Original Research   **Open Access**

A Corpus-Based Lexical Study of the Frequency, Coverage and Distribution of Academic Vocabulary in Islamic Academic Research Articles

Engku Haliza Engku Ibrahim[*]
International Islamic University, Malaysia

Mohamed Ismail Ahamad Shah
International Islamic University, Malaysia

Mayinuer Abudukeremu
International Islamic University, Malaysia

Abstract

This corpus-based lexical study aimed to explore the use of words in Coxhead (2000) Academic Word List (AWL) in academic journal articles in the field of Islamic studies. Around 472,621 word corpus, called the Islamic Academic Research Articles (IARA) corpus, was created for this study. The corpus consisted of 66 research articles written in English that were published in more than 10 different Islamic academic journals. Authentic and academic research articles written on Islam, and from Islamic perspectives, covering a wide range of topics, were selected. The study found that the most frequent 317 AWL words which occurred in the IARA corpus was only 56% of Coxhead's AWL of 570 words. This finding points to the need for a special AWL for students. Findings suggest the need to produce field-specific academic word lists incorporating all frequent academic lexical items necessary for the expression of the rhetoric of the specific research area. Findings also revealed that some of the words which were found in the present study were not found in Coxhead's Academic Word List. This suggests that vocabulary needs of students in Islamic studies are characteristically different from those of students in other disciplines.

Keywords: English for academic purposes (EAP); Corpus based study; Islamic academic research articles corpus (IARA) corpus, Word family.

1. Introduction

The growth of English as the lingua franca of work and study has increased the demand for English for Occupational Purposes. In the acquisition of English as a foreign language for academic purposes, vocabulary acquisition has been identified as a major factor that impedes effective academic writing (Shaw, 1991) has been seen to contribute the most to academic reading more than any other kinds of linguistic knowledge (Saville-Troike, 1984). Learners of English for academic reading and writing are frequently faced with the daunting task of not knowing which words to focus on for efficient learning of the language.

2. Literature Review
2.1. Previous Studies on Academic Vocabulary

The daunting task of knowing which vocabulary to focus on has been somewhat alleviated by the availability of General Service List (GSL) developed from a corpus of 5 million words. The GSL covers the 2,000 high-frequency words of general English. The first 1,000 covers 77% of running words in an academic text and the second 1,000 covers another 5%. Despite its age, there has been no other comparable list that is as comprehensive. However, the criticism of the GSL is that it does not meet the needs of learners who require English for academic and occupational purposes. This need was seen as partially met by the Academic Word List (AWL) (Coxhead, 2000). Prior to her work many research has attempted to draw up a list of academic words such as that of Campion and Elley (1971), (Praninskas, 1972), and Ghadessy (1979); all of which were not done using computers. These studies were the basis on which the University Word List (UWL) was created (Xue *et al.,* 1984). The UWL contains 800 words that are not in the first 2,000 words of the GSL. However, according to Coxhead, these studies were based on small corpora and did not have a wide and balanced coverage of topics; this led to the development of the Academic Word List (AWL) (Coxhead, 2000). The AWL contains 570 word families but provides more text coverage and has a more consistent criteria for word selection (Jing Wang *et al.,* 2008)

The significance of academic vocabulary in its contribution to general text comprehension cannot be denied. Coxhead reports that the AWL covers 10% of the tokens in her corpus of 3.5 million running words of academic texts but only 1.4% of the tokens in her collection of fiction collection. Along the same lines, (Santos, 2004) also shows that academic words amounted to 16% of the words in his text sample. These percentages highlighted are significant following (Laufer, 1998) research findings which points to the fact that in order for a learner to have a reasonable comprehension of the text, there has to be less than 5% of unknown words. In other words, 95% of the

words in the texts need to be known by the learner to enable him to moderately guess unknown words encountered during reading.

The GSL combined with the AWL is said to cover 90% of a learner's target vocabulary. This has been a considerable contribution to providing focus for vocabulary teaching, course development and materials design. Thus far, the AWL has been accepted as the authority in the field of academic vocabulary. It covers the disciplines of Arts, Commerce, Law and Science. In her study, Coxhead (2000) noted that there was a difference in the occurrence of words across these disciplines. The highest was that of Commerce as it covered 12% of the commerce sub-corpora, followed by arts and law (9.3% and 9.4%) and 9.1% from science. The AWL also covers a wide range of subject areas. This has been the point of departure where many claim that the AWL fails to address the needs of learners who have to acquire English for a specific discipline in the most efficient manner.

