International
Journal of Society, Culture & Language
IJSCL

Journal homepage: www.ijscl.net
ISSN 2329-2210 (online)

Developing a Corpus-Based Word List in Pharmacy Research Articles: A Focus on Academic Culture

Farrokh Heidari[1a], Alireza Jalilifar[2a], Anayatollah Salimi[3b]

ARTICLE HISTORY:

Received December 2019
Received in revised form February 2020
Accepted February 2020
Available online March 2020

KEYWORDS:

Academic vocabulary
Pharmacy
PAWL
PRAC
Research article

Abstract

The present corpus-based lexical study reports the development of a Pharmacy Academic Word List (PAWL); a list of the most frequent words from a corpus of 3,458,445 tokens made up of 800 most recent pharmacy texts including research articles, review articles, and short communications in four sub-disciplines of pharmacy. WordSmith (Scott, 2017) and AntWordProfiler (Anthony, 2014) were used to screen words based on frequency, range, dispersion, and specialized occurrence. The developed PAWL contains 750 word families covering 17.69% of the corpus under study. The findings of the current investigation confirm the necessity to compile domain-specific academic word lists to address the needs of non-native researchers and postgraduate students over various disciplines. Such a word list can function as a reference for an EAP lexical syllabus. Pedagogical implications are made for pharmacy researchers, postgraduate students, and material designers, who can use PAWL as a lexical repertoire to set their vocabulary learning/teaching goals.

[1] PhD Candidate, Email: farrokh.heidari@gmail.com
[2] Professor, Email: ar.jalilifar@gmail.com (Corresponding Author)
Tel: +98- 0611-3367008
[3] Associate Professor, Email: anayatsalimi2003@yahoo.com
[a] Shahid Chamran University of Ahvaz, Iran
[b] Ahvaz Jundishapur University of Medical Sciences, Iran

1. Introduction

Researchers all around the world are required to publish to improve their tenure, as their academic success is tied to their success in publishing in English as "the global academic lingua franca" (McKinley & Rose, 2018, p. 1). Similarly, postgraduate students need to present their research through dissertation and finally deliver their investigation by publishing it in English. However, a huge number of scholars and postgraduate students throughout the world are not native speakers of English. This presents a colossal task for the researchers and students alike, since they live, study, and communicate in a non-English context, but are required to present and publish their research in English. Two of the writers of the current paper, for example, have taught at the pharmacy department of a medical school, and have witnessed numerous colleagues and postgraduate students at that department suffering from their insufficient English proficiency to present their research in a language that adheres to the norms of an English academic journal. This immediate need motivated the authors to develop a list of the most frequent pharmacy academic vocabulary that could directly promote fluency and accuracy of non-native researchers and postgraduate students in their academic English reading and writing.

Of all genres within academic writing, it is the research article (RA) that has attracted the most attention as the primary channel for presenting claims of new knowledge (Hewings, 2001, p. 12). Moreover, despite competing electronic publishing alternatives such as websites, the research article continues to be "the pre-eminent genre of the academy" and "the principal site of disciplinary knowledge-making" (Hyland, 2009, p. 67) and what Montgomery (1996, p. 2) describes as the "master narrative" of our time. One reason for this prominence can be the worth given to the highly regarded peer-review procedure as a controlling and regulating system for transforming thoughts into knowledge (Hyland, 2009, p. 68), as authors endeavor to have their arguments become part of the disciplinary consensus where new findings are contextualized within the framework of the previously published literature which efficiently displays current status of the existing consensus (Hewings, 2001, p. 12). Another reason is the prestige ascribed to a genre which creates a means for a scientific generation of facts as more academics around the world are required to publish in high-impact, peer-reviewed journals as a prerequisite for promotion and career development (Hyland, 2009, p.67), a point also acknowledged by other scholars (e.g., Hartley, 2008; Swales, 1990). Similarly, within the field of medicine, and pharmacy, in particular, journal publication in all its various forms (e.g., *research papers*, *review articles*, *comments*, *editorials*, *short communications*, *technical notes*, *case reports*, *and letters*) is regarded as the main criterion against which career development and promotion are judged.

Furthermore, the increasing number of postgraduate non-native English students who are required to read and publish academic articles in English has motivated much research, typically corpus-based research, to find distinctive characteristics of academic genres (Coxhead, 2000; Valipouri & Nassaji, 2013). Many of these studies use electronic corpora as a valid source of information. These corpora are then scrutinized to discover particular patterns of language use, such as specialized or academic vocabulary. The high frequency of academic vocabulary in academic texts is perhaps the reason why it is assumed to be essential to learn for EAP (English for Academic Purposes) learners (Coxhead, 2000; Nation, 2001). Additionally, academic vocabulary acquisition is an essential yardstick of academic proficiency (Kuehn, 1996). Likewise, academic vocabulary knowledge is crucial for effective academic reading, and more importantly, for competent writing in specific academic fields (Corson, 1997). Despite being pivotal to English language learning and teaching for academic purposes, academic words have shown to be challenging to learners (Cobb & Horst, 2004; Shaw, 1991), either because this type of vocabulary does not appear as often as high-frequency words do in general texts (Xu & Nation, 1984), or because they are not taught or explained by the subject-matter instructor because they are not specific enough (Flowerdew, 1993; Nation, 2001). As a result, due to the perceived importance knowledge of vocabulary has on reading and writing skills of second language learners (Nation, 2001), several English academic word lists have been

constructed to assist the learning and teaching of academic vocabulary.

