Content list available at www.urmia.ac.ir/ijltr

Iranian Journal of Language Teaching Research

Urmia University

Profiling vocabulary in psychology journal abstracts: A comparison between Iranian and Anglo-American journals

Is'haaq Akbarian [a,*], Zahra Ghanbarzadeh [a], Mohammad Afzali Shahri [a]

[a] *University of Qom, Iran*

ABSTRACT

Lexical profiling has yielded fruitful results for language description and pedagogy (Liu, 2014), and particularly highlighted the significance of academic vocabulary for EFL learners in this process. This investigation, likewise, attempts to comparatively profile the vocabulary, more particularly the academic vocabulary, in the 'abstract' section of scholarly articles in Iranian and Anglo-American refereed journals in psychology. Iranian journals under study publish articles in Persian but also include an English abstract whereas the latter publish papers in English. For this purpose, a corpus (consisting of 307,126 words), with two sub-corpora of almost similar size and characteristics, was collected from Iranian and Anglo-American journals and analyzed through the software Range. The analyses conducted show a coverage of over 15 percent and the use of over 500 words of the Academic Word List (AWL) in both Iranian and Anglo-American sub-corpora. However, there are variations in academic and nonacademic vocabulary use in abstracts across the two sub-corpora above. Most of the academic words used belong to the beginning AWL sub-lists. Pedagogical implications are made for reading and writing, particularly in EAP contexts.

Keywords: AWL; academic vocabulary; EAP English; journal abstracts; psychology

ARTICLE HISTORY

Received: 14 July 2016 **Revised version received:** 14 Sep. 2016

Accepted: 7 Dec. 2016 **Available online:** 1 Jan. 2017

* Corresponding author: Department of English Language & Literature, University of Qom, Iran
Email address: i-akbarian@qom.ac.ir

10.30466/ijltr.2017.20342

Introduction

The Academic Word List (AWL), developed and validated by Coxhead (2000), has received much attention since its appearance. The list includes 570 word families from a corpus of 3.3 million words from a range of academic disciplines and genres, selected based on the three criteria of frequency, range, and specialized occurrence. Each word family includes head-words plus their inflections and derivations (3,107 words altogether). For instance, the head-word '*access'* contains the following inflected and derived members: *accessed*, *accesses*, *accessibility*, *accessible*, *accessing*, *inaccessible*. The AWL is divided into 10 sub-lists, each with 60 word families, except the last sub-list that has 30 families.

An academic word assumes functions in academic writing (Hirsh, 2010) which a general word cannot. It occurs across different texts and genres, belongs to the academic world and discourse, and covers about 10% of any academic text (Coxhead, 2012).

As Liu (2014) states, the findings of corpus linguistics, more particularly the dimension of profiling vocabulary within this field of study, have produced interesting results as well as options for linguists to describe language and for practitioners or language teachers to teach language efficiently. The many studies conducted on the profiling of academic vocabulary across different genres and in different disciplines conclude that this limited list of vocabulary accounts for a noticeable coverage in academic texts (e.g. Chen & Ge, 2007; Li & Qian, 2010; Vongpumivitch, Huang, & Chang, 2009), and thus a good return for learning. Yet, the review of the related literature reveals that profiling (academic) vocabulary has rarely been approached by the researchers from a comparative perspective thus far. Below, we will survey the literature on the role of the academic vocabulary, in general, and their frequency in the corpora compiled by other researchers in particular, especially with reference to AWL (Coxhead, 2000).

Review of the Related Literature

Some studies concern the significance and/or acquisition of academic vocabulary for EAP/ESP purposes. Coxhead (2012) indicates the importance of academic vocabulary in the life of a university student, showing that writers at university level are aware of the importance of vocabulary and the audience of their writing with specific expectations of their lexical choices.

With approximately 1,016,000 words across different disciplines, Thurstun and Candlin (1998) investigated, among other things, the rhetorical functions realized by academic words across a variety of academic texts. Their findings suggest that "the words chosen [i.e., unknown words introduced to the students and worked on through concordancing in the study] are those that the students need as basic tools for academic writing" (p. 277). In addition, Anderson and Freebody (1981) recognize the link between reading comprehension and vocabulary growth, and report that the students in their classes most often identified academic words as the unknown words in academic texts. Evans and Green (2007) studied the language problems that Cantonese-speaking students encountered at Hong Kong's largest English-medium university. These investigators find, among other obstacles, that students generally had inadequate receptive and productive vocabularies Also, Baker (1988) argues that academic words have a significant role in structuring the writer's argument. Thus, their learning is essential to successful interpretation of the writer's intentions. Meanwhile, other investigations consider academic vocabulary as the most difficult part of academic writing in English for the students to acquire (Li & Pemberton, 1994, cited in Chen & Ge, 2007; Santos, 2002; Shaw, 1991).

Another group of researchers have attempted to determine the frequency and text coverage of academic vocabulary or even tried to develop their own lists of academic words. For instance, Mudraya (2006) collected a corpus with 2,000,000 running words from compulsory textbooks offered to 13 engineering disciplines and established her own academic word list, containing 1,200 word families. She found that some verbs, e.g. *assume*, *define*, *illustrate*, *indicate*, *occur*, *require*, and *sketch* occurred very frequently in the corpus, as similarly as they did in Coxhead's (2000) AWL. She argues for more attention to academic vocabulary for ESP students.

With a corpus of 190,425 words out of 50 English medical research articles and a self-designed computer program, Chen and Ge (2007) reported a 10.073% coverage by the AWL. The AWL word families in the five separate sections of a medical research (i.e. their corpus) distributed in the following proportions: abstract (11.185), introduction (10.258), materials and methods (9.713), results (9.283), and discussion (10,861). The proportion in the *Abstract* was the highest. They also found that 292 (51.2%) word families, out of the 570 items in the AWL, had a frequent use in English medical research articles and that the academic words used had a disperse distribution throughout a whole medical research article. The researchers demonstrate "that academic words are indeed a set of important word items" in medical research articles (p. 513). They argue, however, that AWL does not completely represent the frequent academic words in medical research articles and that academic words have several rhetorical functions to perform in academic texts, especially in medical research articles, as in their study.

