied linguistics and ESP articles, generality of this list as Hyland and Tse (2007) reported will be questioned; hence, it should be revised and completed.

2. Methodology

2.1. The corpus

The selected corpus for this study consisted of 120 research articles (RAs) written in English across two ESP and Asian EFL journals (60 RAs from each journal). These RAs would be downloaded from the http://www.asian-efl-journal.com and http://www.asian-esp-journal.com databases, the Asian’s reputed online journals in ESP and Applied Linguistics including over four thousand articles. Following Nwogu (1997, p. 121, cited in Atai, 2005) reputation is “the esteem which members of an assumed readership held for a particular publication or groups of publications”. After selecting the journals, 80 RAs (40 from each) were randomly selected out of their table of contents as the corpus of this study. Since longer texts give this opportunity to increase the reoccurrence and frequency of words (Coxhead, 1998; Stubbs, 2001) and it may affect the reliability of findings, all the tables, diagrams and bibliographies were removed from the original length of the articles.
2.2. Procedure for data collection
To determine the frequency and range of AWL, all the RAs were copied in Microsoft Word Office 2007. Following Chen's and Ge's (2007) data processing methodology, data processing in this study comprised the normalization, segmentation, and standardization. Normalization includes transformation of words into their basic forms. For example, plural nouns were changed into singular. Segmentation is deriving word family forms. The standardization was exclusion of charts, diagrams, bibliographies, and tables which are not accounted as the parts of lexical analysis.

2.3. Data analysis procedures
The research question aimed to investigate the frequency and distribution of the AWL word forms that are used in two Asian EFL and ESP journal research articles. To answer this research question, the first stage was to obtain a list of AWL word forms that met the criteria set in this study, which included both frequency and range. The frequency criterion was that the word forms have to occur at least 9 times in the entire corpus; the range criterion was that the word forms have to occur at least two times in each journal. The rationale for the selection of these rates of occurrence was based on the Coxhead’s (2000) selection of AWL words. Coxhead’s corpus for the AWL was around 3.5 million words and her selection criteria for the AWL word forms was that each word form in the AWL should occur at least 100 times in the entire corpus and at least 10 times in each of the four disciplines. Because the size of the selected corpus was approximately 320,000 words, which was about 1/11 the size of Coxhead’s corpus, the minimum frequency of occurrence was set to be in agreement with that of Coxhead’s criteria. However, when the Asian EFL and ESP journal articles were analyzed separately, since the corpus was divided into two 160,000 words corpora, the AWL should at least occur 4 times in each corpus.

3. Results
To examine the frequency and distribution of Coxhead's AWL word families, a corpus of 320,310 words from two Asian EFL and ESP journal RAs has been employed. After examining the data, following results have been obtained which are presented below.

**Frequency of AWL word families in the entire corpus**

As it is shown in table 1, AWL account for 14.89% of the entire corpus which is higher than the 11.7% of the AWL reported in Vongpumivitch’s et al. study in Applied Linguistics research articles. This high coverage (14.89%) indicates that AWL plays an important role in RA texts published in two Asian EFL and ESP journals. Frequency analysis indicates that there are 428 AWL word families that occurred at least 10 times in the entire corpus and 3 times in each journal.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Frequency of AWL in Asian EFL/ESP Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asian EFL Journal</td>
</tr>
<tr>
<td>Number of Words</td>
<td>160,110</td>
</tr>
<tr>
<td>Frequency of AWL</td>
<td>19,789</td>
</tr>
<tr>
<td>% of AWL Coverage</td>
<td>12.36%</td>
</tr>
</tbody>
</table>

**Frequency of AWL in two journals separately**

When the analysis is done for each journal separately, results indicate that AWL account for 12.36% of the EFLC and 17.43% for the AESPC (see table 1). The higher coverage of AWL for ESPC indicates that these word families play a crucially important role in ESP articles which is in line with (Coxhead & Nation, 2001). This higher coverage in ESPC can be related to the nature of this field. In contrast, EFL corpus contains a more specific area of research which may lead to application of some limited sorts of vocabularies. Generally, in contrast with other two studies (Chung & Nation, 2003;, Vongpumivitch et al., 2009) that examined the frequency and distribution of Coxhead's AWL in the Applied linguistics journals and books, results of this study show a more coverage.

