

Towards the establishment of an academic word list (AWL) for the abstract section of research articles

¹Nezam Zarei Chamani, ²Abdolreza Pazhakh, ³Alireza Jalilifar, ⁴Bahman Gorjian*

¹Department of TEFL, Science and Research Branch, Islamic Azad University, Khuzestan, Iran

²Department of TEFL, Dezful Branch, Islamic Azad University, Dezful, Iran

³Department of TEFL, Shahid Chamran University of Ahvaz, Iran

⁴Department of TEFL, Abadan Branch, Islamic Azad University, Abadan, Iran

*Corresponding author: bahgorji@yahoo.com

Abstract-This study is a corpus-based lexical study that aims to explore the most frequently-used academic words in the abstract section of RAs in applied linguistics. The corpus consists of 2000 abstracts from ten scientific and prestigious journals, namely, *applied linguistics*, *language learning*, *ELT journal*, *journal of pragmatics*, *system*, *journal of English for Specific Purposes*, *journal of English for Academic Purposes*, *studies in second language acquisition*, *TESOL Quarterly*, *Modern language journal* which have been published between years 2000 and 2011. Using frequency and range as the criteria for word form selection, this study identified 600 AWL word forms that are used more than 14 times in the entire corpus and at least 5 times across the ten journals in the corpus. Results show that AWL words account for 15.35% of the entire corpus. Some of the AWL word forms in this study overlapped with word families included in Coxhead's AWL.

Key words- Academic Word; Applied Linguistics; Research Articles; Abstract Section

1. INTRODUCTION

Research article (RA), as the main channel for presenting scientific findings to researchers all over the world, has received much attention (Bhatia, 2002; Hirano, 2009; Nwogu, 1997; Swales, 1990; Jalilifar, 2007, to name but a few). This scientific genre is the most influential and highly esteemed communicative vehicle in the discourse community of scientific experts (Swales, 1990). Therefore, it is necessary for scientific experts and researchers who want to be the member of discourse community to be able to write acceptable RAs. As Duszak (1997) has noted, "ignorance of, or misconceptions about, the communication styles of others can hinder understanding among academics and ultimately obstruct cooperation and advancement of scholarship". (p.3). Because of the significance of RAs and of English as an international language, many scientists and researchers like to publish their research results in English in international scientific journals. Writing academic discourse plays a significant role in producing and distributing knowledge among people and across borders imposed by differences of language and culture (Jogthong, 2001).

2. REVIEW OF LITERATURE

Vocabulary learning is central to language acquisition, whether the language is first, second, or foreign. Possessing a rich treasury of words will have a direct influence upon the accuracy and quality of students writing and their reading comprehension (Nagy, 1988; Nation, 2001). A wealth of research has documented the

strength of the relationship between vocabulary (Gorjian, Pazhakh & Naghizadeh, 2012) and comprehension. Bronson (2001) argues that limited knowledge of vocabulary makes the acquisition of new knowledge difficult for the learners, especially non-natives. Learners need to have a vocabulary close to 8,000 word families to be able to perform certain tasks such as read a novel, read newspaper, watch movies, and listen to family conversations (Nation, 2001). However, selection of those worthy vocabulary items to be taught is not a simple matter. Nation's (2001) division of vocabulary into "four levels of high-frequency words, academic vocabulary, technical vocabulary, and low frequency words" indicates that some words are more important in various phases of language or for different purposes. Among these, Academic words, falling somewhere between the high-frequency words and technical words, have some important functions. They account for a relatively high proportion of running words in academic texts. An extensive elaboration on academic vocabulary will be presented as follows.

2.1. Academic Vocabulary

The main worth of academic vocabulary lies in the role that these words play in the construction of the scientific experience. Kuehn (1996) observed that knowledge of academic words distinguished academically well-prepared from under-prepared college students from all backgrounds. The findings from many studies (e.g. Coxhead, 2000) indicate that EAP learners with insufficient academic vocabulary knowledge cannot effectively deal with reading materials for various types

of academic tasks they are supposed to fulfill. To solve this problem and also heighten the learners' awareness of academic vocabularies, some academic word lists have been collected and proposed by various researchers. These word lists are discussed below (Gorjian, Pazhakh & Parang, 2012).