As a follow-up to Chen and Ge (2007) who suggest a medical academic word list and inspired by the AWL (Coxhead, 2000), Wang, Liang, and Ge (2008) developed a Medical Academic Word List based on a corpus-based study of medical research articles (i.e., 1,093,011 tokens) across different sub-disciplines in medical science. The list included 623 word families accounting for 12.24% of the running words used in the articles investigated.

Hyland and Tse (2007) studied the frequency, range, preferred meanings and forms, and patterns of collocation in the AWL items. They compiled a corpus from a collection of academic disciplines and genres, including 3.3 million running words. The analyses depicted that the AWL accounted for "an impressive 10.6% of the words in the corpus". The researchers identified "items as *frequent* if they occurred above the mean for all AWL items in the corpus" (p. 240). With this approach, these researchers could regard only 192 word families, roughly one third of the items, in the AWL as frequent. Despite the text coverage of 10.6% of the AWL in their corpus, the individual words in the list perform differently across disciplines in terms of frequency, range, meaning, and collocation. This implies that Hyland and Tse might not have regarded the AWL as general. For them, the list "offers a useful characterization of register-level vocabulary choices" (p. 250) to examine specific practices within their own fields.

With a corpus (i.e. around 6.3 million words out of 25 text types) of financial services industry in Hong Kong and using the software RANGE, Li and Qian (2010) found that the AWL covered 10.46% of the corpus. Also, there was variation in the AWL across the text types, suggesting that the proportion of technical words in the text types was different. However, a strange and interesting result reported by the researchers is that the top 10 AWL "word families achieve a cumulative coverage of 22.03% in the corpus" (p. 405). The total coverage of 10.46% as reported above does not correspond with the cumulative coverage of 22.03% by the top 10 word families.

Motivated by the odd coverage of 22.03% in the financial corpus analyzed by Li and Qian (2010), Neufeld, Hancioglu, and Eldridge (2011) examined their data, illustrating the erroneous output in profiling vocabulary and the hurdles in using the AWL list in filtering academic vocabulary. They considered how RANGE processed non-ASCII (American Standard Code for Information Interchange) characters. They 'cleaned' the Li and Qian's corpus of 6.3 million words to 5,754,441

through applying filters and restricting words to the first 20,000 word families of the British National Corpus. The results showed that the AWL accounted for 11.6 of the coverage in the financial corpus. They showed "the same list of ten words that Li and Qian had identified, but their percentage coverage was miscalculated (roughly by a factor of 10) partly as a result of the fault in text conversion and processing using RANGE" (p. 534).

To criticize the usefulness of a general academic vocabulary, Martínez, Beck, and Panza (2009) attempted a corpus-based study of agriculture research articles to detect the words in the AWL in their corpus in line with Hyland and Tse (2007) and their criterion. Their analysis provided a restricted list of 92 words, giving 9.06% text coverage. They further found through qualitative analysis that some words in the list had genre-specific meanings and behaviors. Some words of the AWL had technical meaning and many general words had academic meaning in their corpus. For instance, *research* and *outcome* occurred 235 and 10 times, respectively in their corpus whereas the words *study/studies* and *results* from the General Service List (GSL) (West, 1953) occurred 1539 and 1270 times, respectively. GSL comprises the 2,000 most widely and frequently used words of English language, including function words.

Vongpumivitch et al. (2009) explored the use of the AWL words in 200 applied linguistics articles, published in five scholarly journals: *Applied Linguistics*, *Language Learning*, *Second Language Research*, *TESOL Quarterly*, and *The Modern Language Journal*. The AWL accounted for 11.17% of the running words in their corpus (1.5 million words). Based on Coxhead's (2000) criteria of frequency and range in word selection, 475 word forms in the AWL appeared more than 50 times in their corpus and *not* less than five times in the five journals under study. Furthermore, they detected 128 non-AWL word forms in the corpus as such. The latter group contained specialized terms on language education as well as research methodology, plus the countries and languages mostly involved in the studies.

Hancioglu (2009) contrasted the lexical profile of 100 abstracts produced by post-graduate EFL students with a group of 100 abstracts that were produced by 'expert' post-graduates who spoke English as a native or second language. 'Novice' or non-native writers used around 3,500 words effectively, whereas the native speakers of English used 2,000 more words. The non-native data exhibited extensive use of higher frequency vocabulary, a tendency to repeat similar items. Non-natives were unable to use appropriate collocations as well as lexico-grammatical patterns. The non-natives relied on fewer lexical items in expressing the same concepts than the native speakers of English did. Thus, 95% of the writing produced by non-natives consisted of only 2,000 words in an academic context. Hancioglu argues that non-natives have to learn more words in terms of breadth and depth, especially academic words, to perform as native speakers of English do.

Cobb and Horst (2004) attempted to identify a French list resembling the AWL, built lists of the 2,000 most frequent word families in French into *Vocabprofil*, a program for profiling lexical frequency online, and tested their coverage potential. They analyzed newspapers, popular expository texts, and medical texts, and detected distinct and consistent profiles for these French texts. The researchers then compared parallel texts in French and English. They showed that the 2,000 most frequent word families in French approximately provided an 85% coverage, only achievable with the 2,000 most frequent English words and the 570 words in the AWL. The researchers tentatively conclude that in French there is no need for an extra list, similar to AWL, to facilitate academic text comprehension as in English. Apparently, the 2,000 most frequent words serve both every day as well as academic purposes.