After counting the AWL in the corpus, about 132 AWs did not receive the criteria for being among the AWL word families; therefore, they were eliminated from the list. The words such as sector, legislate, and commission which are placed among the most frequent words in Coxhead's word list were not found among the AWs in this
corpus. The reason for this can be related to the difference between the selected corpuses for two studies. Although the word legal, for instance, has occurred 68 times in ESP corpus, the score of occurrence for this word in EFL corpus was zero. This low score of occurrence may question the generality of this word list. Therefore, though AWL has high frequency in both EFLC and ESPC, results of this study are somehow in line with Hyland's and Tse's (2007) findings who believed that some of the words in this word list are field specific. Moreover, after deletion of 132 words and analyzing the frequency of AWs, the word list which Coxhead ordered according to their high occurrences needs to be revised and reordered.

Although Coxhead's AWL represented a high degree of words used in the entire corpus in this study, the researcher, after consultations with his instructor and MA classmates, reached this claim that some of the words which are frequently used by researchers in both EFL and ESP RAs have not been included in the Coxhead's AWL. Therefore, lists of all the words that are supposed to be academic words have been prepared. These words, based on the index proposed by Coxhead (2000), have been examined, those which received the criteria were added to the AWL (144 words) and those which did not receive the criteria were omitted from the list. Then, the frequency and distribution of revised AWL (RAWL) were tested in the entire corpus and each corpus separately.

**Frequency and coverage of RAWL in EFLC/ESPC**

As it is indicated in table 3, RAWL accounts for 25.96% of the entire corpus which is higher than Coxhead’ (2000) AWL coverage. In comparison with Vongpumivitch et al. (2009) and Chung and Nation (2003), results obtained from applying RAWL in the present corpus, although cautiously, show that RAWL word families has a higher coverage for Applied linguistics RAs than the Coxhead's (2000) original AWL word families. Comparing cross-disciplinary, results show that RAWL coverage for academic texts is greater than the AWL coverage found by Chen and Ge (2007) in medical RAs (10.073%), by Coxhead and Nation’s study (around 10%) in medicine RAs, Martínez et al. (2009) in agriculture (9.06%), Hyland and Tse (2007) for biology (6.2%) and for computer science (16%).

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Coverage of RAWL in AEFLC/AESPAC</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Asian EFL Journal</td>
</tr>
<tr>
<td>Number of Words</td>
<td>160,110</td>
</tr>
<tr>
<td>Frequency of RAWL</td>
<td>38,634</td>
</tr>
<tr>
<td>% of RAWL Coverage</td>
<td>24.13%</td>
</tr>
</tbody>
</table>

To see if the coverage across two ESP and EFL corpuses is different, RAWL was tested in each corpus separately. Results show that as Coxhead's AWL, the coverage for ESPC is higher than EFLC (27.8% and 24.13%, for ESPC and EFLC respectively), table 2.

As for the individual AWs, after examining two AWL and RAWL, table 3 and 4 demonstrate a list of the top ten most frequently used AWs in the two corpora, with their raw frequencies and their frequencies for per 1000 words. As stated earlier, the total corpus is about 320310 words.

Overall, the top ten most frequently used AWs listed in AWL account for 19.64% (3888/19789× 100) of all the AWs used in the EFLC while the top ten most frequently used AWs listed in RAWL account for 20.83%^{\frac{8050}{38634}} \times 100\) of all the AWs used in EFLC. This shows that both AWL and RAWL contains limited sets of AWs which researchers rely heavily on them. The word strategy is the most frequently used AW in the EFLC according to AWL findings; however, the sixth word in the RAWL ranking. The word learn is the most frequently used AW in EFLC according to RAWL ranking. The high frequency of learn is noticeable because it occurs 1537 times in the entire corpus and 9.59 for per 1000 words. The words strategy, text, task, research, and communicate are the common words into the top most frequently used words into two word list ranks. What is interesting is that the word analyse which Coxhead ranked as the most frequent academic word, is the seventh most frequent word in AWL and absent in the top ten most frequently used words in RAWL.

