Language Teaching Research Quarterly

2019, Vol. 9, 69–82

A Corpus-based Study of Academic Vocabulary in Physiotherapy Research Articles

Mehri Jamalzadeh*, Azize Chalak

Department of foreign languages, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran

Received *17 November 2018* ***Accepted*** *19 February 2019*

Abstract

Scientific English is not homogenous but varies in relation to the contexts. As a contribution to the field, this study developed a list of academic words used frequently in physiotherapy research articles (RAs) and compared it with the distribution of high frequency words in Coxhead's (2000) Academic Word List (AWL) and West's (1953) General Service (GSL). By analyzing a 1.7 million-word corpus, 1450 high frequent academic word families were identified and constituted the Physiotherapy Academic Word List (PAWL). The analysis showed that, of the 570 word families in AWL, 562 occurred frequently in physiotherapy research articles (RAs) and this provided a coverage of 11.51 of the tokens in the corpus. In addition, 406 word families (accounting for 28% of the word families in PAWL) found to be used frequently in PRAC had not been listed in GSL and AWL. The results indicate that AWL (Coxhead, 2000) is not entirely useful for physiotherapy learners because of the narrow coverage of some word families and the shortage of frequently used physiotherapy academic words. The established PAWL may serve as a guide for instructors in curriculum preparation, and for physiotherapy English learners in setting their vocabulary learning goals.

Keywords: *The AWL, The GSL, Corpus Study, Physiotherapy*

Introduction

Scientific English as described by Halliday (2004) refers to a generalized functional variety or register of the modern English language. The variation can be summarized in terms of field, tenor and mode: By field, indicating extending, transmitting or exploring knowledge in the physical, biological or social sciences; by tenor, whether it is addressed to specialists, to learners or to laymen, from within the same group or across groups (e.g., lecturer to students); and by

m82jamalzadeh@yahoo.com doi: 10.32038/ltrq.2018.09.06

mode, he refers to phonic or graphic channel, most congruent or less so and rhetorical function variety —expository, hortatory, polemic, imaginative and so on.

Researchers communicate within specific academic discourse communities through the research article genre. They are the pre-eminent genre of the academy restructuring the processes of thought and reported research to establish a discourse for scientific fact-creation. (Hyland, 2010). As Hyland & Paltridge (2011) puts it, three major developments over the past 20 years have led to interest in academic discourse, and particularly academic writing in English: changes in higher education leading to greater interest to the importance of writing; position of English as the international language of research and scholarship; and the emergence of theoretical perspectives with their emphasis on the centrality of academic discourses knowledge construction.

With the increasing non-native and native academics publishing in English, the study of academic discourse and academic vocabulary, in particular, is inevitable. Academic vocabulary are words reasonably frequent in a wide range of academic genres but relatively uncommon in other kinds of texts (Coxhead & Nation, 2001). Academic vocabulary is one of four levels of vocabulary division by Nation (2001). They are high frequency words, academic vocabulary, technical vocabulary and low frequency words. High frequency general words are core words used very frequently in most language use (Nation &Hwang, 1995). West's (1953) General Service List (GSL) reporting the 2000 most frequent word families are of this kind.

However, Engels, 1968 and Richards, 1974 criticized West's list for its size and its age. The size criticisms question the necessity of the second 1000 words of the GSL because they usually cover only 4-5% of the running words in non-fiction texts compared to the 70% plus coverage of the first 1000. The report on which the GSL is based was prepared in the 1930s and since language changes overtime, the GSL is too old, it contains many words that are not essential and does not contain high frequent current words y (Nation & Hwang, 1995).

Nation & Waring (1997) suggest that EAP students need to first learn the 2000 or so most ‘‘general” words of English, and continue by a set of ‘‘academic” words common to all academic disciplines. General words refer to the 2000-word family GSL mentioned above and academic words to the 570-word family Academic Word List or AWL (Coxhead, 2000). About 90 percent of the running words in an academic text are covered by Academic words and general words together. (Coxhead, 2000; Coxhead & Nation 2001; Nation, 2001)

AWL has been formed upon an analysis of 3.5 million-word corpus of written academic English in 28 sub-disciplines of four main disciplines of Art, Commerce, Law, and Science. AWL includes 570 word families selected in accordance with three criteria: a) frequency of occurrence (occurrence of at least 100 times in the corpus), b) range (occur at least 10 times in each of the 4 disciplines and in 15 or more subject areas) and c) specialized appearance (be outside the first 2000 GSL words). Coxhead categorized the AWL items into 10 sublists according to their frequency. All sublists contain 60 word families, except the tenth sublist, which contains 30 word families. most frequent AWL items in the first sublist, with 3.6% coverage of the corpus; the last sublist comprises the least frequent ones with a coverage of only

0.1% of the corpus. The coverage, however, was not the same for all the 4 subject areas: the list provides the highest coverage over commerce (12.0%) and the lowest over science (9.1%).