Several shortcomings of the AWL have been highlighted. Hyland and Tse (2007) bring attention to the differences in distribution of percentages of the GSL combined with the AWL across different disciplines. In (Coxhead, 2000) corpus, the GSL combined with AWL covered 86.1 % of her corpus when compared to 84.7% of Hyland and Tse (2007). The percentage decreases in Hyland and Tse (2007) science sub-corpus; a case in point that there is still a gap to be filled in order for a learner to have complete comprehension of a discipline-specific text. A second problem identified with the AWL concerns word meaning and use. Wang *et al.* (2004) brought attention to homographs in the list that could affect the criteria of frequency for inclusion in the AWL list. In addition, Hyland and Tse (2007) noted that some of the words in the AWL behave different semantically in different disciplines. In other words, meanings are specific to their respective discipline, thus cannot be assumed to only have one meaning that covers all the different disciplines. Along the same lines, (Fuentes, 2001) conducted a study that showed a distinct difference between technical and academic word behavior in the field of Information Science Technology. Additionally, Lam (2001) reported that learners had problems in comprehending academic vocabulary in technical texts.

This realization gave rise to numerous studies that focus on single disciplines which attempt to draw up a list in addition to the AWL to include technical words specific to the various disciplines. Martinez *et al.* (2009) conducted a study of academic vocabulary in research articles in the field of agriculture from a corpus of 826, 416 words derived from 218 journal articles in the field of Agriculture. Their research revealed that only 9.06% of the AWL words are covered in their corpus. These studies conclude that some words carry specific meanings and behaviours in their genre and thus warrant an additional list to the AWL. Their studies also revealed some words that are in the AWL carry technical meanings in their corpus. Mudraya (2006) established a corpus for engineering students. (Vongpumivitch *et al.,* 2009a) carried out a frequency analysis of words in Applied Linguistics research Papers and found 475 AWL word forms and 128 non-AWL word forms in their corpus of 1.5 million words. Similarly, Chen (2007) found that the AWL does not have the coverage nor dispersion of medical academic vocabulary in the medical research articles included in their corpus of 190 425 running words. The assumption underlying these studies is that there are special features of the vocabularies of these single disciplines that have prevented students from the various disciplines from complete comprehension of the texts they are reading.

This present study attempts to extend the AWL to include words that are frequently found in Human Science Islamic Journal articles, in particular vocabulary that are used for Islamic references.

The objectives of this study were to:

1. find out the frequency, coverage and distribution of Coxhead's AWL words in the Islamic academic research articles (IARA) corpus according to the word types, tokens and word families.
2. investigate on the existence of non–AWL frequent content words and other specific lexical items that function as academic words in the present study.

The findings of this research is significant to aid teachers in materials development and putting sequence in teaching.

3. Methodology

This research analyzed frequent AWL lexical items as well as other non-AWL lexical items which are common in research articles published in Islamic academic journals in order to determine what impact the number of high frequency words found could have on the future of how word lists are compiled and used for pedagogical purposes. The methodology used stresses the need to use actual authentic data and takes 'whole text' as the rationale of lexical analysis.

Information was collected primarily from two sources that are the International Islamic University Malaysia (IIUM) publications and Non-IIUM publications. The main reason for choosing articles from different sources was to ensure language variety in the lexical items. Thus, articles that were general in context and published in the last thirty years (from 1975 onwards), were selected in the present study. To ensure that the collected data are research articles and originate from various sources, a collection of approximately 900 articles and essays published by more than 10 peer–reviewed Islamic academic journals was used as a population. Of the 900 articles, 66 were selected, which makes it a sampling rate of 7.3%. From this population, research articles were randomly selected. As the study reports on frequent AWL words in a particular genre, that is, the research article, book reviews, conference reports, forums and essays were excluded.

In addition, to ensure a representative sample of language varieties, the researcher used *Stratified Random Sampling* (SRS), which first involves dividing the whole populations into relatively homogeneous subgroups (so-called strata) and then taking samples of each stratum at random (McEnery *et al.,* 2006). In this research the articles were stratified according to whether they are IIUM or non-IIUM articles and also according to the types of journals

and the subject areas articles from IIUM publications as well as foreign publications from English-speaking countries. In this research the articles were stratified according to whether they are IIUM publication or non-IIUM which is foreign publications from English-speaking countries and also according to the pure Islamic subjects and multi- disciplines special reference to Islam.

An additional criterion for the selection of articles was that it had to be medium length articles of approximately 2000 words. The basic sampling principle in the compilation of IARA corpus McEnery *et al.* (2006) and Chen (2007) was to randomly select not only the titles from the bibliographical sources but also the particular subject of a text using a random-number table. This sampling principle was dictated by practical considerations, such as the availability of material, or whenever a single text did not yield the required 2,000 words. To avoid language bias from an author, articles were chosen from different journals, edited by different editors and written by different authors. It was decided that only one article was to be chosen from an author even though there were many good articles written by the same authors.