As it is shown in table 4, the same job has been done for ESPC. The top ten most frequently used AWs in ESPC...
based on two AWL and RAWL word lists along with their raw frequency, percentage of the overall AWs and RAWs, and frequency of them for per 1000 words have been indicated in Table 4. The top ten most frequent words in AWL account for 12.76% of all the AWs of ESPC, while top ten most frequent words in RAWL account for 11.88% of all the AWs used in the ESPC.

Table 3
Top Ten Most Frequently Used AWs and RAWs in Asian EFL Journal

<table>
<thead>
<tr>
<th>Rank</th>
<th>AWs</th>
<th>R.F</th>
<th>%</th>
<th>F.P.W</th>
<th>Rank</th>
<th>RAWs</th>
<th>R.F</th>
<th>%</th>
<th>F.P.W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>strategy</td>
<td>691</td>
<td>3.49</td>
<td>4.31</td>
<td>1.</td>
<td>Learn</td>
<td>1537</td>
<td>3.97</td>
<td>9.59</td>
</tr>
<tr>
<td>2.</td>
<td>Text</td>
<td>522</td>
<td>2.63</td>
<td>3.26</td>
<td>2.</td>
<td>Teach</td>
<td>1397</td>
<td>3.61</td>
<td>8.72</td>
</tr>
<tr>
<td>3.</td>
<td>Task</td>
<td>452</td>
<td>2.28</td>
<td>2.82</td>
<td>3.</td>
<td>Read</td>
<td>955</td>
<td>2.47</td>
<td>5.12</td>
</tr>
<tr>
<td>4.</td>
<td>Research</td>
<td>448</td>
<td>2.26</td>
<td>2.79</td>
<td>4.</td>
<td>Study</td>
<td>821</td>
<td>2.12</td>
<td>5.12</td>
</tr>
<tr>
<td>5.</td>
<td>Communicate</td>
<td>422</td>
<td>2.13</td>
<td>2.63</td>
<td>5.</td>
<td>Test</td>
<td>805</td>
<td>2.08</td>
<td>5.02</td>
</tr>
<tr>
<td>7.</td>
<td>Analyze</td>
<td>291</td>
<td>1.47</td>
<td>1.81</td>
<td>7.</td>
<td>Text</td>
<td>522</td>
<td>1.35</td>
<td>3.26</td>
</tr>
<tr>
<td>8.</td>
<td>Assess</td>
<td>246</td>
<td>1.24</td>
<td>1.53</td>
<td>8.</td>
<td>Task</td>
<td>452</td>
<td>1.16</td>
<td>2.82</td>
</tr>
<tr>
<td>10.</td>
<td>Respond</td>
<td>222</td>
<td>1.12</td>
<td>1.38</td>
<td>10.</td>
<td>Communicate</td>
<td>422</td>
<td>1.09</td>
<td>2.63</td>
</tr>
<tr>
<td></td>
<td>total</td>
<td>3888</td>
<td>19.64</td>
<td></td>
<td></td>
<td>total</td>
<td>8050</td>
<td>20.83</td>
<td></td>
</tr>
</tbody>
</table>

(R.F: raw frequency, %: of the overall AWs and RAWs used, F.P.S: frequency for per 1000 words)

By comparing the results to the EFLC, it can be concluded that the dispersion and variability of AWs in ESPC is higher than EFLC. Academic word text with raw frequency of 740 and 4.61 frequency for per 1000 words is the top most frequent word in AWL ranking and the academic word study with raw frequency of 837 and 5.22 frequency for per 1000 words is the top frequent word in RAWL. Across two word lists, text, analyze, participate, and research are the common words in the top ten most frequently used AWs.