Hyland & Tse (2007) pointed out that individual lexical items on the list often occur and behave in different ways depending on the disciplines in terms of range, frequency, collocation, and meaning. As they point out "the different practices and discourses of disciplinary communities reduces the usefulness of such lists" and suggested, "that teachers help students develop a more restricted, discipline-based lexical repertoire" (Hyland & Tse, 2007, p. 235).

According to Hyland and Tse (2007), the best way to prepare students for their academic studies is to provide them with an understanding of the features of the discourses they will encounter in their particular courses. As Coxhead, 2017 puts it, "Studying at university can mean exposure to several million running words a year through reading textbooks, source books, content and learning-based websites, and other academic sources of information." p. 90. Word list research has also been driven by the needs of particular groups of language learners and to help set learning goals (Nation, 2016).

Literature Review

Some studies have developed academic word lists across disciplines. Campion and Elley (1971) and Praninskas (1972) developed the earliest lists of frequently used words students would encounter during their academic studies in a range of different university disciplines. Lynn (1973) and Ghadessy (1979) developed lists of difficult words for students who needed to read academic texts. The University Word List (UWL) contains 836 high frequencies non-GSL words across a wide range of disciplines was produced by Xue and Nation (1984) for university level students. They compiled the list from Campion and Elley (1971), Praninskas (1972), Lynn (1973), and Ghadessy (1979).

Recently Coxhead (2000) developed the aforementioned academic word list whose frequency accounted for approximately 10% of tokens in academic texts. Coxhead (2000) set up a corpus of 3.5 million running words chosen from various academic journals and college course books in four primary subjects: the arts, commerce, law and natural science. The AWL list that contains 3,112 individual items, does not include words in the most frequent 2,000 word families in the English language and West's (1953) GSL.

Field-specific academic word lists or discipline-based lexical repertoires have been produced for different disciplines. Kwary and Artha (2017) created and tested a word list called the Academic Article Word List for Social Sciences. Lei Lei and Dilin Liu (2016) developed a new medical academic vocabulary list based on the results of a series of comparative analyses. Hsu (2013), Wang, Liang & Ge (2008), Chen & Ge, (2007), developed medical academic word lists too. Wang et.al presented a word list including 623 non-GSL word families occurring frequently across medical research articles. A coverage of 12.24% was provided over their corpus. Just 342 of the 623 high frequency word families identified, coincided with those listed in AWL. Chen and Ge (2007) found that only 292 of 570 AWL's word families were frequent in medical research articles and 179 AWL items either did not occur or occurred infrequently in the corpus.

Valipouri & Nassaji, (2013) developed an academic word list for chemistry discipline. They examined a 4 million-word corpus of chemistry research articles. They identified 1400 word families used with a reasonable frequency in the corpus. Muñoz (2015) and Martinez et. al.'s (2009) academic word lists included 1941 word families occurring frequently in agriculture corpus. Li and Qian (2010) profiled the presence of the AWL in a financial services corpus. It was found that just 28.42% of AWL items or to put it more exact, 162 word families were frequently used in their corpus. Hyland and Tse (2007) compared the use of academic words in textbook chapters, academic book reviews, master's theses, and doctoral dissertations. Collectively, these studies outline a critical role for domain-specific word lists. They call for more research that examines field-specific corpora in order to develop such word lists for students studying in different disciplines.

Ardasheva, and Tretter, (2017) examined current vocabulary literature to identify effective, research-based vocabulary instruction principles and developed the SVS program suitable to the learning needs of high school newcomers. Watson-Todd (2017) created a list of opaque words for teaching engineering English at a Thai university by comparing the meanings of words in the context against the main meanings given in the online dictionaries that students often rely on. Masrai and Milton (2018) demonstrated that the learning of the AWL appears to be strongly influenced by the frequency of these words in general corpora and that the AWL test very strongly resembles a test of overall vocabulary size.