The articles selected from these academic journals were on Islamic related subjects and they were multi-disciplinary in nature from the various fields of history, geography, political science, economics, anthropology, sociology, law, literature, religion, philosophy, international relations, medical, environmental and developmental issues. Articles chosen were stored electronically, if they were available in the electronic form, while the others were scanned into PDF versions. Then, each article was transferred into Microsoft Word and saved in a plain text format (txt). After having all the articles in plain text format, they were compiled into a computer file and entered into the Word Smith Tools (WST) 5.0 hh scott 2009 programme software for analysis.

As the focus of this research is on words from the AWL, and considering that studies on academic vocabulary have mostly used word families as the unit of analysis, the researcher identified the families of the most frequent words in the current corpus. This is because word families are considered a significant measurement unit in concepts of the learning burden (wNagy *et al., 1989*). They are also considered an important unit in lexical studies West (1953); Xue *et al.* (1984) and Coxhead (2000).

For the purpose of this paper, the units of analysis are word families, tokens, and types. The concept of a word family is used to classify a group of words so that its meanings can be understood when the meaning of the base form in the group is familiar to a learner. The basic idea here is that there is a core or central form and meaning from which certain derived forms with their meanings are closely related. Thus, comprehending regularly inflected or derived members of a word family does not require much burden, if learners have the knowledge of base words and basic word building processes (Bauer and Nation, 1993). For instance, the headword *accumulate* and its other family members are *accumulated*, *accumulating*, *accumulation* and *accumulates*. From the example *accumulate*, it can be seen that a word family consists of a headword, its inflected forms and its closely related derived forms, even if the part of speech is not the same (Coxhead, 2000). Additionally, to ensure a rigorous word selection, the researcher followed (Hyland and Tse, 2007) criterion, which considered as frequent those words that, occurred above the mean of the total number of academic words. The mean for full text was identified and this built the families of those items that were above the mean.

Types, on the other hand, are defined as single word forms. We can count the words in the sentence 'It is very difficult to read it quickly' in another way. If we see the same word again, we do not count it again. Thus the sentence of eight tokens consists of seven different words or 'types' as the word 'it' occurred twice (Coxhead and Nation, 2001). Tokens are defined as the number of occurrences of each 'type'. One way is simply to count every word form in spoken or written text, and if the same word form occurs more than once, then each occurrence is counted. Thus the sentence 'It is very difficult to read it quickly' would contain eight words, even though two of them are the same word form, *it.* Words which are determined in this way are called 'tokens' or sometimes 'running words' (Coxhead and Nation, 2001).

4. Findings and Discussion

The first objective of the research was to find out the frequency, coverage and distribution of Coxhead's AWL words in the Islamic academic research articles (IARA) corpus according to the word types, tokens and word families.

4.1 Word Frequencies in the Islamic Academic Research Articles (Iara) Corpus

The overall distribution of tokens and types in this corpus is shown in Table 4.1. The items were sub-grouped into general words (GSL), academic words (AWL) and "other". These included mainly technical words, but also other words, such as formulas, adjectives, loan words, proper names and field specific words.

Table-4.1. Coverage of lexical items in the IARA Corpus

Items	AWL	GSL	Other	Total
Tokens	38496	274827	159298	472621
Percentage (%)	8.15	58.15	33.7	100
Types	2219	1900	18036	22155
Percentage (%)	10.02	8.58	81.4	100

4.1.1. Token

The overall distribution of tokens in the present corpus is presented in the Table 4.1. The GSL (58.15%), the AWL (8.15%), and others (33.7 %) provided a cumulative coverage of 100% for the IARA corpus. The combined GSL and the AWL covered only 66.3% for the whole corpus.

4.1.2. Type

As for types, the total number used in the corpus was 22,155. Their distribution across vocabulary areas was different, indicating lower variability in the GSL and higher in the other section as shown in Table 4.1. There are 3,107 types in Coxhead's AWL (Martinez et al., 2009). The sub-group of AWL word types revealed that, of the total of 3,107 types in the AWL, only 2,219 occurred in the IARA corpus, which means that 888 items of the AWL (28.6%) did not occur at all in the corpus of the present study. However, AWL occurrence in this corpus was higher compared to GSL word types that only had 1,900 word types in the corpus. The GSL (8.58%) and the AWL (10.02%) provided a cumulative coverage of 18.6% for the whole corpus. The rest was 81.4% to hold the highest percentage in the corpus.