In summary, after analyzing the data obtained from the application of Coxhead's (2000) word list across two EFLC and ESPC, it was revealed that this word list covers high frequency of words used in both corpora. However, this coverage in ESPC was higher than EFLC. Moreover, some new academic words added to the list and some words which did not acquire the criteria have been omitted from the list. The new list was examined into the entire corpus and each corpus separately. Results support the high coverage of new revised academic word list (RAWL) within or across two corpora.

4. Discussion and conclusion
Increased number of research shows that vocabulary learning and teaching is one of the main concerns in different fields of study and contexts (Nation & Waring, 1997; Moirand & Nation, 2002; Nation, 2006; Mudraya,
2006; Ward, 2009). Among all, some researchers such as Coxhead (2000) attempted to define and assemble a comprehensive and general AWL generalizable to all the academic contexts. Following such tendencies, some have tried to examine the generality of provided AWL by Coxhead (Chen & Ge, 2007) and some others have attempted to create a new word list specific for their fields of study (Wang et al., 2008). Along with above mentioned studies, in this study it was attempted to examine the generality of Coxhead's (2000) AWL across two Asian EFL and ESP journals. Moreover, the study aimed to find if there are words in the corpus which can be called academic words but not mentioned in the AWL word families. For this aim, a corpus of 320310 running words of RAs published in Asian EFL/ESP journals from 2004 to 2009 was used. Downloading electronic files of 80 research articles (40 from each journal) from http://www.asian-efl-journal.com and http://www.asian-esp-journal.com, the researcher copied all the articles in the Microsoft Office Word 2007 and tried the AWL in the corpus. As results in Tables 1, 2, 3, and 4 show, both Coxhead's AWL and RAWL, which contain new added academic words to the previous word list, have a high coverage of the entire corpus and two EFLC and ESPC separately. These findings demonstrate that AWL and RAWL play important role in academic writing; therefore, they lend support to Coxhead's (2000) findings. Comparing to two other studies aimed to explore and examine AWL within Applied Linguistics academic texts, findings of this study resulted in a higher coverage of Coxhead's (2000) AWL. The reason for this difference may be one of the limitations of this research. For instance, Vongpumivitch's et al. (2009) study was conducted with four outstanding journals in the area of Applied Linguistics, while this study employed its corpus from two Asian EFL and ESP journals which may lead to less inconsistent results. Findings of this study are also consistent with Hyland and Tse (2007) concerns about the speciality of some of the words in AWL. Many of the words which were supposed to be academic words across all the academic texts and corpora were not occur in the entire corpus of this study, while in Coxhead's list they received high value. Therefore, it can be concluded that as Hyland and Tse (2007) state these words occur in special contexts which ESP teachers should be aware of them.

One of the primary purposes of this study was to help both ESP and writing teachers in their attempts to teach academic vocabularies. As Ellis (1990) states, courses involving direct attention to the teaching materials lead to better understanding. Synthesizing findings of this study with other studies concerned with teaching and learning AWs, be cross-disciplinary or within the area of ESP and EFL, both ESP and EFL teachers can incorporate findings of these studies in their syllabus to effectively teach academic words and also to analyze the degree to which their students use these words in their academic texts. Since coverage of both AWL and RAWL for ESPC was higher than EFLC, it seems that ESP teachers should shoulder a heavier responsibility in terms of familiarizing their students with these academic words.

Findings of this study can also contribute to academic material and syllabus designers in both ESP and EFL areas. They can incorporate teaching of these words in their books and timetables. As Cheng et al. (2003) in their study titled "The language learner as language researcher: putting corpus linguistics on the timetable" stated, our experience has shown that it is possible and worthwhile, even in a fairly short time period, to introduce students to corpus linguistics through combining elements of existing, and we would imagine, standard subjects on an undergraduate English language programme" (p. 183), ESP and EFL syllabus designers can familiarize students with lexical corpus-based studies.

References