In the present study, we decided to examine a large written corpus of academic research articles in the field of physiotherapy to develop an academic word list that could be useful for physiotherapy students. An academic word list exclusively for physiotherapy students can be taught and directly studied in the same way as the words from the GSL. A physiotherapy academic word list can also play an important role in helping EFL physiotherapy students learn academic English more effectively. Therefore, the study aimed to establish a Physiotherapy Academic Word List (PAWL) of the most frequently-used physiotherapy academic vocabulary across different sub-disciplines in physiotherapy by examining a written corpus of academic research articles in this field.

This study also seeks to determine whether and to what extent the words identified as high frequency in the physiotherapy corpus have also been identified as high frequency in West's (1953) General Service List (GSL) and Coxhead's (2000) Academic Word List (AWL) and vice versa. in other words, the second aim is to check the overlap and the difference between our list and the current well-known general and academic word lists, and hence highlight its usefulness for physiotherapy students. Thus, the following e research questions were posed:

1. What are the most frequently used academic words in a corpus of physiotherapy research articles ?
2. How do the most frequently Physiotherapy research articles academic words compare against Coxhead AWL and West GSL word lists?

Methodology

The Corpus

A specialized corpus was built to represent research articles genre, science register, and discipline of physiotherapy. In the discipline of physiotherapy, there are five main subjects: Orthopedic, Geriatric, Neurological, Cardiopulmonary and Pediatric. The samples in the corpus were chosen equally from the five subject areas. It was designed following the criteria proposed by Sinclair (1991, 2005) and Barnbrook (1996), considering representativeness, specificity of corpus, use of whole documents, and availability in electronic form. The relevant research articles were obtained from the electronic journals of *International Journal of Neurologic Physical Therapy*, *Journal of Orthopaedic & Sports Physical Therapy (JOSPT)*, *Pediatric Physical Therapy*, *Journal of Physiotherapy Research*, *The Cardiopulmonary Physical Therapy Journal*, *Journal of Cardiac and Pulmonary Rehabilitation* and *Physiotherapy* (Elsevier) with full text. The articles included in the corpus were published between 2016 and 2018. As recommended by Swales (1990), all the articles included in the corpus had identifiable Abstract, Introduction, Method, Result and Discussion sections (IMRD). Thus, if an article did not follow IMRD format, they were not selected.

A three-step selection was followed to choose the sample physiotherapy research articles for the corpus. Firstly, only research articles focusing on empirical studies, written in the identifiable Introduction, Method, Result, and Discussion sections, were included in the PRAC. secondly, the research articles chosen had to have been published between 2013 and 2016. Thirdly, the length of the chosen articles must be longer than 2,000 running words and shorter than 10,000 running words. After this three-step selection, 60 articles were selected from each subject area and a total of 300 articles were chosen for the PRAC. The average length of a research article in the PRAC is 3,996 running words, the shortest containing 2,134 running words and the longest containing 8,013 running words. The articles were collected in their electronic version with their reference lists, appendices, captions, footnotes, and acknowledgments removed (Swales, 1990). The corpus represents a genre, experimental research articles, and a field, that is, physiotherapy. The results show that the PRAC contains 1,768,176 running words.

Table1
Description of the Corpus

Corpus	Number of articles	Number of tokens
PRAC	300	1,768,176

Instruments

The computer software used for the analysis was AntWordProfiler. It is a much more modern version of the program with numerous extra features and is available from Laurence Anthony's Website. (Nation, 2014). This tool generates vocabulary statistic and frequency information about a corpus of texts loaded into the program. It compares the files against a set of vocabulary level lists based on the research of Paul Nation. (Anthony, 2014).

Data Processing

The research articles were in a PDF format and had to be converted into a .txt for further processing. After obtaining the corpus, the software called AntWordProfiler available from Laurence Anthony's Website was used for analysing the vocabulary load of texts. The special version used in this study can tell how much and what vocabulary occurs in the particular group of GSL/AWL texts. Most of the previous studies on developing academic vocabulary have used Range software. AntWordProfiler Vocabulary Analysis programs, is a much more modern version of the program with numerous extra features (Nation,2014). Previously, Jamalzadeh, 2017 used another corpus analysis toolkit of the same series called AntConc for conducting a corpus-based study on cohesive conjunctions on medical research articles and it was found as an efficient and really user-friendly program. Thus using the program, the number of occurrences of each word, its range (i.e. in how many different texts each word has occurred) as well as the words shared with the AWL and GSL word lists were determined.