The studies reviewed above indicate how significantly the AWL contributes to the coverage of the texts considered. Though the list only accounts for an average of 10 percent, the coverage is quite

indispensable, due to the role of academic vocabulary in academic discourse. The survey also reveals that most of the corpora compiled and analyzed emerge from published, peer-reviewed, and scholarly sources. An exception is Hancioglu (2009) that contrasts native and non-native abstract writings, which are not refereed and published either. However, Hancioglu focuses primarily on vocabulary in general, but *not* on academic vocabulary. Furthermore, the survey above points out the scarcity of research on any comparison of lexical profiling between refereed non-Anglo-American and Anglo-American academic texts, for instance the profile of vocabulary across Iranian and Anglo-American research outlets, particularly on displaying the performance of academic vocabulary in the two channels of appearance of academic writing.

Given the points above, a suitable potential for such a comparative investigation in academic writing is the abstract of a scientific journal article. An efficient abstract includes five essential things or elements: (a) a background statement, (b) purpose of the study, (c) data source (participants or materials), (d) methods for data collection and analysis, (e) general results, and (f) conclusions and related implications (Perry, 2011). Academic vocabulary has a large contribution to materializing these five essential elements in the abstract. Along with a cluster of linguistic features (lexico-grammatical features), academic words can perform a number of functions in structuring the argument as well as the components of research articles, as some researchers (Baker, 1988; Hirsh, 2010; Thurstun & Candlin, 1998) indicate. Thus, the abstract somehow summarizes the contents of any manuscript and, to some extent, benefits from the functions performed by academic vocabulary.

Note that scientific rigor, soundness, and research originality, as well as scientific writing, contribute to a paper's publication. Scientific writing is both a question of language and of discourse (e.g. argumentation). Yet, comparing Iranian and Anglo-American journals might also reveal some facts about differences in vocabulary use, in general, and the behavior of academic words, in particular, across the two outlets that could not be detected otherwise. Also, the comparison might help discover key words in the corpora that would possibly differentiate one corpus from the other.

We selected the genre of journal articles because, as Baker (1988) observes, "Scientific journal articles in general are among the obvious examples of the role of English as an international language" (p. 93). These journal articles are written and read by an increasingly large number of native and non-native speakers of English. Besides, due to the variety of topics in journal articles, the abstracts, if collected in large numbers, provide a more suitable corpus on one section of the article. A corpus with such a variety, in turn, enables researchers to explore the academic vocabulary, or even to create a more extensive and less controversial list of academic words.

Purpose of the Study

Anglo-American journals are more referred to and cited, and thus are more visible, than Iranian ones. For instance, there is more tendency among Iranian researchers in the field of applied linguistics to cite papers in Anglo-American journals than those published in Iran. A comparison of the citations by Iranians of the first author's papers and those of his colleagues in other universities (on the publication date) in Google Scholar provide evidence to this phenomenon.

The intention of this investigation, therefore, is to detect any possible similarities and differences that might exist in lexical profiling, especially profiling the academic words, in the abstract section of Iranian and Anglo-American research outlets. Due to the fact that the abstract section of any research publication, after the title, might (de)motivate readers to read or not to read the remainder of the document (APA, 2009), the current research will focus on the abstract section of journal

articles. Normally, the searching readers evaluate the suitability of the papers to their interest through considering the title and abstract.

Note that this study will be limited to Iranian and Anglo-American journal abstracts in psychology. The reason for focusing on the discipline of psychology is that, as the review of the literature in this area shows, the studies in vocabulary research, conducted up until now, have not dealt with any lexical profiling on psychology separately. There is no previous study that identifies a particular academic word list in the field of psychology, and neither is there a study that compiles such a specific word list in English for Iranian learners. Also, a search in Google Scholar testifies to this scarcity.

Consequently, the current research is novel since there is no study on comparing academic writings in terms of (non)academic words, especially in a very specific area, supervised and published nationally by Iranian psychology experts in an EFL context with those supervised and/or produced in Anglo-American journals by native speakers of English. More particularly, the current study attempts to answer the following research questions (RQs):

RQ 1: Based on the comparison of lexical profiles, do the abstracts of journal articles in psychology, published in Iranian and Anglo-American journals, differ from one another?

RQ 2: Do AWL word forms provide the same lexical coverage in article abstracts in psychology across Iranian and Anglo-American journals?

Method

The Corpora

We selected to study the *Abstract*, considering the following issues: Firstly, to alleviate any bias existing in word counts mainly based on longer texts (Coxhead, 2000), we used short texts. Secondly, longer texts might be more related to one topic (Coxhead, 2000), whereas short texts allow more variety in terms of topic. For instance, a corpus of 6,000 words, on average, might equal the vocabulary of one article with only one topic but almost that of 30 abstracts with 30 different topics in the same field. Consequently, the words in these two corpora might differ from one another. Thirdly, corpus collected from numerous different authors will contribute to a more balanced and unbiased type of academic vocabulary list. Fourthly, each of the seven subject areas within each of the four sub-corpora (i.e., arts, commerce, law, and science, compiled by Coxhead (2000)), included approximately 125,000 running words. This might indicate that a similar corpus size on the abstracts for a given specific subject might prove sufficient for analyzing the frequency coverage of the AWL. More importantly, Chen and Ge (2007) had a corpus of 190,425 words from 50 medical articles for all the components of an article. Furthermore, comparing a sub-sample with the full version of a sub-corpus under study, Adolphs and Schmitt (2004) found that there is "no obvious relationship between corpus size and the magnitude of lexical coverage" (p. 47). Given the specificity of our corpus, this suggests that corpus size does not have a great effect on lexical coverage in our case.

The articles were differentiated on the basis of whether they are produced and supervised nationally by Iranians from those written and/or supervised internationally in line with Wood's (2001) criteria; the authors' names should be names native to the country concerned and also be affiliated with an institution in their respective countries. Furthermore, the editors or editorial members of Iranian journals under study are all Iranians. However, Anglo-American journals are internationally run and supervised by editorial members from English-speaking countries or edited by native speakers

of English. That is, the English used in the Anglo-American journals under study is supervised and edited by a team of well-recognized and important scholars, including native speakers of English, whereas Iranian journals under study are run by only a team of Iranian scholars. Moreover, the articles should be reviewed by Iranian or international/English-speaking referees. Note that the papers in Iranian journals are only reviewed by Iranian reviewers, but not English-speaking reviewers, as confirmed by the editors of the journals under study.