2.2. Academic word lists

Jalilifar (2007) notes that "a well-structured abstract is one which states all the four components, which are fundamental and obligatory, that is, introduction, method, results, and conclusion." Therefore, writing a useful abstract needs proficiency of the textual and rhetorical features that form an effective abstract. Additionally, abstract is the first section of a RA. A well-documented abstract with the use of appropriate vocabularies can satisfy the readers to pursue the reading of the article (Zhang, Wu, Wei & Wang, 2011).

3. METHODOLOGY

3.1. The corpus

In a descriptive study like the present study, the larger the corpus the more reliable it would be. After consultation with professors, about 2000 abstracts of the prestigious journals in applied linguistics, with 400.310 running words, were selected as the corpus of the study (Table 1). The abstracts have been selected from ten scientific and prestigious journals, which have been published in years 2000 - 2011. These journals are: 1) *Applied linguistics*, 2) *ELT Journal*, 3) *Language learning*, 4) *Journal of Pragmatics*, 5) *modern language Journal*, 6) *Studies in second language acquisition*, 7) *English for Specific Purposes*, 8) *System* 9) *TESOL Quarterly*, 10) *English for Academic Purposes*. These journals are assumed to be internationally renowned and cover a wide range of topics in the field of second language teaching and learning.

This study only comprises abstract sections as the major characteristics of research corpus are to be categorized in a manageable way. Some characteristics of the research corpus should be mentioned here: (a) each article of the corpus presents a report of an empirical research (b) the corpus articles are assumed to be written by both native and non-native speakers of English (c) the present study has been focused only on the abstract section of applied linguistics articles; thus, the corpus only comprises abstract sections.

Table 1. Composition of research corpus

Journals	Number of articles
Applied Linguistics	200
ELT Journal	200
English for specific Purposes	200
Journal of Pragmatics	200
Modern Language Journal	200
Language Learning	200
Studies in Second Language Acquisition	200
System	200
TESOL Quarterly	200
English for Academic Purposes	200

3.2. Data collection procedure

In order to make this study manageable, we focused on ten journals published by Thomson Reuters and rated by impact factor. For selecting the journals, the researcher consulted with professors and experienced applied linguistics professionals. All the abstracts of RAs published in the focused journals in years 2000–2011 were selected as the corpus of the study. The articles that the researcher picked are all written in English and followed any typical format. To determine the frequency and range of the word types, Range program available as a free downloadable zip file at <http://www.vuw.ac.nz/lals/staff/paul-nation/nation.aspx> was taken. This program help the teachers identify academic words in texts of different size. Coxhead and Byrd (2007) note that teachers can use this tool to find out whether the words in the text occur in the first (one in the program) and the second (two in the program) 2000 words (GSL), in the Coxhead's AWL (third wordlist in the program) or not in any list. By using range and frequency information, teachers know which words need to be addressed first in class (Coxhead & Byrd, 2007).

Since in this study word types rather than word families were under the focus, after collecting the data from the Range program, those word types which reached the criteria were picked on as the academic words and ranked in the order of frequency. For the standardization of the applied linguistics RAs included in the corpus, the charts, diagrams, bibliographies and other components in texts, which were not able to be processed by computer analyzing programs (Zhang & Wu, 2011 b) or should not be included in the lexical analysis in the chosen applied linguistics RAs, were removed so as to eliminate the factors unrelated to the lexical analysis and to ensure that the texts stored in the corpus be readable by computer software (Zhang & Wu, 2011 a).

3.3. Data analysis procedure

The study aimed to investigate the frequency and distribution of the AWL word forms that are used in the abstract section of the RAs in Applied Linguistics. To answer the research questions, the first stage was to obtain a list of AWL word form that meets the criteria set in this study. The frequency criterion was that the word

forms had to occur at least fourteen times in the entire corpus. The rationale to select these rates of occurrence is based on the Coxhead's (2000) selection of AWL words. Coxhead's corpus for the AWL was around 3.5 million words and her selection criteria for the AWL word forms were that each word form in the AWL should occur at least 100 times in the entire corpus and at least 10 times in each of the four disciplines. The obtained frequencies will show what vocabularies used in abstract section is considered as the academic word. The collection of these words will establish an AWL for the abstract section of RAs in applied linguistics.