For a word family to be included in the list, Coxhead's (2000) procedure was followed. Coxhead's corpus for the AWL consisted of 3.5 million words and the non-GSL word families that occurred 100 times in the entire corpus and at least 10 times in each of the 4 disciplines were included in her list. Since words with a frequency of 100 times in the whole corpus were selected, it becomes almost 28.5 times in a million words. There were 1,775,162 words, so it was decided that the cut-off frequency of members of a word family should be equal to or higher than 51 times in the whole corpus to be included in the list. With regard to range, only word families appearing in all five sub-disciplines that occurred at least five times in each sub-discipline were included in the list. All the content words that met the set criteria were identified. Function words such as pronouns, auxiliaries, articles, and numbers were excluded from the analysis as these were considered too general. Chung and Nation's (2003) rating scale was used to distinguish technical from academic words, and it was decided that words judged to be unknown or not used in other sub-disciplines should be considered as too technical.

Results

Establishment of a Physiotherapy Academic Word List

Identifying words that were frequently used by academic writers in the field of physiotherapy concerned one of the main questions of this study. To answer this question, the two criteria of range and frequency used by Coxhead (2000) were applied. As noted earlier, all words with a frequency of 51 or more in the whole corpus, which occurred at least five times in all the five physiotherapy sub-disciplines, were considered as frequent.

Altogether, 1620 content word families met the frequency criteria. Of the 1620 word families, 145 were considered as too technical and thus were eliminated from the list. Excluding some of the basic function words, some prepositions and numbers finally, 1450 word families were left, which constituted Physiotherapy Academic Word List (PAWL). It includes 704 GSL word families plus 340 AWL word families plus 406 non-GSL/non-AWL word families.

AWL and PAWL word forms used in the PRAC

The coverage of GSL, AWL, and non-GSL/non-AWL words (word families and word tokens) in PRAC corpus is depicted in Table 1. (As the table shows, the 2000 most frequent word families of GSL accounted for 2,480,109 tokens (66.81%) of the corpus. Within the 1st 1000 most frequent GSL word families, 970 word families occurred in the corpus, out of which 520 word families appeared with the criterion frequency (higher than 51 times) in the corpus. Excluding those function words stated above, 490 word families from the 1st 1000 GSL words were found in PAWL. Within the 2nd 1000 most frequent GSL, 742 word families occurred in the corpus. However, only 214 of them met the frequency criterion. Out of the 2000 GSL word families, 704 occurred (with the criterion frequency) in PRAC. As for AWL words, of the 570 AWL word families, 562 occurred in physiotherapy corpus, out of which 340 met the cut-off frequency criterion. The frequency counts of tokens of these word families were 203,469 accounting for a coverage of 11.51% of PRAC. The AWL coverage in PRAC is higher than the coverage in both Coxhead's science corpus (i.e. 9.1%), which consisted of different scientific disciplines, and in Coxhead's multi-disciplinary corpus (10%). Out of the 1450 word families in PAWL, 406 (28%) were not among GSL and AWL. These words comprised 301,222 tokens (a total of 17.03%) of PRAC. The coverage of these words in the corpus and the order of their frequency of occurrence are shown in Table 1.

Table 2
The coverage of different base word lists over PRAC

Word lists	Tokens	% of PRAC	Headwords in PAWL
1st GSL	1011623	57.21	490
2nd GSL	117085	6.62	214
AWL	203469	11.51	340
Non-GSL/AWL	435990	24.66	406
Total	1768167	100	1450

Table 2 displays the 10 most frequent word families and their frequency in PAWL. The most frequent word in PAWL is the *physical* word family, which occurred more than 9,000 times in PRAC.

Table 3
The first 10 most frequent words in PAWL

Word	Frequency	Word	Frequency
physical	9777	participants	4406
therapy	7689	group	4152
study	7444	sports	4051
patient	5334	Journal	3938
pain	4737	Health	3841

Discussion

The present study was an attempt to identify frequently used words in physiotherapy research articles and develop a word list for physiotherapy students. Furthermore the list was compared with the two general and academic word lists (GSL and AWL) to explore their coverage. One issue about developing field-specific word lists is about whether to exclude GSL or general academic words from the list or not. Some researchers are against the exclusion of these words because they are general as many of them are highly frequent and have sometimes different meanings, uses and collocations in specialized contexts (Billuroglu & Neufeld, 2005, 2007; Martinez et al., 2009; Paquot, 2007; Ward, 2009). Billuroglu and Neufeld (2005), for example, put aside the idea of classifying words as 'general' or 'academic' and combined all the words in a series of word lists (e.g. GSL, AWL, Brown corpus, etc.) and built Billuroglu – Neufeld list (BNL) with 2709 word families. For our purpose, we did not exclude GSL or general academic words. So we constructed a word list based on frequency and range of all words that appeared in the corpus.