2. Theoretical Framework

The significance of academic vocabulary has motivated several studies to create word lists that represent the most common academic lexical items. There are two approaches to do this. One is the common core approach, adopted by pioneering researchers, which aims to provide discipline-crossing academic word lists appropriate for learners in various fields (Campion & Elley, 1971; Coxhead, 2000; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972), and the other is the early specialization approach which intends to create discipline-specific academic word lists based on independent domains suitable for the learners of that particular field. Based on the lists developed by Campion and Elley (1971), Ghadessy (1979), Lynn (1973), and Praninskas (1972), Xu and Nation (1984) produced a University Word List (UWL) which was extensively used for 15 years. However, in an attempt to construct a more robust and more representative academic word list, Coxhead (2000) introduced her Academic Word List (AWL) in a seminal article. The AWL was assembled from a corpus of 3.5 million tokens of written academic texts in 28 subject areas of four major fields of Law, Science, Commerce, and Art. The AWL includes 570 word families selected based on three yardsticks of *frequency* (occurrence of at least 100 times in the entire corpus), *range* (occurrence of at least 10 times in each of the four disciplines and in 15 or more subject areas), and *specialized occurrence* (items must be outside the first 2000 *General Service List* (GSL) words). The AWL (Coxhead, 2000) accounts for roughly 10.0% of all the running words in academic texts (Coxhead's corpus) but merely 1.4% of all the running words in a fiction corpus of the same size, which indicates that the list contains mainly academic words. The 10 percent coverage of the academic corpus, however, varies among the four disciplines: AWL has the highest coverage over commerce (12.0%) and the lowest over science (9.1%), with the other two being 9.3% (arts) and 9.4% (law).

Quite recently, however, some scholars (e.g., Hyland & Tse, 2007) have raised questions about the commonly believed assumption that there is a single common core academic vocabulary, and therefore have questioned the effectiveness of a domain-crossing academic vocabulary list, such as the AWL, according to research findings which show vocabulary use and behavior differ significantly across academic fields concerning frequency, range, meaning, and collocation. As a result, they contend that it is required to construct academic word lists independently for different academic domains. This has led to the prominence of the approach in which discipline-specific academic word lists are created based on independent domains suitable for the learners of that particular field. Informed by this necessity, several corpus-based studies have established word lists within the context of specific disciplines, e.g., engineering (Mudraya, 2006; Ward, 2009), chemistry (Valipouri & Nassaji, 2013), environmental science (Liu & Han, 2015), agriculture (Martinez, Beck, & Panza, 2009), nursing (Yang, 2015), and medicine (Wang, Liang, & Ge, 2008). In particular, Wang et al. (2008) created a Medical Academic Word List (MAWL) using corpora from 32 subject areas related to the field of medicine, representing high frequency words in their corpus which were not among the 2000 general English words as represented in West's (1953) GSL.

As shown by previous studies, a general discipline-crossing academic word list such as the AWL is by no means complete in expressing the most frequent academic words in a specific discipline, and a necessity to create more restricted discipline-based academic word lists is stressed because students in different disciplines have different lexical requirements. Furthermore, the AWL words tend to be more useful for expressing ideas and viewpoints (Coxhead & Nation, 2001), which is characteristic of social sciences, but less useful for emphasizing the description and depiction of results (Coxhead & Nation, 2001), which is characteristic of natural sciences. As a result, the necessity to construct a domain-specific academic word list within natural sciences is further stressed. As mentioned earlier, within natural sciences, a few academic word lists have been constructed; however, to the best of our knowledge, no list has yet been constructed in the discipline of pharmacy. It is interesting to note that even the MAWL has not taken into account any subfield of pharmacy. This is

because Wang et al. (2008) have adopted in their corpus the RAs from the 32 subject areas listed in the field of Medicine and Dentistry from *ScienceDirect Online*; whereas, subfields of pharmacy are listed under a separate discipline in *ScienceDirect Online*, namely *Pharmacology, Toxicology and Pharmaceutical Science*. As a result, the corpus used in Wang et al.'s study lacks pharmacy subject areas and does not include pharmacy-related journals or articles. This might bias the words selected to be included in the final list, as MAWL is not fully representative of medical vocabulary, since the field of pharmacy was omitted.

While, a pharmacy-based academic vocabulary list is yet to be developed, there has been somewhat related research in pharmacy (Grabowski, 2013; Grabowski, 2015). However, these studies have focused on keywords and lexical bundles, and not on academic vocabulary. Furthermore, a closer inspection of the composition of the corpora employed in Grabowski's studies reveals key differences between the results of those studies and those of our study. In Grabowski's (2013) research, 463 patient information leaflets, 146 summaries of product characteristics, 240 clinical trial protocols, and *only* 26 research articles on pharmacology have been used to compile a 2,478,992 million-word corpus. Also, Grabowski (2015) used 463 patient information leaflets, 136 summaries of product characteristics, 240 clinical trial protocols, and 86 chapters from pharmacy textbooks to amass a 2,230,161 million-word corpus. It is interesting to note that the number of research articles used in Grabowski' 2013 corpus is merely 26, and in his 2015 investigation is zero. Moreover, the majority of the corpora employed in these two studies include patient information leaflets, summaries of product characteristics, and clinical trial protocols. While analysis of such texts can provide valuable insights into linguistic variation in pharmacy, they are, as Grabowski himself acknowledged (2015, p.24), more pertinent to and beneficial for "pharmacists or pharmacy technicians" than researchers or postgraduate students. For researchers and postgraduate students who are required to present and publish their research, analysis of a corpus assembled from relevant research articles can be far more beneficial.