Compilation of the Psychological Abstracts Corpus

The Psychological Abstracts Corpus (hereafter PAC) was specially compiled for the study. The study focused only on the final published version of the abstracts in Iranian and Anglo-American journals, and *not* the versions before publication. Therefore, PAC consisted of a large number of the abstracts published in 11 journals in the field of psychology; six Iranian journals and five Anglo-American journals that were randomly selected for the corpus out of a group of the outlets suggested by two specialists in the field. Both specialists served as the faculty members of the University of Qom and had a PhD in psychology, one from the University of Tehran and the other from the University of Isfahan. The first had around 12 years whereas the second had about eight years of teaching experience at the university at BA and MA levels. They had published in Iranian and international journals, and served as the reviewers for prestigious journals. Thus, they had sufficient information about the outlets in their fields.

The previous issues of some Iranian journals are not online, and thus not available for analysis. Therefore, it was decided to have six Iranian journals to compile almost the same amount of the corpus as that of the corpus from the five Anglo-American journals. Only empirical studies were included since they contain abstracts.

The freely available abstracts were compiled by the second author from the internet so as to obtain the PAC corpus. The titles and keywords, adjacent to the abstracts, were excluded from the analysis for the words in both titles and keywords also appear in the abstracts due to the importance of the issues discussed in the articles.

It is compulsory to provide an English abstract along with a Persian one for the articles published in Persian in Iran. Iranian data included lots of spelling mistakes, due to low English proficiency of the authors or translators or typists, which were corrected before analysis, in line with Vongpumivitch et al. (2009) and Wang et al. (2008). Proper names of people and places as well as acronyms, in the two corpora, were also put in a file to be used as a stop list within the software RANGE, and thus excluded from the analyses. The final PAC corpus has approximately 307,126 words and consists of the following sub-corpora:

Iranian Journals

The corpus for Iranian Journals came from Scientific Information Database (http://www.sid.ir/En/index.asp). Unlike Anglo-American journals, some issues or volumes did not exist online in the database above. Also, the journals were not consistent in the number of issues per volume or year. Table 1 displays an overview of the number of volumes and issues considered, and of words from each Iranian journal:

- *Journal of Psychology (JP)*
- *Journal of Psychology (Tabriz University) (JPTU)*
- *Iranian Journal of Psychiatry and Clinical Psychology (IJPCP)*

- *Developmental Psychology (Journal of Iranian Psychologists) (JIP)*
- *Journal of Psychology and Education (JPE)*
- *Studies in Education and Psychology (SEP)*

Table 1
Iranian Journals

Iranian Journals				
Journal	No. of volumes	No. of issues	No. of words before using stop list	No. of words after using stop list
IJPCP	5	19	46,866	44,830
JIP	5	20	17,707	16,662
JP	6	17	19,090	18,437
JPE	7	20	26,632	25,902
JPTU	6	19	22,981	22,190
SEP	9	15	23,831	23,088
Total	38	110	157,107	151,109

Anglo-American Journals

The data were taken from ScienceDirect (http://www.sciencedirect.com). Table 2 shows volume and issue numbers considered, and the number of words each Anglo-American journal provides to the final PAC corpus:

- *Acta Psychologica (AP)*
- *Contemporary Educational Psychology (CEP)*
- *Journal of Applied Developmental Psychology (JADP)*
- *Journal of Experimental Social Psychology (JESP) New Ideas in Psychology (NIP)*
- *An International Journal of Innovative Theory in Psychology*

Table 2
Anglo-American Journals

Anglo-American Journals				
Journal	No. of volumes	No. of issues per vol.	No. of words before using stop list	No. of words after using stop list
AP	4 (Vols. 138/2009 – 141/2012)	12 (Issues. 1 – 3)	34,510	33,829
CEP	5 (Vols. 33/2008 – 37/2012)	20 (Issues 1 – 4)	23,052	22,688
JADP	3 (Vols. 31/2010 – 33/2012)	18 (Issues 1 – 6)	17,355	17,036
JESP	2 (Vols. 47/2011 – 48/2012)	12 (Issues 1 - 6)	58,497	57,947
NIP	6 (Vols. 25/2006 – 30/2012)	18 (Issues 1 – 3)	22,853	22,517
Total	20	80	156,267	154,017

Data Analysis

To analyze the corpus, the software RANGE (Heatley, Nation, & Coxhead, 2002) was used. The analysis adopted word family as the unit of consideration. 'Word form' is defined as any string of letters, bounded by space (Sinclair, 1991). For instance, *create*, *creation*, *creative(ly)*, *creator(s)*, and *created* are separate word forms, belonging to the same word family.

The software RANGE is utilized to compare the vocabulary profile of up to 32 different texts concurrently, based on the following four levels: Level one and two include the 2,000 most frequent words in GSL (West, 1953), level three contains the 570 frequent academic words (Coxhead, 2000), and level four consists of the words, not found in the levels above.

For each word, RANGE gives a figure or a number: (1) of the texts the word appears in (range), (2) of the times the actual headword type occurs in the texts (headword frequency), (3) of the times the word and its family members appear in the texts (family frequency), (4) for each of the texts in which the word is used (frequency). The software can also compare or profile the lexical coverage of the texts against certain vocabulary lists, provide the percentage of the lexical coverage by the lists, and/or detect the similarities and differences of vocabulary use across several writings. Please refer to the instructional manual of the software RANGE for further use.

Results and Discussion

This section provides a profile of the tokens, types, and word family as well as the type/token relationship across the corpora. *Tokens* are the number of running words in a text, while *types* are the number of different words.