The cumulative percentage of types is shown in Table 4.2. There were two types in the AWL that occurred more than three hundred times, *tradition* and *research*, and there were thirteen that occurred between 200 and 300 times, *process, authority, physical, legal, approach, philosophy, concept, theory, economic, rational, revelation, role,* and *context*. The high-frequency of these words or members of word families reflects the important role of academic vocabulary in Islamic academic texts. These fifteen types represent 0.68% of the AWL's word types in the frequency list. On the other hand, according to Table 4.2, the last 917 types had between one and four occurrences and represented 41.33% of the AWL's word types used. The first fifteen types, however, accounted for 3,652 tokens, that is, a total of 9.49% of the AWL words in the corpus, while the last 917 types accounted for only 1994 tokens, that is, 5.18% of the tokens used.

Table-4.2. Frequency of academic word tokens and types used in the Islamic academic research article corpus

| Tokens	Types	Percentage types	Cumulative types	Cumulative percentage types
≥ 300	2	0.09	2	0.09
200-299	13	0.59	15	0.68
100-199	49	2.21	64	2.89
50-99	123	5.54	187	8.43
21-49	289	13.02	476	21.45
10-20	383	17.26	859	38.71
5-9	443	19.96	1302	58.67
4	164	7.39	1466	66.06
3	178	8.02	1644	74.08
2	229	10.32	1873	84.4
1	346	15.59	2219	100

4.2. Frequency of Word Families

This section sought to determine the most frequently used AWL in Islamic academic writing, which is the main objective of the research. This aim was achieved by calculating the frequencies of all word-forms of each AWL word family in this corpus of Islamic research articles. As described in Section 3.1, word families in this corpus were established by following (Hyland and Tse, 2007) frequency criterion, hence, the researcher first identified the mean for academic words which was 17. This was the criterion that was used in developing the frequent word families of those items. Words occurring fewer than 17 times were omitted from the list (Table 4.3).

**Table-4.3.** Frequency and distribution of academic word families above the mean in IARA corpus

Academic word families	With head words		Without head words		Total	Percentage
	numbers	%	numbers	%	numbers	%
1 member family	110	45.27	57	78.08	167	52.85
2 member families	84	34.57	9	12.33	93	29.43
3 member families	24	9.88	6	8.22	30	9.49
4 member families	14	5.76	0	0	14	4.43
5 member families	8	3.29	1	1.37	9	2.85
6 member families	3	1.23	0	0	3	0.95
Total	244	100	73	100	317	100

Figure-4.1. Distribution of academic word families above the mean in the IARA corpus
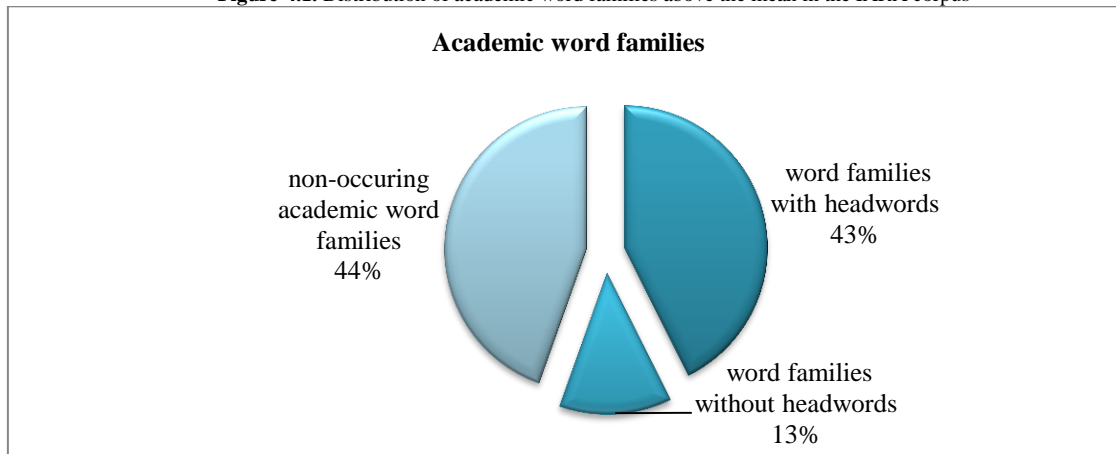
The resulting list contain 317 word families arranged in the sequence of those occurring most frequently (310 times) to those occurring the least frequently (17 times). As shown in Table 4.3 and Figure 4.1, among the 570 AWL word families, only 244 word families with AWL headwords (43%), and 73 word families without headwords (13%) appeared more than 17 times in the corpus, which were considered as frequently used academic words. Thus, the total number of word families above the mean for this corpus was 317 of Coxhead's AWL. It means that only 56% of the AWL word families occurred in this corpus, whereas the other 253 word families (44%) did not appear at all, which were presumed not frequently used academic words in Islamic academic research articles. This occurrence list was much higher than Hyland and Tse (2007) and Martinez *et al.* (2009) results, who found 192 and 92 frequent families respectively, while much lower than (Vongpumivitch et al., 2009b) study around 475 (83% of the AWL) in their overall corpus. The result of this study on the distribution of the AWL word families in IARA corpus is 317 (56%), consistent with the AWL distribution in academic writing reported in Chen and Ge's study around 292 (51.2%) (see Figure 4.2). It is interesting to note that the corpus size in the IARA corpus, (66), and the corpus in (Chen, 2007) study, (50), are quite close. This closeness in the corpus size in these two studies could have contributed to the similarities in the AWL distributions in both the corpuses.