To address this need, the current study aims to establish the first pharmacy academic word list. Such a word list can function as a reference for an EAP lexical syllabus. Additionally, the findings of this study may be used as a valuable source for additional investigation into the creation of domain-specific academic word lists. Accordingly, we addressed the following research questions in the current study:

1. What are the most frequent academic words in a large corpus of pharmacy research articles that are not among the first 2000 words of English as represented in the GSL (West, 1953)?
2. Do the frequently occurring words in the corpus of pharmacy RAs also occur frequently in the AWL word list?
3. To what extent are the MAWL (Wang et al., 2008) words employed in the pharmacy RA corpus?
4. Are there any words that are not highly frequent in AWL and MAWL but appear with high frequency in the corpus of pharmacy research articles?
5. Are the words that appear with high frequency in the corpus of pharmacy RAs also identified as high frequency words in a general corpus such as BNC?

3. Methodology

A corpus-based approach for studying lexical items includes emphasis on the representativeness of text samples and the computational tools for exploring distributional patterns across discourse contexts (Biber, Conor & Upton, 2007, p. 3). In other words, priority is given to describing the most common uses of the most common words on the supposition that if a word is detected to occur frequently enough in the past, then it is probable to be important in the future as well (Hyland, 2006, pp. 75-76). This allows us to predict and generalize based on a representative sample. Therefore, the present study adopted a quantitative approach using computational tools for identifying the pharmacy-specific word lists.

3.1. Data Collection

All the texts used to compile the corpus employed in this study were downloaded from *ScienceDirect Online.* In its database, we selected the *pharmacy* domain, which consists of five subdomains: *Drug Discovery,*

Pharmaceutical Science, Pharmacology, Toxicology, and General. Altogether, 194 pharmacy journals were listed in the *ScienceDirect Online* database at the time of data collection, of which we decided to use all that were published in English in 2016 or afterwards, and 117 titles met that criterion. We downloaded the latest issue of all 117 pharmacy journals available and put together all the research articles (1323), review articles (227), and short communications (136), which amounted to 1686 texts. Following that, one of the co-authors of this study, who is an experienced pharmacy researcher, analyzed all the texts and reshuffled the articles into *four* subfields (*Drug Discovery*, *Pharmaceutical Science, Pharmacology, and Toxicology*) based on their abstracts and titles. Also, articles that pertained to more than one subfield were omitted so that the remaining texts could be confidently assigned to the related subfield. Finally, 200 texts were randomly selected to represent each of the four subfields, including 160 research articles, 25 review articles, and 15 short communications, representing their ratio in the initial 1686-text collection; 80% for research articles, 12.5% for review articles, and 7.5% for short communications. All the texts were published in 2016 and 2017 to take into account and reflect the latest academic vocabulary employed in pharmacy. The specialized corpus assembled for the present study contains 3,458,445 running words from 800 texts of pharmacy. To the best of our knowledge, this is one of the largest corpora employed in the establishment of English domain-specific academic wordlists.

3.2. Data Processing

In the present investigation, data processing consisted of standardization of the texts and normalization of the words in the texts. Text standardization included the removal of charts, diagrams, images, references, and basically everything that was either irrelevant to text analysis or could not be processed by the software. For normalization of words, WordSmith (Scott, 2017) and AntWordProfiler (Anthony, 2014) were employed to sort out words based on their frequency, range, and dispersion, and compare the coverage of our final wordlist with that of other benchmark lists. Following that, Familizer (Cobb, 2018b) was utilized to classify words as a family. A word family comprises a headword along with its inflected and derived forms (Nation, 2001, p. 11). This can benefit learners since knowing a headword facilitates the comprehension of its inflected and derived forms (Coxhead, 2000, p. 218).

3.3. Word Selection Criteria

Coxhead's (2000) principles (non-GSL words, range, and frequency) were employed in the current study with some adaptation. Similar to Coxhead, we decided to omit the GSL words as we were attempting to create an academic wordlist. In her study, Coxhead opted for selecting wide-ranging words that appear in at least half of the 28 subject areas. In our study, we decided to include words that appear in at least 3 of the 4 subfields available. This increased range criterion ensures that the candidate high frequency words are not stacked in one or two subfields and are more smoothly dispersed over the corpus. Following Coxhead's (2000) suggestion that range is a more important word selection factor than frequency, we decided to use the dispersion index available in the WordSmith software as an added measure. In this study, the dispersion was set at 40%. This number was achieved after several rounds of testing as a higher index would omit many useful academic words, and a lower index would include many technical words that were frequent in only one or two of the subfields. The minimum frequency of a word in our wordlist was 100, which is exactly similar to that of Coxhead's study, since our corpus (3,458,445) was almost exactly the same size as Coxhead's (3,500,000).

In sum, the following four criteria were employed to select the words to be included in the final wordlist:

1. *Specialized occurrence*: The included word families were required to be outside of the first 2000 most frequent words of English in the GSL (West, 1953).
2. *Range*: Word families had to appear in at least 3 of the 4 subfields.
3. *Dispersion*: Word families had to have a minimum 40% dispersion index in WordSmith.
4. *Frequency*: Members of a word family had to appear at least 100 times in the whole corpus.