Lexical Profile in Iranian Sub-corpus

There are totally 151,109 tokens, or words, in Iranian sub-corpus (Table 3). The first 1,000 most frequent words account for 66.06% of the words in the data whereas, the second level provides coverage of 6.77% of the tokens. However, AWL word forms account for 15.19% of the entire Iranian sub-corpus. The first two levels of most frequent words together cover 72.83% of the tokens in Iranian data (Table 3). Adding the percentage (15.19%) of the AWL, the coverage will amount to 88.02% of the tokens. Finally, words not found in either of the lists above make up 11.98% of the tokens in the data.

Table 3
Lexical frequency profile in Iranian sub-corpus

Word list	Tokens/%	Types/%	Families
One	99828/66.06	1950/26.14	817
Two	10234/ 6.77	706/ 9.46	374
Three	22947/15.19	1352/18.12	502
Not in the lists	18100/11.98	3452/46.27	?????
Total	151109	7460	1693

The third column on 'types' shows another aspect of the data; the first level gives 1,950 (26.14%) words to Iranian sub-corpus. However, the second level gives only 706 types of words (9.46%), approximately half the number of the AWL level with 1,352 word types, that cover 18.12% of the data. The 3,452 word types of level four, not found in the list, account for the highest coverage (46.27%) of the total types existing in Iranian sub-corpus.

As Table 3 shows, there are totally 1,693 word families in Iranian sub-corpus as the software RANGE does not give the number of word families in 'Not in the lists' level. Out of this number, 817 (48.26%) word families belong to the first level whereas only 374 (22.09%) word families in the second have been depicted in our profile. According to Table 3, 502 (29.65%) of the 570 word families in the AWL have been used in Iranian sub-corpus of PAC.

Lexical Profile in Anglo-American Sub-corpus

There are 154,017 tokens in the Anglo-American sub-corpus (Table 4). The first two levels together produce 66.08% and 5.69% of the lexical coverage for the data, respectively. In other words, the two lists combined give coverage to 71.77%. However, the AWL words in the third level cover 16.66% of the Anglo-American sub-corpus of the entire PAC. The fourth level 'Not in the lists' account for the remaining percentage of 11.57% words.

Table 4
Lexical frequency profile in Anglo-American sub-corpus

Word list	Tokens/%	Types/%	Families
One	101780/66.08	2240/22.19	841
Two	8759/ 5.69	977/ 9.68	461
Three	25663/16.66	1639/16.24	527
Not in the lists	17815/11.57	5238/51.89	?????
Total	154017	10094	1829

The third column on the percentage of word 'types' reveals another aspect of the profile of the words running in the Anglo-American sub-corpus. There are 10,094 word types in this section, with the first level including only 2,240 (22.19%) word types and the second level having the least number of word types, 977 (9.68%). However, the 1,639 AWL word types make up 16.24% of the total percentage of the word types existing in the data. The highest percentage of the word types in the Anglo-American sub-corpus, i.e. 51.89%, goes to the word types that are not found in either of the lists above (Table 4).

There are 1,829 word families in the general list of the most frequent vocabulary and the AWL word families, used in the Anglo-American data collected (Table 4). Out of this figure, the first and second levels comprise about 45.98% and 25.21%, respectively, whereas the AWL word families include about 28.81% of the word families.

Given the percentages above, the academic words in our corpus cover a considerable proportion of texts. Though higher in coverage, our finding corresponds with the noticeable coverages reported in Chen and Ge (2007), Wang et al. (2008), Li and Qian (2010), and Vongpumivitch et al. (2009). Our findings are also in line with the coverage reported by Hyland and Tse (2007) and Martínez et al. (2009) though they had different perspectives in lexical profiling and are more concerned with criticizing the AWL (Coxhead, 2000).

Comparison of the Two Profiles

Comparing the two sub-corpora (i.e., Iranian versus Anglo-American) of PAC reveals yet another aspect of the profile of the data (Table 5). Apparently, almost the same percentage of the tokens and word families of the first level occurred in both Iranian and Anglo-American sub-corpora, but not the same percentage of word types of that level existed in the two sub-corpora of PAC corpus.

This point raises the interesting observation that it seems as though the writers in Anglo-American journals used nearly the same number of families as the Iranians, but had utilized more word types in their writings. Thus, for example, the Iranians only had used 'process' and 'proceed', while the Anglo-American journal writers had produced 'processual' and 'procedure' as well. In terms of coverage, 1,950 (26.14%) of the types in Iranian sub-corpus gave 3.95% more coverage than the 2,240 (22.19%) ones in Anglo-American sub-corpus. It means that there was more variation of word use among the authors in Anglo-American journals; Iranian writers comparably used fewer words of the first level than the researchers publishing in Anglo-American journals. As to the second level, Iranian writers had used more tokens, but fewer word types and word families than the latter group of authors. We can infer that using fewer tokens, but more word types and word families, points to the richer lexicon and higher vocabulary proficiency of the writer and his or her competence in using the words.

Table 5
Comparing word frequency profiles across Iranian and Anglo-American sub-corpora

Word list	Sub-corpus of PAC	Tokens/%	Types/%	Families
One	Iranian	99828/66.06	1950/26.14	817
	Anglo-American	101780/66.08	2240/22.19	841
Two	Iranian	10234/ 6.77	706/ 9.46	374
	Anglo-American	8759/ 5.69	977/ 9.68	461
Three	Iranian	22947/15.19	1352/18.12	502
	Anglo-American	25663/16.66	1639/16.24	527
Not in the lists	Iranian	18100/11.98	3452/46.27	?????
	Anglo-American	17815/11.57	5238/51.89	?????
Total	Iranian	151109	7460	1693
	Anglo-American	154017	10094	1829

Comparing Iranian and Anglo-American sub-corpora of the data in terms of the AWL words (i.e. level three) apparently shows more variation on the part of the latter for the researchers in Anglo-American journals with respect to the tokens, types, and families in using academic vocabulary. In other words, not only do the authors in Anglo-American journals have richer lexicon (i.e. a strong vocabulary repertoire with variety in word use and knowledge of all word family members), but they are more adept and skillful in using academic vocabulary due to their significance in structuring the conventions, arguments, and elements of research. In spite of the variation, both groups of the authors have used a high percentage of the 570 academic words, with Iranian authors using 502 (88.07%) of the AWL words and the authors in Anglo-American journals using 527 (92.46%) of the words (Table 5). Thus, our finding provides stronger evidence than even Mudraya (2006), and Chen and Ge (2007) for the usefulness of the AWL.