Figure-4.2. Comparing distributions of academic word families with other *studies*
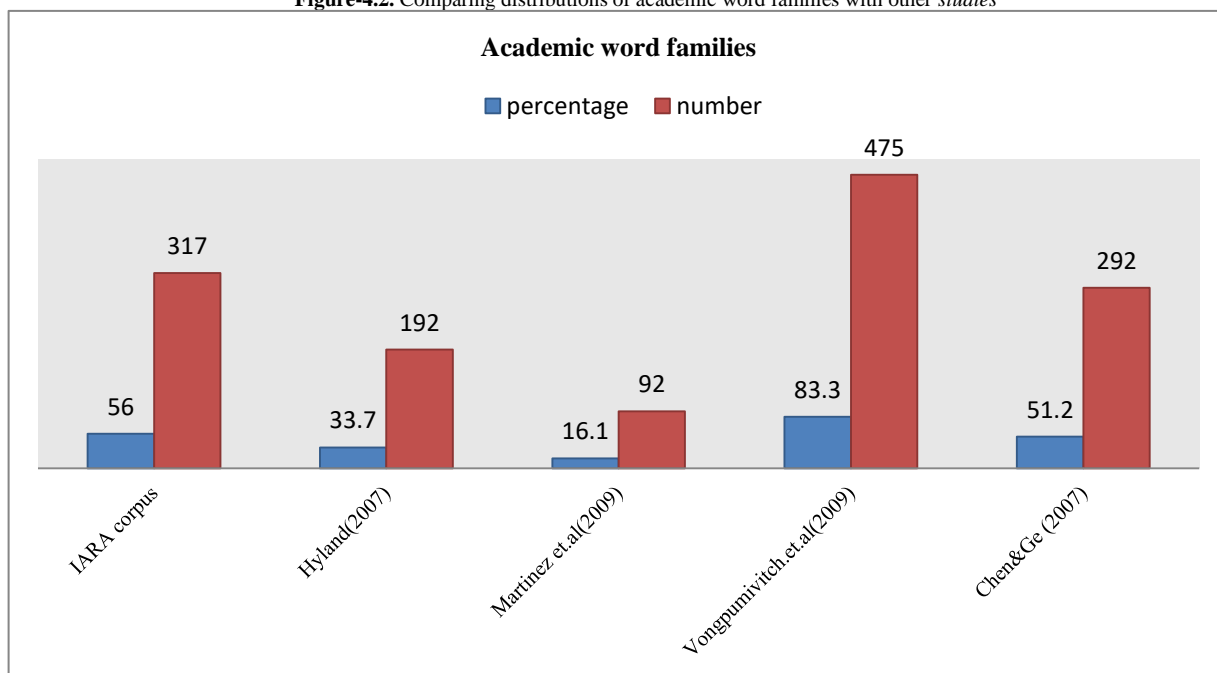
It is also interesting to observe that the majority of the families had only one member, in other words, only one form of the word was used. The families with only one member which occurred with Coxhead's AWL headwords were 110 and without head words were 57. As shown in Table 4.3, there were 167 word families that reached around 53% of the total academic word families in the Islamic academic corpus, as shown in Table 4.3.

In the IARA corpus there were only 3 word families that occurred with single form. Examples of these word forms are *research*, *physical* and *legal* which accounted for 15% of the 20 most frequent families. Besides, other five families occurred with 2 word forms (25%), 7 families with 3 word forms (35%), three families with 4 word forms (15%) in the IARA corpus. The highest numbers as shown in Table 4.4, with seven members of a family and five members of a family both occurred in the present corpus. Each of the word family accounted for 10% of the top 20

most frequent families. Thus, the IARA corpus had headwords with the highest number of the family members compared to the other corpus.