3.4. Pharmacy Academic Word List (PAWL) Development

The frequency, range, and dispersion of the words in the corpus were calculated by a computer software. The word selection conditions were then implemented to identify our candidate words. Words that occurred in only one or two of the four subfields, had a lower-than-40% dispersion, or appeared in the GSL were removed. Moreover, the remaining words had to appear at least 100 times in the corpus. Finally, the criteria-fulfilling words were analyzed by one of the co-authors of this study, the experienced pharmacy researcher, to remove any word with technical meaning (research question 1). Using WordSmith (Scott, 2017), our domain-specific PAWL was compared with Coxhead's (2000) domain-crossing AWL to find out the words exclusive to our list and the words shared between the two (research question 2). We also aimed to discover to what extent MAWL (Wang et al., 2008) word-families were used in our corpus, which was examined using AntWordProfiler (Anthony, 2014) (research question 3). Moreover, using WordSmith, we found out which words were present in PAWL, but not in AWL and MAWL (research question 4). Finally, to check if our list was truly academic, it was necessary to compare it against a general corpus (research question 5). To do so, Coverage Calculator (Cobb, 2018a) was employed to examine PAWL coverage against BNC Sampler Written.

4. Results

Our Pharmacy Research Article Corpus (PRAC) consisted of 3,458,445 tokens and 75,446 types in 800 texts in four sub-disciplines of pharmacy. Each sub-discipline consisted of 200 articles, including 160 research articles, 25 review articles, and 15 short communications.

After analyzing the PRAC using WordSmith (Scott, 2017), erroneous items and function words were deleted, and the remaining words were turned into families by Familizer (Cobb, 2018b), presenting us 10129 word families. Following that, the above-mentioned four word-selection criteria were applied. Accordingly, there were 1862 families that occurred more than 100 times with +40 dispersion and appeared in at least 3 of the 4 sub-fields. Moreover, the word families that appeared in the first 2000 GSL were deleted, leaving us with 950 word families, which were subsequently reviewed by the experienced pharmacy researcher of this study to remove technical words. That left us with 750 word families, which represented the final version of PAWL (Question 1). Table 1 shows the statistical results of the top 50 word families in PAWL.

Table 1
The Top 50 Word Families in PAWL

Number	Headword	Frequency	Coverage	Number	Headword	Frequency	Coverage
1	cell	15703	0.46	26	interact	2394	0.07
2	drug	8120	0.24	27	factor	2352	0.07
3	analyze	7060	0.21	28	structure	2347	0.07
4	significant	6107	0.18	29	clinic	2338	0.07
5	protein	5591	0.17	30	potential	2337	0.07
6	data	5093	0.15	31	similar	2293	0.07
7	concentrate	4948	0.15	32	DNA	2281	0.07
8	inhibit	4672	0.14	33	receptor	2239	0.07
9	method	4454	0.13	34	assay	2134	0.07
10	expose	3927	0.12	35	administer	2119	0.07
11	induce	3878	0.12	36	gene	2103	0.07
12	dose	3837	0.12	37	react	2074	0.06
13	patients	3551	0.11	38	role	2018	0.06
14	compound	3408	0.10	39	activate	2011	0.06
15	process	3240	0.10	40	range	1976	0.06

16	release	3098	0.09	41	tumor	1949	0.06
17	acid	3022	0.09	42	involve	1900	0.06
18	molecule	2999	0.09	43	chemical	1824	0.06
19	function	2834	0.09	44	phase	1803	0.06
20	obtain	2694	0.08	45	complex	1763	0.06
21	tissue	2633	0.08	46	parameter	1717	0.05
22	response	2632	0.08	47	demonstrate	1700	0.05
23	previous	2573	0.08	48	medium	1663	0.05
24	species	2547	0.08	49	inject	1659	0.05
25	formula	2459	0.08	50	stress	1600	0.05

Table 2
Distribution of PAWL's 750 Word Families over AWL and MAWL

	PAWL families	percentage
Shared with AWL	322	42.9%
Absent in AWL	431	57.4%
Shared with MAWL	427	56.9%
Absent in MAWL	326	43.4%
Absent in both AWL and MAWL	277	36.9%

As can be observed in Table 2, only 322 (42.9%) of the 750 word families in the PAWL overlapped with the 570 word families in the AWL. This noticeable difference corroborates the argument that more discipline-specific wordlists are required (e.g., Hyland & Tse, 2007). Additionally, there were 431 (57.4%) word families that appeared in PAWL but were absent in AWL, which further highlights the necessity to develop domain-specific academic wordlists. Table 3 shows the coverage of PAWL and AWL over PRAC. AWL covers 9.28% of our pharmacy corpus, which is in line with AWL's coverage over corpora of various disciplines; science 9.1%, art 9.3%, law 9.4% and commerce 12% (Coxhead, 2000, p. 222). However, PAWL's coverage is 17.69%, which is almost twice as much as that of AWL. This further undermines the usefulness of general academic wordlists and reaffirms the need for a more restricted academic lexical repertoire (Question 2).

Table 3
PRAC Coverage Comparison of PAWL and AWL

	PRAC coverage
PAWL	17.69 %
AWL	9.28%
MAWL	16.07%

As a further purpose of this study, we compared the performance of our newly developed discipline-specific PAWL with that of MAWL, another domain-specific wordlist in a different but related field. Only 427 (56.9%) word families in PAWL overlapped with MAWL. In other words, there were 326 (43.4%) word families that appeared in PAWL but were absent in MAWL. This clear difference shows that while MAWL was developed for medical texts, compared to PAWL, it might fail to reflect the word coverage of pharmacy texts. Table 3 shows the coverage of PAWL and MAWL over PRAC. MAWL covers 16.07% of our pharmacy corpus. However, PAWL's coverage is 17.69%, which is approximately 2% more than that of MAWL. This further confirms the usefulness of a pharmacy-specific academic wordlist (Question 4). Furthermore, there were 277 (36.9%) word families that were exclusive to PAWL and did not appear in AWL nor in MAWL, suggesting the uniqueness of PAWL (Question 3).