3.4. Word selection criteria

Coxhead (2000) also reported that in her AWL word selection, range was the first criterion and frequency the second because a word count based mainly on the frequency would have been biased by longer texts and topic related words. This principle was also used in the current study. Only word families covering 5 journals would be included in the AWL, while word families occurring with very high frequency but covering fewer than 5 journals would be excluded. In sum, all the finally included word families in the AWL met the following word selection criteria:

1. Specialized occurrence: the word families included had to be outside the first 2000 most frequently occurring words of English, as represented by West's GSL (1953).

2. Range: Members of a word family had to occur at least in 5 journals.
3. Frequency: Members of a word family had to occur at least 14 times in the corpus of applied linguistics research articles.

4. RESULTS

To examine the frequency and distribution of AWL word families, a corpus of 400,310 words from the abstract section of ten journals has been employed. Totally 2282 word families were found to have occurred 14 times (frequency). After the elimination of the GSL word families (1112 word families), 1170 word families were left and 600 word families occurred in 5 or more journals (subject areas) under study (range). By our word selection criteria plus the expert opinion of our consulted experienced applied linguistics professors, these 600 word families were ultimately chosen and formed the AWL of the abstract section of research articles in applied linguistics, which appeared 61,462 times totally. As it is shown in Table 1, AWL account for 15.35% of the entire corpus which is higher than the 11.7% of the AWL reported in Vongpumivitch, Huang, and Chang (2009) study in Applied Linguistics research articles. This high coverage (15.35%) indicates that AWL plays an important role in the abstract section of RAs in applied linguistics.

Table 2. Frequency and coverage of AWL words in the RAs abstracts

No. of words in the corpus	400310
Frequency of AWL	61,462
Percentage of AWL coverage	15.35%

In the abstract section of RAs, the most frequently used word was *research*, which appeared 1241 times and appeared in all ten journals in the corpus, while the least frequently used one was *intake*, which appeared 14 times and appeared in 5 journals in the corpus. Table 3 shows the statistical results of the top 30 most frequently used word families in the abstract section of research articles in applied linguistics. Total frequency of top 30 academic word families was 11,920. These words occurred in a wide range across the journals in the study. 29 out of 30 top word families covered 10 journals under the study.

They account for about 18.40% of the academic word types in the corpus. Interestingly, 23 out of 30 top words and 288 Out of 600 academic word types of the corpus were in Coxhead's AWL. The word families in the abstract section of RAs occurred in a wide range of the subject areas in our corpus. Of the 600 word families in the list, 214 (35.66%) occurred in all the 10 journals and 93 (15.05%) occurred in 9 journals, and 82 (13.66%) occurred in 8 journals, 79 (13.16%) occurred in 7 journals, 67(11.16%) occurred in 6 journals, and 65(10.83%) occurred in 5 journals.

Table 3. Top 30 most frequently used word types in the corpus

Rank	Word Types	R.F	%	R	Rank	Word Types	R.F	%	R
1.	Research	1241	2.01	10	16.	Role	356	0.57	10
2.	Result	728	1.18	10	17.	Interaction	344	0.55	10
3.	Analysis	722	1.17	10	18.	Instruction	341	0.55	10
4.	Data	610	0.99	10	19.	Vocabulary	329	0.53	10
5.	Academic	603	0.98	9	20.	Lexical	316	0.51	10
6.	Discourse	584	0.95	10	21.	Proficiency	300	0.48	10
7.	Classroom	525	0.85	10	22.	Communication	286	0.46	10
8.	Context	447	0.72	10	23.	Process	270	0.43	10

9.	Strategy	445	0.72	10	24.	Comprehension	270	0.43	10
10.	Acquisition	437	0.71	10	25.	Corpus	266	0.43	10
11.	Linguistic	417	0.67	10	26.	Education	265	0.43	10
12.	Feature	392	0.63	10	27.	Text	265	0.43	10
13.	Participant	377	0.61	10	28.	Information	257	0.41	10
14.	Approach	375	0.61	10	29.	Implication	257	0.41	10
15.	Task	361	0.58	10	30.	Focus	256	0.41	10