The words 'Not in the lists' above are nearly similar with respect to the percentage of the coverage of the tokens in Iranian and Anglo-American sub-corpora (18,100/11.98% versus 17,815/11.57%, respectively), with 285 more tokens existing and a negligible 0.41% further coverage in Iranian data. However, word types used at this level by the authors in Anglo-American journals are more numerous and various in comparison to the word types used by Iranian authors. According to Table 5, the authors in Anglo-American journals have used 1,786 more word types than the Iranians (Iranian authors = 3,452/46.27% vs. the authors in Anglo-American journals = 5,238/51.89%). The results support the suggestion by Hancioglu (2009) that non-natives have to learn more words, especially academic words, to perform more fluently and accurately as native speakers of English do in academic world, especially on writing.

Table 6 comparatively displays the first 20 words in the first and second most frequent levels as well as academic words across the two sub-corpora of PAC. The table shows a great deal of

similarity in the use of the word types across Iranian and Anglo-American sub-corpora for both levels, despite the rankings of the word types. For instance, the word type 'study' of the first level has been used in both Iranian and Anglo-American sub-corpora of the data but the ranking of the word is different.

Table 6
Comparison of the first 20 word types in the first and second levels across Iranian and Anglo-American sub-corpora of PAC

Types found in base list one						Types found in base list two						Types found in base list three (AWL)					
Iranian data			Anglo-American data			Iranian data			Anglo-American data			Iranian data			Anglo-American data		
Type	*Range*	*Freq.*	*Type*	*Range*	*Freq.*	*Type*	*Range*	*Freq.*	*Type*	*Range*	*Freq.*	*Type*	*Range*	*Freq.*	*Type*	*Range*	*Freq.*
The	6	9805	The	5	7600	Scale	6	527	Performance	5	260	Significant	6	686	Abstract	5	1047
Of	6	7804	Of	5	5876	Health	6	402	Behavior	5	258	Research	6	613	Participants	5	664
And	6	7783	And	5	4849	Anxiety	6	364	Information	5	229	Analysis	6	563	Research	5	413
In	6	4275	In	5	3790	Education	6	288	Examined	5	201	Method	6	518	Task	5	344
To	6	2709	To	5	3731	Sampling	6	270	During	5	182	Data	6	461	Negative	5	216
Were	6	2146	A	5	3048	Skills	6	246	Discussed	5	180	Selected	6	442	Positive	5	207
Was	6	1871	That	5	2451	Sample	6	244	Model	5	162	Factors	6	365	Perceived	5	198
A	6	1756	For	5	1470	Treatment	6	219	Compared	5	156	Positive	6	300	Implications	5	197
With	6	1677	On	5	1286	Female	6	205	Attention	5	140	Academic	6	277	Processing	5	191
That	6	1280	With	5	1284	Educational	6	204	Behaviors	5	100	Negative	6	274	Goal	5	180
This	6	1259	Were	5	1090	Aim	6	196	Skills	5	93	Mental	6	265	Theory	5	175
For	6	1191	As	5	1070	Male	6	175	Educational	4	86	Factor	6	251	Achievement	5	170
Students	6	1097	Is	5	1016	Performance	6	160	Self	5	86	Depression	6	241	Individuals	5	164
On	6	1096	this	5	990	Information	6	153	Models	5	82	Conclusion	5	229	Role	5	163
Study	6r	1088	Are	5	903	Satisfaction	6	153	Examine	5	79	Styles	6	225	Evidence	5	159
Is	6	1045	We	5	879	Model	6	147	Practice	5	79	Validity	6	218	Processes	5	158
Results	6	923	Was	5	870	Behavior	6	144	Behavioral	5	78	Strategies	6	209	Context	5	157
Between	6	915	By	5	862	Females	6	139	Multiple	5	67	Indicated	6	208	Response	5	151
Group	6	797	Study	5	773	Males	6	133	Parents	3	67	Participants	6	205	Goals	5	146
As	6	751	An	5	760	Compared	6	129	Risk	4	66	Analyzed	6	202	Motivation	4	142

Further investigation is needed to determine the degree of similarity and difference between the two sub-corpora in terms of what functions academic words perform, which words they collocate or go with, and so on. There might be reasons for the different rankings of different academic words in the two sub-corpora. However, it is beyond this paper. Readers at this point can refer to the appendix to see the similarities and differences of the occurrence of academic words in the two sub-corpora (See the Appendix). Another topic for further study can be concerned with considering other abstracts from other Asian regions. This piece of further research would allow the future researchers to compare their findings with the patterns observed in Anglo-American and Iranian sub-corpora under study.

To sum up, the comparison above indicates that Iranian abstracts under study seem to contain more frequent vocabulary whereas the word types compared suggest that a higher number and variety of different words are used in Anglo-American abstracts. This suggests that Iranian psychology researchers have not obtained a rich vocabulary repertoire in comparison to the authors in Anglo-American journals in the field whose works collectively contributed to our Anglo-American sub-corpus of the data. It might point out that Iranian authors, whose works were selected for the study, are less proficient in using different word types and members of the same word family, the more we move toward the less frequent word levels. Note that our generalization only applies to similar samples under investigation in the field of psychology, but not other fields/disciplines.