PAWL had an average text coverage of 17.69% in our corpus. The following paragraph was randomly selected from a pharmacy article (Abe & Sasaki, 2016) in our corpus. Thirty-three of the total 176 words in this passage are included in PAWL, which are bold-faced. PAWL coverage of this passage is 18.75%, which is very similar to the results of the present study.

> The **cytotoxic** effect of the **DNA** alkylating agent is due to the massive production of the AP **site** by alkylation of the nucleobases, therefore, blockage of the AP **sites** repair pathway is expected to **enhance** the **cytotoxic** efficacy of such agents. In one **approach**, small-**molecular inhibitors** for APE1 have been **developed** as a **therapeutic** agent.6,7 On the other hand, Lhomme and co-workers **demonstrated** that the AP **site**-binding **molecules** potentiated the effect of the anticancer **drug**, bischloroethylnitrosourea (BCNU), in vitro and vivo.8 Binding of their **molecules** to the AP **sites inhibited** the repair and caused the **apparent** synergy with BCNU. Recently, Tell and co-workers also suggested that the endonuclease activity of APE1 was **modulated** by the nucleolus **protein** nucleophosmin (NPM1) and that the **inhibition** of association between APE1 and NPM1 **inhibited** the AP **site cleavage** to increase the anti-**proliferative** activity.9 Thus, the AP **site** binding **molecules** are attractive as a tool for the study of the AP **sites** repair pathway and also as a **potential candidate** to **enhance** the antitumor effect of the preexisting **DNA** alkylating agents.

Finally, to verify if our list was truly academic, we compared it against a general corpus. Using Coverage Calculator (Cobb, 2018a) revealed that PAWL covers only 4.87% of BNC Sampler Written, which is completely in line with AWL's 4.90% and MAWL's 4.16% coverage of the same corpus (Table 4). This low coverage of an academic wordlist over a general corpus confirms the academic nature of the wordlist. To further corroborate our results, we also compared PAWL against BNC Med using Cobb's (2018a) Coverage Calculator (Table 4). PAWL covered 14.76% of BNC MED, compared to AWL's 8.30% and MAWL's 11.67% coverage of the same corpus. These results further confirm that domain-crossing wordlists cannot be optimally effective and useful, and that domain-specific wordlists need to be developed. Moreover, it was surprising to see that our pharmacy wordlist had notably higher coverage (14.76%) over BNC Med than that of MAWL (11.67%), which is essentially a medical wordlist. This suggests that PAWL performs better in terms of coverage than MAWL not only in pharmacy texts, but also in medical texts (Question 5).

Table 4
BNC Coverage Comparison

	BNC Written Sampler	BNC Med
PAWL	4.87%	14.76%
AWL	4.90%	8.30%
MAWL	4.16%	11.67%

5. Discussion

In this study, the authors developed a pharmacy academic wordlist. To the best of our knowledge, this is the only academic word list that has been produced in the field of pharmacy. Furthermore, the corpus employed in this study is one of the largest of its kind with 3,458,445 running words. Moreover, unlike other studies (e.g., Wang et al., 2008; Yang, 2015) that used only research articles to develop academic wordlists, we employed not only *research articles*, but also *review articles* and *short communications*. The results of the present research will serve as a base for non-native pharmacy researchers and postgraduate students to develop a new understanding of this type of vocabulary and assist them in better using of English for Academic Purposes.

The findings of the current research can culturally benefit the members of the pharmacy

discourse community (i.e., pharmacists, pharmacy researchers, pharmacy students and practically anyone who needs to read, write and research within this discourse community). This is, evidently, because reading, writing, and researching within a given academic community, with its own communicative and social purposes, require its members to have an adequate command of the academic vocabulary that enables effective communication within the community and among its members.

The results drawn from the current study generate several important implications for researchers and postgraduate pharmacy students, EAP teachers and materials developers. Firstly, researchers and postgraduate pharmacy students can directly benefit from PAWL by learning the most frequent academic vocabulary of their field of study since knowledge of the base words can assist the learning of other derived and inflected forms in the word families (Bauer & Nation, 1993, p. 253). This knowledge can assist them to present and publish their research in English more effectively. Secondly, EAP teachers can use PAWL as a lexical basis to inform their approach to teaching reading and writing skills for pharmacy students. They can either incorporate this academic repertoire as their main classroom practice or employ it as supplementary material. Either way, knowledge of PAWL can improve the confidence of pharmacy students when reading or writing in English. Finally, EAP materials designers can use PAWL, either explicitly or implicitly, in developing pertinent reading/ writing tasks and in designing pharmacy course books.

The present study focused on individual words and did not investigate how these lexical items might collocate with one another. Future studies may inspect the patterns on which these academic words group. Also, our corpus was comprised solely of academic articles. Future studies can incorporate other sources such as textbooks or spoken corpora to determine whether or not our developed wordlist responds to genre changes in the same discipline.

Acknowledgment

The authors would like to thank Tom Cobb for his assistance during the data analysis phase of this study.