Note. R.F: raw frequency, %: of the overall corpus, R: range

After counting the AWL in the corpus, about 132 AWs did not receive the criteria for being among the AWL word families; therefore, they have eliminated from the list. The words such as *legislate* and *commission* which are placed among the most frequent words in Coxhead's word list did not place among the AWs in this corpus. The reason for this event can be related to the difference between the selected corpora for two studies. Although Coxhead's AWL represented a high degree of words used in the entire corpus in this study, the researcher, after consultations with his instructor and MA classmates, reached to this claim that some of the words which are frequently used by researchers in abstract section of RAs have not been included in the Coxhead's AWL. Therefore, lists of all the words that are supposed to be academic words have been prepared (Gorjian, Alipour & Saffarian, 2012). These words, based on the index proposed by Coxhead (2000), have been examined, those which received the criteria added to the AWL and those which did not receive the criteria have been omitted from the list. These new academic words which are not in Coxhead's AWL are highlighted in the Appendix.

The following passage randomly selected from an Applied Linguistics RAs abstract in our corpus can give us a picture of the academic words used and distributed in such texts. The academic words are in boldface.

This study **assessed** the **impact** of completing English for **Academic** Purposes (EAP) course on the writing of **postgraduate** learners. We begin this paper by describing a course offered for credit to **postgraduate international** students at a university in Australia, and then report on a large-scale study ($n=69$) which **investigated** the **improvement** (if any) in the writing of students **enrolled** in this course. The students **undertook** the course at the beginning of their **postgraduate** studies at the university and concurrently with their degree **programs**. Two cohorts were **identified**: students with an IELTS **score** on entry to the University of 6.0 (or **equivalent**) and those with a score of 6.5. **Data** for the study were the in-class writing produced by the students at the beginning and at the end of **semester**, and a short **questionnaire**. Among the 196 words in the above passage, 43 words belonged to the academic word types of the abstract section in applied linguistics RAs. The coverage of academic words in this text was 21.93% that a bit consistent with the results of the study.

5. DISCUSSION AND CONCLUSION

The comparison of the most frequent items in this corpus with the findings of Chen and Ge (2007) revealed low coincidence, which further reinforce the idea that differences are based on specificity. When the researcher

compared the 20 most frequent words in Chen and Ge's corpus, and those in this corpus, the researcher found a coincidence of only two items, analysis and data. The researchers also compared the 20 most frequent words in this corpus with the 20 most frequent words in the Agro corpus of Martinez, Beck & Panza (2009). Again, there was low coincidence (Zhang, Wang, & Wu, Huo, 2011). The result(s) of this study can contribute to those MA and PhD students who wish to publish in international journals. The result(s) of this study can help the researchers use these academic words in writing the abstract section of RAs. This study can help the researchers who have difficulty in writing the abstract section of RAs; it can also augment undergraduate students' awareness of academic words and this familiarization helps them to write effective essays. More importantly, findings of this study have created an AWL for the abstract section of RAs, which is important for non-native writers in the field of EFL.

REFERENCES

- [1] Bhatia, V. K. Applied genre analysis: Analytical advances and pedagogical procedures. In A.M. Johns (Ed.), *Genre in the classroom* Multiple perspectives (pp. 279-283). New Jersey: Lawrence Erlbaum, Mahway. (2002).
- [2] Bronson, M. C. *Genre is a verb: Research on academic writing in critical perspective: An unpublished qualifying paper*, retrieved 10 February from [http:// www. Enformy. Com/Bronson Qpup. Html](http://www.Enformy.Com/BronsonQpup.Html). (2001).
- [3] Chen, Q., & Ge, G. C. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26, (2007), 502-514.
- [4] Coxhead, A. A new academic word list. *TESOL Quarterly*, 34 (2), (2000), 213-238.
- [5] Coxhead, A., & Nation, P. The specialized vocabulary of English for academic purpose. In J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252-267). Cambridge: Cambridge University Press. (2001).
- [6] Coxhead, A., & Byrd, P. Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing*, 16, (2007), 129-147.
- [7] Duszak, A. Cross-cultural academic communication: A Discourse community view. In A. Duszak (Ed.), *Culture and Style of Academic Discourse* (pp. 11-39). New York: Mouton de Gruyter. (1997).
- [8] Gorjian, B., Pazhakh, A. R., & Parang, K. An investigation on the effect of critical thinking (CT) instructions on Iranian EFL learners' descriptive writing: A case of gender study. *Advances in Asian Social Science*, 1(1), (2012), 114-118.