An Analysis of the PAWL

This study found that the coverage of AWL word forms in the Physiotherapy Research Articles Corpus was 11.51. This is higher than 9.3% coverage of AWL found in Hyland and Tse's (2007) science sub-corpus, 9.1% of Coxhead's (2000) science sub-corpus, and 9.06% in Martinez et al.'s corpus of agricultural research articles. However, AWL coverage in PRAC was slightly lower than the 11.7% coverage over applied linguistics corpus in Vongpumivitch, Huang, & Chang's (2009) study. These findings endorse that AWL items are not equally useful for students in specific fields

Regarding the first research question altogether, 1450 word families with a reasonable frequently were identified in the corpus. These words are called physiotherapy Academic Word List (PAWL). Word families of *physical* and *therapy* and *study* appear most often (9993, 9635 ,7473 times, respectively). In this study, the 10 word families with the highest frequencies account for 3.79% of running words in the corpus. Conversely, *transfer, substance* and *slow* word families appear least often (251, 250, 250 times respectively). In this study, the 10 word families with the lowest frequencies account for only 0.1.2% of the running words in the corpus.

Comparing the PAWL with the AWL and the GSLs

The second question in this research was about comparing the most frequently Physiotherapy research articles academic words against Coxhead AWL and West GSL word lists. The PAWL list was compared with the two general and academic word lists (GSL and AWL) to explore their coverage. Of its 1450 word families, the PAWL shares 1044 with the AWL and the GSLs. As for GSL words, out of 2000 GSL word families, only 704 occurred with high frequency. This suggests that, for graduate physiotherapy students who need to read research articles, only about one third of GSL word families may be worth learning. While Following Coxhead (2000), we calculated the most frequent 60 word families in our corpus and also compared them with those

in Coxhead's sublists. the first 60 academic word families in PAWL accounted for 230,117 tokens of the PRAC. They accounted for 5.70% of the corpus. The headwords of these word families are shown in Table 4. The numbers below the words indicate the corresponding Coxhead's AWL sublist. Coxhead divided the AWL into ten rank-ordered sublists, according to decreasing word family frequency. Except sublist 10, each sublist contains 60 items. Eighteen PAWL word families appear in the AWL's sublist 1, thirteen in sublist 2, eight in sublist 3, six in sublist 4. The remaining four word families appear separately in sublists 5 through 8. For example, *task* and its family members are in sublist 3 of Coxhead's AWL. Within the 60 word families, the words coinciding with items in sublist 1 of Coxhead's AWL are shown in bold type.

Among the first 60 word families, 18 coincided with Coxhead's sublist 1, 8 items fewer than the reported figure in Martinez et al.'s (2009) corpus of agricultural research articles, and 17 items less than the reported figure in Hyland and Tse's multi-disciplinary corpus. Some low frequency word families in Coxhead's study, such as *visual, duration* and *manual* (included in the seventh and eighth sublists of the AWL) were among the first most frequent word families in CAWL. Some word families identified as frequent in Coxhead's sublist 1 such as *assessment, assume* and *analysis*, did not appear in our corpus. However, *data, research, significant and function* from sublist 1 were among the first most frequent word families in PAWL. These findings suggest that words occurred with different frequencies in physiotherapy research articles than those identified in academic word lists such as AWL. Word families included in the PAWL are field-specific academic words rather than technical words. Admittedly, the "division between technical and non-technical vocabulary is far from distinct" (Mudraya, 2006, p. 238). One distinguishing factor is 'range', as "technical words occurred frequently in a specialized text or subject area but did not occur or were of very low frequency in other fields" (Nation, 2001, pp. 18–19), whereas academic vocabulary has been observed to be common across academic texts from different fields (Coxhead & Nation, 2001). Huang and Yang (1984) hold that words should generally be treated as technical terms if they only appear in one or very few subject areas. Conversely, if words appear in many subject areas of a particular discipline, they should not be considered as technical terms. There are generally two types of academic word lists. One type, represented by the AWL, includes academic words that are commonly used across disciplines. The other type focuses on a particular discipline, such as electronics (Farrell, 1990) or medicine (Wang et al., 2008), and contains academic words from this particular discipline. The second type is a field-specific academic word list which includes words commonly found in different subject areas of a particular discipline, while a technical word list includes words that are poorly distributed within a discipline despite their high frequency. In other words, it is necessary to consider 'range' when compiling word lists (Nation, 2001). For example, although people outside the medical field tend to consider them technical terms, words such as lesion and vein are included as medical academic vocabulary in Wang et al.'s (2008, p. 451) Medical Academic Word List because "they are general purpose medical words frequently used across different medical subject disciplines". Similarly, the PAWL includes words commonly used across the five subject areas of the physiotherapy research articles corpus. These words cover a wide range