References

Abe, Y. S., & Sasaki, S. (2016). DNA cleavage at the AP site via b-elimination mediated by the AP site-binding ligands. *Bioorganic & Medicinal Chemistry*, *24*, 910–914.

Anthony, L. (2014). *AntWordProfiler v.1.4.1* [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from http://www.laurenceanthony.net/software

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279.

Biber, D., Connor, U., & Upton, T. A. (2007). *Discourse on the move: Using corpus analysis to describe discourse structure*. Philadelphia, PA: John Benjamins.

Campion, M., & Elley, W. (1971). *An academic vocabulary list*. Wellington, New Zealand: New Zealand Council for Educational Research.

Cobb, T. (2018a). *Coverage Calculator v.1.0* [Computer Software]. Retrieved from https://www.lextutor.ca/cover/

Cobb, T. (2018b). *Familizer v.2.0* [Computer Software]. Retrieved from https:// www.lextutor.ca/familizer/

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 15–38). Amsterdam, Netherlands: John Benjamins.

Corson, D. (1997). The learning and use of academic English words. *Language Learning*, *47*(4), 671–718.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew, & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 252–267). Cambridge, England: Cambridge University Press.

Flowerdew, J. (1993). Concordancing as a tool in course design. *System*, *21*(2), 231–244.

Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*, *17*, 24–27.

Grabowski, Ł. (2013). Register variation across English pharmaceutical texts: A corpus-driven study of keywords, lexical bundles and phrase frames in patient information leaflets and summaries of product characteristics. *Procedia-Social and Behavioral Sciences*, *95*, 391-401.

Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes*, *38*, 23-33.

Hartley, J. (2008). *Academic writing and publishing: A practical handbook*. New York, NY: Taylor & Francis.

Hewings, M. (2001). Introduction. In M. Hewings (Ed.), *Academic writing in context: Implications and applications: Papers in honour of Tony Dudley-Evans* (pp. 9-16). Birmingham, England: University of Birmingham Press.

Hyland, K. (2006). *English for academic purposes: An advanced resource book*. London, England: Taylor & Francis.

Hyland, K. (2009). *Academic discourse: English in a global context*. New York, NY: Continuum.

Hyland, K., & Tse, P. (2007). Is there an academic vocabulary? *TESOL Quarterly*, *41*(2), 235–253.

Kuehn, P. (1996). *Assessment of academic literacy skills: Preparing minority and limited English proficient (LEP) students for post-secondary education*. Fresno, CA: California State University (ERIC Document Reproduction Service No. ED415498).

Liu, J., & Han, L. (2015). A corpus-based environmental academic word list building and its validity test. *English for Specific Purposes*, *39*, 1–11.

Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*, *4*(1), 25–32.

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes*, *28*(3), 183–198.

McKinley, J., & Rose, H. (2018). Conceptualizations of language errors, standards, norms and nativeness in English for research publication purposes: An analysis of journal submission guidelines. *Journal of Second Language Writing, 42*, 1-11.

Montgomery, S. (1996). *The scientific voice*. New York, NY: The Guildford Press.

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, *25*(2), 235–256.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press.

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

Scott, M. (2017). *WordSmith Tools v.7.0* [Computer Software]. Stroud: Lexical Analysis Software. Retrieved from https://lexically.net/wordsmith/

Shaw, P. (1991). Science research students' composing processes. *English for Specific Purposes*, *10*(3), 189–206.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. New York, NY: Cambridge University Press.

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes*, *12*(4), 248–263.

Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, *27*(4), 442–458.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170–182.

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

Xue, G., & Nation, P. (1984). A university word list. *Language Learning and Communication*, *3*, 215–229.

Yang, M.-N. (2015). A nursing academic word list. *English for Specific Purposes*, *37*, 27–38.

Appendix

PAWL Headwords List Alphabetically

Numbe	Headword	Numbe	Headword	Numbe	Headword	Numbe	Headword
1	**abandon**	42	**approach**	83	breast	124	**code**
2	abnormal	43	**appropriate**	84	**brief**	125	cognition
3	**abstract**	44	**approximate**	85	buffer	126	cohort
4	abundant	45	aqueous	86	**bulk**	127	colon
5	abuse	46	**area**	87	burden	128	column
6	accelerate	47	aromatic	88	calcium	129	combine
7	**access**	48	array	89	calibrate	130	**communicate**
8	**accompany**	49	artery	90	cancer	131	**community**
9	**accumulate**	50	**aspect**	91	candidate	132	compartment
10	**accurate**	51	assay	92	**capable**	133	compete
11	**achieve**	52	assembly	93	**capacity**	134	**complement**
12	acid	53	**assess**	94	capsule	135	**complex**
13	**acquire**	54	**assign**	95	capture	136	compliance
14	activate	55	**assume**	96	carbon	137	**component**
15	acute	56	asthma	97	cardiac	138	**compound**
16	**adapt**	57	atmosphere	98	carry	139	**comprehensiv**
17	add	58	atom	99	cascade	140	compress
18	addict	59	**attach**	100	catalyse	141	**comprise**
19	**adequate**	60	attenuate	101	**category**	142	computer
20	adhere	61	**attribute**	102	cavity	143	**concentrate**
21	**adjust**	62	automobile	103	cell	144	**concept**
22	administer	63	**available**	104	cellulose	145	**conclude**
23	adolescent	64	axis	105	centrifuge	146	**conduct**
24	adsorption	65	bacterium	106	cerebral	147	confocal
25	**adult**	66	barrier	107	**challenge**	148	**conform**
26	adverse	67	baseline	108	chamber	149	conjugate
27	aerosol	68	batch	109	**channel**	150	connect
28	**affect**	69	bead	110	**chemical**	151	consecutive
29	**aggregate**	70	**bias**	111	chemistry	152	**consent**
30	**aid**	71	binary	112	chemotherapy	153	consequence
31	alcohol	72	bioactive	113	chloride	154	**considerable**
32	algorithm	73	bioavailabilit	114	cholesterol	155	**consist**
33	allergy	74	biochemical	115	chromatograph	156	consistent
34	**alter**	75	biology	116	chronic	157	**constant**
35	amorphous	76	biomarker	117	circulate	158	**constitute**
36	**analyse**	77	biosynthetic	118	**classic**	159	**construct**
37	antibacteria	78	biotechnology	119	clear	160	**consume**
38	antibiotic	79	blank	120	cleave	161	**contact**
39	antigen	80	blend	121	clinic	162	contaminate
40	apparatus	81	blot	122	cluster	163	**context**
41	**apparent**	82	**bond**	123	coat	164	contraction