This research also compared the top 100 families in the IARA Corpus with those in Coxhead's Sublist 1, which includes the 60 most frequent families in the AWL. Table 4.4 shows the 100 families in this list, with the first 60 families shaded. The words that coincide with Coxhead's items in Sublist 1 are shown in italics. Based on the table 4.12, it was found that within the first 60 families, only 31 coincided with Coxhead's, with four items less than the number of coinciding items examined by Hyland and Tse (2007), while 5 items more than the number of coinciding items listed by Martinez in relation to their corpus. In addition, within the top 100 families, 38 coincided with Coxhead's Sublist 1, only 3 items more than the 35 words that coincided in Hyland and Tse's comparison of their top 60 items with Coxhead's. Finally, when considering all the academic word families in this entire corpus (see Table 4.4 and Table 4.5 for more precise information), the researcher determined that 53 (48 with AWL headwords +5 without AWL headwords) belonged to Coxhead's Sublist 1. Again, this list is much higher than Martı́nez et al.'s result, who found that only 33 belonged to Sublist 1 according to their whole corpus.

Table-4.4. The top 100 academic word families in the IARA Corpus and in Coxhead's Sublist 1*

1	Tradition	2	Research	3	Process	4	Authority
5	physical	6	*legal*	7	*approach*	8	philosophy
9	*concept*	10	*theory*	11	*economy*	12	rational
13	reveal	14	*role*	15	*context*	16	*issue*
17	*create*	18	culture	19	*analysis*	20	*principle*
21	*individual*	22	text	23	aspect	24	contemporary
25	*source*	26	hence	27	found	28	element
29	community	30	institutions	31	perspective	32	*establish*
33	framework	34	*major*	35	*interpret*	36	fundamental
37	status	38	*data*	39	*method*	40	*specific*
41	*structure*	42	*similar*	43	*constitute*	44	ethical
45	*evident*	46	ultimate	47	objective	48	*period*
49	*define*	50	mental	51	positive	52	vision
53	whereas	54	*derive*	55	*factor*	56	external
57	communication	58	*significant*	59	instance	60	conduct
61	conclude	62	*involve*	63	classical	64	despite
65	phenomenon	66	evolution	67	final	68	empirical
69	notion	70	technology	71	attitude	72	furthermore
73	logic	74	interaction	75	military	76	*respond*
77	goal	78	civil	79	*consists*	80	*environment*
81	foundation	82	intelligence	83	category	84	revolution
85	focus	86	achieve	87	conflict	88	likewise
89	dimension	90	relevant	91	internal	92	*percent*
93	*identify*	94	negative	95	perceive	96	primary
97	imply	98	*benefit*	99	ideology	100	impact

Note: *The words in italics occur in both lists.   31+7

Table-4.5. Academic word families in the Islamic academic research article corpus

No of Sublist	Word families[a]	Word families[b]	Total word families	Percentage families %	Cumulative families	Cumulative percentage families %
Sublist 1	48	5	53	16.72	53	16.72
Sublist 2	36	10	46	14.51	99	31.23
Sublist 3	28	9	37	11.67	136	42.9
Sublist 4	33	9	42	13.25	178	56.15
Sublist 5	31	8	39	12.30	217	68.45
Sublist 6	18	5	23	7.26	240	75.71
Sublist 7	24	13	37	11.67	277	87.38
Sublist 8	9	7	16	5.05	293	92.43
Sublist 9	13	4	17	5.36	310	97.79
Sublist 10	4	3	7	2.21	317	100
Total	244	73	317	100		

Note: Word families[a] = AWL word families which occurred with Coxhead's headwords.
   Word families[b] = AWL word families which occurred without Coxhead's headwords.

It can be seen that most of the frequently-occurring words in the IARA corpus come from Coxhead's first two sublists (Sublist 1 and Sublist 2). In fact, 31.23% of the 317 word families that occur more than 17 times in the present study come from Coxhead's Sublist 1 and Sublist 2.

The second objective of this research was to investigate on the existence of non–AWL frequent content words and other specific lexical items that function as academic words in the present study.

4.3. Non-AWL Content Word Forms in the IARA Corpus

Frequency analysis shows that there are many non-AWL content word forms that occur at least 50 times in the IARA corpus and five times in each of the articles. It can be assumed that it is useful for learners to know these frequently-occurring non-AWL content words, especially since these words reflect some important concepts in the field of Islamic studies.