- [9] Gorjian, B., Pazhakh, A. R., & Naghizadeh, M. Comparative study of conjunctive adverbials (CAs) in native researchers' (NRs) and non-native researchers' (NNRs) experimental articles. *Advances in Asian Social Science*, 1(2), (2012), 224-247.
- [10] Gorjian, B., Alipour, M., & Saffarian, R. The effect of multisensory techniques on reading comprehension among pre-intermediate EFL learners: The case of gender. *Advances in Asian Social Science*, 1(2), (2012), 192-196.
- [11] Hirano, E. RA introductions in English for specific purposes: A comparison between Brazilian Portuguese and English. *English for Specific Purposes*, 28, (2009), 240-250.
- [12] Hirsh, D. *The vocabulary demands and vocabulary learning opportunities in short Novels*. Unpublished master's thesis, Victoria University of Wellington, New Zealand. (1993).
- [13] Jalilifar, A. R. (Hedging as a pragmatic strategy: Variation across cultures. *TELL*, 1(3), 2007), 43-69.
- [14] Jogthong, C. *Research Article Introductions in Thai: Genre Analysis of Academic Writing*. An MA thesis, Morgantown, West Virginia. (2001).
- [15] Kuehn, P. *Assessment of academic literacy skills: Preparing minority and limited English Proficient (LEP) students for post secondary education*. Frenso, CA: California State University (ERIC Document Reproduction Service No. ED 415489). (1996).
- [16] Martinez, I. A., Beck, S. C., & Panza, C. B. Academic vocabulary in Agricultural research articles: A corpus-based study. *English for Specific Purposes*, 28, (2009), 183-198.
- [17] Nagy, W. *Teaching vocabulary to improve reading comprehension*. Newark, DE: International Reading Association. (1988).
- [18] Nation, P. *Learning vocabulary in another language*. Cambridge: Cambridge University Press. (2001).
- [19] Nwogu, K. N. The medical research articles: Structure and function. *English for Specific Purposes*, 16 (2), (1997), 119-138.
- [20] Swales, J. M. *Genre analysis: English in academic and research setting*. Cambridge: Cambridge University Press. (1990).
- [21] Tahririan, M. H., & Jalilifar, A. R. Generic analysis of theses and dissertation abstracts: variation across cultures. *IJAL*, 7(2), (2004). 121-143.
- [22] Vongpumivitch, V., Huang, J., & Chang, Y. Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in the applied linguistics research articles. *English for Specific Purposes*, 28, (2009), 33-41.
- [23] West, M. (1953). *A general service list of English words*. London: Longman, Green and Co.
- [24] Zhang, Y, Wang, L.S, & Wu, L, Huo, Y. Multi-channel diffusion tensor image registration via adaptive chaotic PSO. *Journal of Computers*, 6 (4), (2011), pp. 825-829.
- [25] Zhang, Y, Wu, L, Wei, G, & Wang, S, L. A novel algorithm for all pairs shortest path problem based on matrix multiplication and pulse coupled neural network. *Digital Signal Processing*, 21 (4), (2011), pp. 517-521.
- [26] Zhang, Y & Wu, L. Crop Classification by forward neural network with adaptive chaotic particle swarm optimization. *Sensors*, 11 (5), (2011 a), pp. 4721-4743.
- [27] Zhang, Y & Wu, L. A novel algorithm for APSP problem via a simplified delay pulse coupled neural network. *Journal of Computational Information Systems*, 7 (3), (2011 b), pp. 737-744.



Bahman Gorjian obtained his PhD in TEFL from Research and Science Center of Khurasgan University in Isfahan Province. In 2006, he was appointed Assistant Professor at TEFL Department, Abadan Branch, Islamic Azad University, Abadan, Khuzestan Province, Iran. He is currently working within several research areas covering TEFL, applied linguistics, psycholinguistics, individual differences and teaching and testing EFL. He can be reached at: bahgorji@yahoo.com

Cell phone: 00989161310917