Thus, this investigation produces evidence, suggesting that a general list of academic words is limited, but instead gives a good return for learning if we take into adequate and economical account the coverage they provide (8–12%, based on the corpus-based studies cited above, and over 15% of the running words in our study). This argument and emphasis on the general aspect of the list is against the position taken by Hyland and Tse (2007) and Martínez et al. (2009) who recommend a discipline-specific list of academic words. Note that it is more economical for an EAP learner to attempt learning a limited number of academic words that are encountered in many disciplines than a list of the academic words more commonly used in a specific discipline.

With regard to the point above, there is implication for both reading and writing; Considering the verification of the coverage by Chen and Ge (2007), Li and Qian (2010), Vongpumivitch et al (2009), and other studies surveyed, academic vocabulary comprises between 8–12 percent of the tokens in any text. Note also that the results from our own small-scale study were even considerably higher. Thus, language learners will practically be better off in reading many, if not all, texts of their related fields through only learning a limited number of words within a short time. At the same time, as Hirsh (2010) illustrates, these very words will serve as effective signposts for EFL learners to organize their concepts efficiently in writing for academic purposes. The list of academic words will actually develop cohesion and coherence within the texts the learners write for their academic purposes. That is, academic words will help develop linkage between sentences and content elements both at intra- and inter-sentence level as well as between paragraphs. Consequently, in reading and writing scholarly papers, the academics in different departments in EAP contexts might benefit from a general list of academic vocabulary, compiled from a very representative corpus across the respective fields and disciplines, including psychology.

In addition, the most frequently- and immediately-consulted sections of a scholarly journal article, especially in an EAP context, are the paper title and the *abstract*. Logically, the more we are able to detect academic words with a high percentage of coverage in the abstracts, the better we serve the readers and researchers in EAP contexts. In most foreign language contexts, journal articles publishing in languages other than English require an English abstract from the contributors who submit articles to be considered for publication. Given that, providing a list with a high coverage of the running words in journal abstracts is theoretically and pedagogically of utmost importance.

It is, therefore, a favor to EAP learners to provide them with a list containing a well-researched and limited number of academic words, with few shortcomings, if such a list is developed in future. Until then, however, this paper takes sides with adopting a general list of academic vocabulary, mainly the AWL, though it humbly acknowledges the shortcomings that some researchers have raised against it. Yet, consider that the existence of objections or alternative positions that emerge with different objectives and different conceptualizations of vocabulary only adds to the richness of the arguments on vocabulary. That being acknowledged, the present researchers, however, suggest that an efficient list would work better than an ideal one for EAP instructors. The language instructors are under the pressure of many factors, especially economic ones, to design such an efficient list for their own purposes. Furthermore, aside from EAP instructors, the researchers confirm and encourage the use of such a list by the university students at undergraduate and graduate levels, novice student researchers, as well as the researchers with a need to read or publish, especially abstracts in English, either in the journals published in their home countries or in Anglo-American ones.

As shown in the current and previous research, there are pros and cons associated with the AWL. Therefore, we also believe that EAP practitioners can be eclectic in the sense that they can use both discipline-specific and general academic word lists at the same time. It will depend on the disciplinary homogeneity of their students: do they come from one discipline only, from several unrelated disciplines, from several related disciplines, and so on?

All in all, it is of great use to have an academic word list. It might explain why some researchers find value and usefulness in attempting to develop a list of academic vocabulary in languages other than English. An example is Cobb and Horst (2004) who have developed an academic word list in French. The list, however, differs from the AWL (Coxhead, 2000) in that the 2,000 most frequent French words in Cobb and Horst's list serve both every day and academic purposes.

Irrespective of whether language instructors use discipline-specific or general academic word lists, we also suggest extensive reading to boost the process of vocabulary instruction as academic vocabulary, according to Krashen (2013), is a late-acquired aspect of language among others. Krashen even states that some aspects of language, including academic vocabulary, "will be acquired if the student gets more comprehensible input. If it is an aspect of academic language, it will be acquired by reading" (p. 28).

Language practitioners, therefore, need to reconsider their practices in instructing English vocabulary. Currently, some volumes (e.g., Gardner, 2013) systematically respond to this need (Akbarian, 2015). Future research will approve the appropriate intensive and extensive activities for that purpose.

References

Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 39–49). Amsterdam, Netherlands: John Benjamins Publishing Compancy.

Akbarian, I. (2015). Review of Describing and explaining grammar and vocabulary in ELT. *Iranian Journal of Language Teaching Research*, *3*(2), 123–127.

American Psychological, Association. (2009). *Publication manual of the American psychological association*: American Psychological Association (APA).

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. T. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77-117). Newark, DE: International Reading Association.

Baker, M. (1988). Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles. *Reading in a Foreign Language, 4*(2), 91-105.

Chen, Q., & Ge, G. C. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes, 26*, 502-514.

Cobb, T., & Horst, M. (2004). Is there room for an academic word list in French. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15-38). Amsterdam, Netherlands: John Benjamins Publishing Company.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly, 34*, 213-238.

Coxhead, A. (2012). Academic vocabulary, writing and English for Academic Purposes: Perspectives from second language learners. *RELC Journal, 43*(1), 137-145. doi: 10.1177/0033688212439323

Evans, S., & Green, C. (2007). Why EAP is necessary: A survey of Hong Kong tertiary students. *Journal of English for Academic Purposes*, *6*(1), 3–17. doi.org/10.1016/j.jeap.2006.11.005

Gardner, D. (2013). *Exploring vocabulary: Language in action*. London, England: Routledge.

Hancıoğlu, N. (2009). *Incorporating corpus data into an advanced academic: Thesis writing course* (Doctoral dissertation). Eastern Mediterranean University. Cyprus. Retrieved from http://www.scribd.com/doc/14596786/Write-Like-an-Academic-the-Thesis

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). RANGE [Computer software]. Wellington, New Zealand: Victoria University of Wellington. Retrieved from http://www.victoria.ac.nz/lals/staff/ paul-nation/nation.aspx.