One of the findings of this study is that many of the non-AWL content words, which occurred frequently, include specialized terms employed in Islamic academic research. Some examples of the non-AWL words are: *Islamic/Islam, Muslim/Muslims, Religious, Ibn, Qur'an/Quranic, Prophet/prophets, ALLAH, Hadith, interlanguage, metalinguistic, Religious, Sciences, spiritual, Muhammad, secular/secularization, forgiveness, ILM, beliefs/believe/believed/believers, intellect/ intellectual, worldview, rights, values, discourse, fiqh, guidance, teachers/teachings, apostasy, sign/signs, sunnah, Christian/Christians, learning, doctrine, punishment, theological, comparative, follows, merely, scholar/scholars, affairs, paradise, era, modernization/modernity, prayer, ontological,* and *acceptable*. This finding is similar to Vongpumivitch *et al.* (2009b) who also found some field-specific words being used in their corpus of applied linguistics, such as *morphology/morphological*, *phonological*, *pragmatic*, *semantic/semantically*, and *syntactic/syntax*. Furthermore, the IARA corpus also contains several adjectives, and nouns that indicate country or language, such as *Arab/Arabic, Muhammad, Ibn-Sina, Al-Attas, Turkish, Books, SWT, Moses, Adam, Imam, Jews, cognitive, democracy, European, Ijtihed, consciousness, Mankind/Humankind/Human, and Malaysia*. These words occur frequently, perhaps due to the diversity of research participants or the various languages that were the subject of Islamic studies research.

Finally, the list of non-AWL content word forms also include some practical terms in research methodology. These can be found in academic research papers in Islamic studies as well as other fields in social sciences, such as *noted, said, based, means, meant, discussed, refers, argued, developed, considered, studies* and *according*. These terms are commonly used in both qualitative and quantitative research paradigms. In summary, the non-AWL content word list that we have compiled shows words that are commonly used in Islamic academic studies but are not included in the Coxhead (2000) AWL. These words provide learners a window to the specialized content areas in the field as well as important concepts in academic research.

4.4. Loan Words in the IARA Corpus

The findings revealed that there are many loan words in IARA corpus that are not easily found in the other academic research articles written in English. It can be suggested that in Islamic academic settings, English for Specific Purpose students do not only have to master academic vocabulary, but also need to master some loan words. It can also be assumed that these loan words play a vital role in the comprehension of Islamic academic readings as well as to complete academic writings successfully, since they constitute a substantial percentage of academic research writing. It is not surprising that the use of loan words has long been a matter surrounded by controversy. Different language communities might have different attitudes towards the use of loan words. In monolingual settings, speakers of one language may use words belonging to another language when they fail to retrieve an equivalent way of expressing the same notion in their own language, or they may use loan words on purpose, to evoke meanings that go beyond the mere propositional content of the words used. While the former is seen by purists as a sign of language impoverishment and loss, the latter is frequently associated with erudition and language enrichment. For example, words like *Quran* and S*unnah, Islamic/ Islam*, *Muslims*, *Knowledge, God, Science, Holy Prophet (s.a.w.)/Prophet* and *ALLAH* were used widely in this study. Interestingly, some loan words are explained or translated in English in the Islamic research articles. For example, words like *ijma`* (consensus of the scholars), *maslahah* (deriving and applying a juridical ruling that is in the public interest), *qiyas* (analogy), *ijtihad* (independent reasoning), *talaq* (divorce), *fiqh* (jurisprudence) and *naskh* (abrogation).

5. Conclusion

This study is a response to a call for useful, more valid high frequency word lists, particularly those intended for pedagogical purposes (Read, 2000). It provides empirical findings to understanding (Coxhead, 2000) AWL word frequency, coverage and distributions in the IARA corpus associated with word forms and families. This study also emphasized the crucial role of AWL as envisioned in the academic fields. Even though it is understood that the AWL plays an important role in general academic context, there is a need to apply a field- specific academic vocabulary to help and encourage ESP learners in higher institutions who need to read and write academic articles in their specific fields. The aim of this research is to bring attention to increase the research efforts in this area. The researcher's overall objective is to identify the necessary and high frequency AWL words that play academic functions in a field specific discipline.

For that reason, the main objectives of this study were to find out the frequency, coverage and distribution of Coxhead's AWL words in the Islamic academic research articles (IARA) corpus according to the word types, tokens and word families. This study also aims to find out if there are non–AWL frequent content words and other specific lexical items that function as academic words in the IARA corpus. Based on the findings it can be seen that AWL items have high text coverage (around 8.15% by tokens and 10.02% by types) in IARA corpus. It supports Coxhead (2000b) Coxhead (2000) and Coxhead and Nation (2001) claim that AWL word covers around 8% to 10% of academic texts across a wide range of subject disciplines. This clearly demonstrates that AWL words are indeed a group of essential lexical items in the Islamic studies.