(continued)

Number	Headword	Number	Headword	Number	Headword	Number	Headword
165	**contrary**	206	develop	247	elevate	288	**factor**
166	**contrast**	207	**deviate**	248	**eliminate**	289	favour
167	**contribute**	208	**device**	249	elute	290	**feature**
168	convention	209	diabetes	250	embed	291	feed
169	**convert**	210	diagnose	251	embryo	292	fetus
170	cord	211	diameter	252	**emerge**	293	**final**
171	**core**	212	diet	253	emit	294	**focus**
172	coronary	213	differential	254	emotion	295	**formula**
173	correlate	214	**differentiate**	255	**enable**	296	fraction
174	**correspond**	215	diffract	256	encapsulate	297	fragment
175	counter	216	diffuse	257	endogenous	298	**framework**
176	**couple**	217	digest	258	**energy**	299	frequency
177	crease	218	digital	259	engineer	300	front
178	**create**	219	dilute	260	**enhance**	301	**function**
179	**criteria**	220	**dimension**	261	enrich	302	**fundamental**
180	**crucial**	221	disorder	262	**ensure**	303	fungus
181	crude	222	disperse	263	**environment**	304	**furthermore**
182	crystal	223	**display**	264	enzyme	305	fuse
183	**culture**	224	disrupt	265	equation	306	gastric
184	cumulative	225	dissolve	266	equilibrium	307	gel
185	**cycle**	226	distil	267	equipment	308	**gender**
186	cytotoxic	227	**distinct**	268	**equivalent**	309	gene
187	**data**	228	**distribute**	269	**error**	310	**generate**
188	**decade**	229	**diverse**	270	**establish**	311	**generation**
189	**decline**	230	DNA	271	**estimate**	312	genetic
190	defence	231	dock	272	ethics	313	genome
191	degrade	232	**document**	273	**ethnic**	314	genotype
192	demography	233	**domain**	274	**evaluate**	315	genre
193	**demonstrate**	234	dominant	275	evaporate	316	gland
194	dense	235	donor	276	evidence	317	global
195	dependence	236	dopamine	277	evolution	318	**goal**
196	depict	237	dose	278	exam	319	**grade**
197	deplete	238	drug	279	**exclude**	320	gradient
198	deposit	239	dual	280	excrete	321	granulate
199	**depress**	240	**duration**	281	exert	322	**guideline**
200	**derive**	241	dye	282	**exhibit**	323	gut
201	descriptor	242	**dynamic**	283	exogenous	324	healthcare
202	**design**	243	dysfunction	284	**expose**	325	height
203	**despite**	244	**economy**	285	**external**	326	**hence**
204	**detect**	245	electronic	286	**extract**	327	herb
205	deter	246	**element**	287	**facilitate**	328	herein

(continued)

Number	Headword	Number	Headword	Number	Headword	Number	Headword
329	**highlight**	370	**instance**	411	lipid	452	mutate
330	histology	371	**institute**	412	liver	453	nanoparticle
331	histopathology	372	**instruct**	413	local	454	nasal
332	hormone	373	insulin	414	**locate**	455	negative
333	humid	374	intact	415	**logic**	456	negligible
334	hybrid	375	intake	416	magnet	457	nervous
335	hydro	376	**integrate**	417	**maintain**	458	**network**
336	hypertension	377	**intense**	418	**major**	459	neurology
337	**hypothesis**	378	**interact**	419	mammal	460	**neutral**
338	**identical**	379	interest	420	map	461	**nevertheless**
339	**identify**	380	interface	421	mark	462	**normal**
340	**illustrate**	381	**intermediate**	422	maternal	463	novel
341	**image**	382	**internal**	423	matrix	464	**nuclear**
342	immune	383	**interpret**	424	**mature**	465	obese
343	**impact**	384	**interval**	425	maximum	466	**objective**
344	impair	385	**intervene**	426	**mechanism**	467	**obtain**
345	**implement**	386	interview	427	**mediate**	468	**obvious**
346	**implicate**	387	intestine	428	**medical**	469	**occur**
347	**incidence**	388	**intrinsic**	429	**medium**	470	online
348	**incorporate**	389	invasion	430	membrane	471	onset
349	incubate	390	**investigate**	431	**mental**	472	optic
350	**index**	391	**involve**	432	mesh	473	optimal
351	**indicate**	392	ion	433	metabolic	474	**option**
352	**individual**	393	irradiate	434	**method**	475	oral
353	**induce**	394	**isolate**	435	micro	476	organism
354	induct	395	isomer	436	microscope	477	**outcome**
355	infant	396	**issue**	437	microwave	478	outpatient
356	infarct	397	**item**	438	**migrate**	479	**output**
357	infect	398	kidney	439	**minimise**	480	**overall**
358	infrared	399	kinetic	440	**minor**	481	overnight
359	infuse	400	kit	441	mobile	482	oxide
360	ingredient	401	**label**	442	**mode**	483	oxygen
361	inhale	402	laboratory	443	**modify**	484	**panel**
362	**inhibit**	403	larva	444	modulate	485	**paradigm**
363	**initial**	404	laser	445	moisture	486	**parallel**
364	**initiate**	405	lateral	446	molecule	487	**parameter**
365	inject	406	**layer**	447	**monitor**	488	pare
366	**injure**	407	**legal**	448	mood	489	**participate**
367	**input**	408	lesion	449	morphology	490	**passive**
368	**insert**	409	linear	450	mortal	491	pathology
369	**insight**	410	**link**	451	muscle	492	patients