Hirsh, D. (2010). *Academic vocabulary in context*. Bern: Peter Lang.

Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly, 41*, 235-253.

Krashen. S. (2013). Reading and vocabulary acquisition: Supporting evidence and some objections. *Iranian Journal of Language Teaching Research*, *1*(1), 27–43.

Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System, 38*, 402-411. doi: 10.1016/j.system.2010.06.015

Liu, D. (2014). *Describing and explaining grammar and vocaburlary in ELT: Key concepts and effective practices*. New York: Routledge.

Martínez, I., Beck, S., & Panza, C. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*, 183-198. doi: 10.1016/j.esp.2009.04.003

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes, 25*, 235-256.

Neufeld, S., Hancioğlu, N., & Eldridge, J. (2011). Beware the range in RANGE, and the academic in AWL. *System, 39*, 533-538.

Perry, F. L. (2011). *Research in applied linguistics: Becoming a discerning consumer*. New York: Taylor & Francis.

Santos, M. G. (2002). Examining the vocabulary skills of language minority students in community college study context. TESOL 2002 Salt Lake City, UT-April 11, 2002. #1957: Supporting the ESL student beyond ESL programs (HE).

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

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, England: Oxford University Press.

Thurstun, J., & Candlin, C.N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes, 17*, 267-280.

Vongpumivitch, V., Huang, J., & Chang, Y.. (2009). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes, 28*, 33-41. doi: 10.1016/j.esp.2008.08.003

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a Medical Academic Word List. *English for Specific Purposes, 27*, 442-458. doi: 10.1016/j.esp.2008.05.003

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

Wood, A. (2001). International scientific English: The language of research scientists around the world. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 71–83). Cambridge, England: Cambridge University Press.

Acknowledgement

Special thanks are due to *IJLTR* anonymous reviewers for their constructive comments on the first version as well as the revision of our manuscript. This study was supported by the University of Qom (Registered No. 3340KP).

Is'haaq Akbarian, PhD, is a faculty member of the University of Qom, Qom, Iran. He has published papers in a number of local and international scholarly journals, including *System*, *English for Specific Purposes*, *Journal of Asia TEFL*, and *Iranian Journal of Applied Linguistics (IJAL)*.

Zahra Ghanbarzadeh has BA in English Literature from the University of Qom, Iran. She is a freelance translator and teaches in language institutes in Qom.

Mohammad Afzali Shahri, a member of the University of Qom, Iran, received his MPhil in TEFL from the University of Isfahan. He is mostly interested in cultural issues, teacher training, materials development, and teaching skills. Consequently, He attempts to apply his findings in these regards to his teaching career and textbooks.

Appendix

Coxhead's (2000) AWL sub-lists and the 100 most frequently-occurring AWL word forms in each sub-corpus of the PAC

Iranian data:

1. ***analysis***/analyzed, ***approach***, areas, assess/*assessed*/**assessment**, ***concept***, consisted/consistency, creativity, ***data***, factor/***factors***, *identity*, ***indicate***/*indicated*, ***individual***/*individuals*, **major**, **method**/methods, **percent**, **period**, ***process***, ***research***, ***role***, ***significant***/*significantly*, **structure**, ***theory***, variable/**variables**/variance

2. ***achievement***, ***aspects***, **computer**, **conclusion**, conducted, **design**, evaluate/evaluated/**evaluation**, **normal**, **obtained**, *participants*, ***perceived***, ***positive***/*positively*, selected, ***strategies***, **survey**

3. **components**, **criteria**, **items**, ***negative***/negatively, ***physical***, reliability, **sex**, **validity**

4. *attitude*, **communication**, **dimensions**, *goal*/***goals***, ***implications***, **internal**, investigate/*investigated*, **job**, predict/***predicted***, statistical, ***status***, **stress**

5. ***academic***, **adjustment**, ***affect***, **medical**, ***mental***, ***orientation***, *psychological*, style/**styles**, **version**

6. ***assigned***, attachment, ***gender***, **index**, **intelligence**, ***motivation***, ***revealed***

7. *adult*, couples, ***grade***, **intervention**

8. **random**/randomly

9.

10. **depression**

Anglo-American data:

1. analyses/***analysis***, ***approach***, *assessed*, ***concept***, **consistent**, **context**, ***data***, **evidence**, ***factors***, *identity*, ***indicate***/*indicated*, ***individual***/*individuals*, ***process***/processes/processing, ***research***, **response**/responses, ***role***, ***significant***/*significantly*, **similar**, **specific**/specifically, theoretical/theories/***theory***

2. ***achievement***, ***aspects***, **complex**, **cultural**, **features**, **focus**, **impact**, *participants*, ***perceived***/perception/perceptions, ***positive***/*positively*, **potential**, **previous**, strategy/***strategies***

3. **demonstrate**/demonstrated, **framework**, **initial**, **interaction**/interactions, ***negative***, **outcomes**, **partner**, ***physical***, **task**/tasks

4. *attitudes*, **contrast**, *goal*/***goals***, **hypothesis**/hypothesized, ***implications***, *investigated*, **mechanisms**, ***predicted***, **prior**, ***status***

5. ***academic***, ***affect***, awareness, **conflict**, ***mental***, ***orientation***, **perspective**, *psychological*/**psychology**, **target**/targets, **whereas**

6. **abstract**, accuracy, ***assigned***, **discrimination**, **furthermore**, ***gender***, ***motivation***, ***revealed***

7. ***adults***, ***grade***, **inferences**, **paradigm**,

8. **bias**, **implicit**, **manipulated**

9. mediated, **visual**

10.

Note: The most frequent family members in AWL (Coxhead, 2000) appear in bold and those shared in both Iranian and Anglo-American sub-corpora of the data are in italics. However, the words in bold and italics have both of the characteristics above.