The present study has provided a list of Coxhead's AWL word families that are frequently used in IARA corpus. Out of the 570 AWL word families, only 317 (56%) were found to be frequently used in the IARA. The result indicates that Coxhead's academic word list does not represent an overall picture of the frequent words in the Islamic field. The ranking positions of quite a number of AWL words found in the present corpus are very different from those in the Coxhead's AWL itself. Thus it can be concluded that some academic words of high-frequency in Coxhead's corpus are not used as often in the IARA corpus, and vice versa.The study also provides some non-AWL content words that are frequently used in articles related to Islamic studies. This clearly indicates that attention should also be given to other academic words in the Islamic field. In addition, there is also an urgent need to establish a useful and valid Islamic academic word list in the near future.

The most outstanding result was that the list of frequent words from the AWL in the corpus was found to be even more restricted than (Vongpumivitch *et al.,* 2009b) applied linguistics lists but much higher than (Martinez *et al.,* 2009) agriculture corpus and Chen (2007) medical articles lists. This list, which is the outcome of this study, cater for the specific needs of ESP learners with special reference to Islamic studies.

Acknowledgement

References

Bauer, L. and Nation, P. (1993). Word families. *International journal of Lexicography,* 6(4): 253-79.

Campion, M. E. and Elley, W. B. (1971). *An academic vocabulary list.* NZCER: Wellington.

Chen, Y. K. (2007). Adaptive sampling enhancement for Hotelling's T2 charts. *Eur. J. Oper. Res.,* 178(3): 841–57.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly,* 34(2): 213–38.

Coxhead, A. (2000b). A new academic word list.

Coxhead, A. and Nation, I. S. P. (2001). *The specialized vocabulary of English for academic purposes.in j flowerdew and M. Peacock, editors, research perspectives on english for academic purposes.* Cambridge University Press: Cambridge

Fuentes, A. C. (2001). Lexical behavior in academic and technical corpora: implications for esp development electronic version. *Language Learning & Technology,* 5(3): 106-29. Available: http://llt.msu.edu/vo15num3/curado/

Ghadessy, M. (1979). Frequency counts, word lists, and materials preparation: a new approach. *English Teaching Forum,* 17(1): 24-27.

Hyland, K. and Tse, P. (2007). Is there an academic vocabulary ? *Tesol Quarterly,* 41(2): 235-53.

Jing Wang, Shao-lan Liang and Ge., G.-c. (2008). Establishment of a medical academic word list. *English for Specific Purposes,* 27: 442–58.

Lam, J. (2001). *A study of semi-technical vocabulary in computer science texts, with special reference to ESP teaching and lexicograph.* Language Centre, Hong Kong University of Science & Technology. 3:

Laufer, B. (1998). The development of passive and active vocabulary in a second language: same or different? *Applied linguistics,* 19(2): 255-71.

Martinez, A. I., Beck, S. C. and Panza, C. B. (2009). Academic vocabulary in agriculture research articles, A corpus based study. *English for Specific Purposes,* 28(1): 183–98.

McEnery, T., Xiao, R. and Tono, Y. (2006). Corpus-based language studies, An advanced resource book. *Taylor & Francis*:

Mudraya, O. (2006). Engineering English, A lexical frequency instructional model. *English for Specific Purposes,* 25(2): 235-56.

Praninskas, J. (1972). *London American university word list.* Longman. 10.

Read, J. (2000). *Assessing vocabulary.* Cambridge university press: Cambridge. 1-85.

Santos, M. G. (2004). Some findings on the academic vocabulary skills of language-minority community college students. *Focus on Basics,* 6(D): 7-9.

Saville-Troike, M. (1984). What really matters in second language learning for academic achievement? *Tesol Quarterly,* 18(2): 199-219.

Shaw, P. (1991). Science research students composing processes. *English for Specific Purposes,* 10: 189–206.

Vongpumivitch, V., Huang, J. Y. and Chang, Y. C. (2009a). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research paper. *English for Specific Purposes,* 2: 235–56.

Vongpumivitch, V., Huang, J. y. and Chang, Y. C. (2009b). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research paper. *English for Specific Purposes,* 2: 235–56.

Wang, J., Nation, R. L., Evans, A. M. and Cox, S. (2004). Isolated rat kidney perfused with dextran and bovine serum albumin, a stable model for investigating renal drug handling. *Journal of pharmacological and toxicological methods,* 49(2): 105-13.

West, M. (1953). *A general service list of english words.* Longman: London.

wNagy, W., Anderson, R., Schommer, M., Scott, J. A. and Stallman, A. (1989). Morphological families in the internal lexicon. *Reading Research Quarterly,* 24: 262–81.

Xue, Guoyi, Nation and I.S.P. (1984). A University List. *Language Learning and Communication,* 3(2): 215-29.