(continued)

Number	Headword	Number	Headword	Number	Headword	Number	Headword
493	peak	534	precursor	575	recruit	616	score
494	pediatric	535	**predict**	576	refer	617	secrete
495	pellet	536	pregnant	577	regenerate	618	**section**
496	penetrate	537	**preliminary**	578	regimen	619	**select**
497	**perceive**	538	prescription	579	**region**	620	sensitize
498	**percent**	539	prevalent	580	regress	621	sensor
499	perception	540	**previous**	581	**regulate**	622	**sequence**
500	perfusion	541	**primary**	582	**relax**	623	**series**
501	**period**	542	**principal**	583	**release**	624	serum
502	periphery	543	**principle**	584	**relevant**	625	session
503	permeate	544	**prior**	585	**rely**	626	set
504	**persist**	545	probe	586	remodel	627	**sex**
505	**perspective**	546	procedure	587	**remove**	628	sham
506	pharmaceutical	547	**process**	588	renal	629	shear
507	pharmacokinetic	548	profession	589	replicate	630	**shift**
508	pharmacology	549	progress	590	**require**	631	signal
509	pharmacy	550	**project**	591	**research**	632	**significant**
510	**phase**	551	proliferate	592	**reside**	633	**similar**
511	**phenomenon**	552	prolong	593	residue	634	**simulate**
512	phosphate	553	**promote**	594	resin	635	simultaneous
513	**physical**	554	**proportion**	595	resist	636	**site**
514	physician	555	**prospect**	596	resolution	637	slice
515	physiological	556	protein	597	**resource**	638	sodium
516	pilot	557	**protocol**	598	respire	639	software
517	placebo	558	psychiatry	599	response	640	soluble
518	plasma	559	**publish**	600	**restrict**	641	**source**
519	plastic	560	**purchase**	601	**retain**	642	spatial
520	platform	561	pure	602	**reveal**	643	species
521	plot	562	**qualitative**	603	**reverse**	644	**specific**
522	**plus**	563	quantify	604	robust	645	specimen
523	polar	564	questionnaire	605	rodent	646	spectroscope
524	**policy**	565	radiate	606	**role**	647	spectrum
525	polymer	565	**radical**	607	rotate	648	spontaneous
526	pool	567	radioactive	608	**route**	649	**stable**
527	pore	568	**random**	609	routine	650	standard
528	**portion**	569	**range**	610	rural	651	**statistic**
529	**positive**	570	**ratio**	611	saline	652	**status**
530	potency	571	**react**	612	saturate	653	sterile
531	**potential**	572	reagent	613	scaffold	654	steroid
532	precipitate	573	receptor	614	scan	655	stimulate
533	**precise**	574	**recover**	615	**scheme**	656	store

(continued)

Number	Headword	Number	Headword	Number	Headword	Number	Headword
657	strain	681	systemic	705	transient	729	**vehicle**
658	strand	682	tablet	706	transition	730	vein
659	**strategy**	683	**target**	707	translocate	731	velocity
660	**stress**	684	**task**	708	**transmit**	732	**version**
661	**structure**	685	**team**	709	**transport**	733	versus
662	subjective	686	**technique**	710	treat	734	**via**
663	**subsequent**	687	**technology**	711	**trend**	735	viable
664	**substitute**	688	temporal	712	triplicate	736	vial
665	substrate	689	tensile	713	tumour	737	victim
666	suicide	690	terminal	714	type	738	virus
667	**summary**	691	**theory**	715	**ultimate**	739	viscous
668	superior	692	therapy	716	**undergo**	740	**visible**
669	**supplement**	693	**thereby**	717	**underlie**	741	**visual**
670	suppress	694	thermal	718	**uniform**	742	vital
671	surfactant	695	threshold	719	**unique**	743	vitamin
672	surgery	696	tissue	720	unit	744	**volume**
673	**survey**	697	**topic**	721	uptake	745	volunteer
674	**survive**	698	toxic	722	**utilise**	746	weight
675	susceptible	699	**trace**	723	vaccine	747	**whereas**
676	**suspend**	700	tract	724	vacuum	748	write
677	**sustain**	701	**tradition**	725	**valid**	749	yeast
678	symptom	702	transcript	726	**vary**	750	zone
679	syndrome	703	**transfer**	727	vascular		
680	synthesis	704	**transform**	728	vector		

Note: AWL words are bold and MAWL words are underlined.