and occur frequently in the physiotherapy research articles corpus. Therefore, PAWL words are field-specific academic words rather than technical words.As seen in Table 2, the PAWL covers the physiotherapy research articles corpus better than the AWL. To sum up, compared to the AWL, the PAWL includes more word families that cover a wider range, appear more often and can better reflect lexical features of physiotherapy research articles.

Table 4
The first most frequent 60 AWL word families in the PAWL compared with those in AWL sublists

1.	PHYSICAL 3	2.	JOURNAL 2	3.	**DATA 1**	4.	**RESEARCH 1**
5.	**SIGNIFICANT 1**	6.	FUNCTION 1	7.	OUTCOME 3	8.	RANGE 2
9.	MEDICAL 5	10.	CRITERIA 3	11.	**AVAILABLE 1**	12.	INDEX 6
13.	**SECTION 1**	14.	PREVIOUS 2	15.	**SPECIFIC 1**	16.	VOLUME 3
17.	ACADEMY 5	18.	**INDIVIDUAL 1**	19.	PRIMARY 2	20.	VISUAL 8
21.	DURATION 9	22.	MANUAL 9	23.	**SIMILAR 1**	24.	DESIGN 2
25.	**PERIOD 1**	26.	STATUS 4	27.	TASK 3	28.	SURVEY 2
29.	IMPACT 2	30.	CAPACITY 5	31.	POTENTIAL 2	32.	**METHOD 1**
33.	**PROCESS 1**	34.	**APPROACH 1**	35.	ITEM 2	36.	OVERALL 4
37.	VERSION 5	38.	POSITIVE 2	39.	**ASSESS 1**	40.	COMMUNITY 2
41.	**ROLE 1**	42.	NORMAL 2	43.	INITIAL 3	44.	PROTOCOL 9
45.	RATIO 5	46.	**IDENTIFY 1**	47.	**FACTOR 1**	48.	SEX 3
49.	APPROPRIATE 2	50.	AUTHOR 6	51.	INTERNAL 4	52.	TARGET 5
53.	PRIOR 4	54.	**AREA 1**	55.	ACCESS 4	56.	CONSENT 3
57.	MINIMAL 9	58.	GENDER 6	59.	PHASE 4	60.	INTERVAL 6

Coverage of Unlisted Words in the Physiotherapy Corpus

Unlisted words are the words that appear in neither the AWL nor the GSL. Table 2 shows that unlisted words cover 24.66% of the running words in PRAC. This study aims to highlight field-specific academic words, so unlisted words are only compared with those in the AWL here. Two factors potentially explain why unlisted words have relatively high coverage in PRAC. First, the AWL does not include some academic words that are commonly used in physiotherapy research articles and some AWL word families seldom appear in the PRAC. Table 5 shows some examples of unlisted vocabulary that appear PRAC, and their word frequency is quite high. For example, *therapy* occurs 7473 times in the PRAC. In contrast, a number of AWL vocabulary

items have extremely low frequency. The word *immigrate* is a case in point: It appears only once. For students majoring in physiotherapy, words like *immigrate* are of little or no help in reading professional literature and writing research essays. Second, in contrast to Coxhead's (2000) Academic Corpus, the PRAC is a specialized one and contains a relatively higher number of field-specific terms that occur often and widely in the discipline of physiotherapy.

Table 5
The first most frequent 68 unlisted words in the physiotherapy corpus

therapy	7473	patients	5191	orthopaedic	2831	score	1781
gait	1691	Muscle	1617	physiotherapy	1468	rehabilitation	1435
chronic	1323	Disability	1212	pediatric	1194	cardiovascular	1131
Baseline	1122	hip	1109	unauthorized	999	cognitive	927
Mobility	862	syndrome	859	postural	858	symptoms	851
dementia	832	diabetes	812	cerebral	800	acute	780
randomized	756	palsy	743	extremity	715	ankle	695
spine	683	therapists	682	concussion	659	obesity	633
Surgery	625	frailty	622	neurology	615	correlation	613
spinal	599	van	596	clinicians	574	geriatric	573
impairment	564	sessions	564	diagnosis	552	session	551
Median	547	infants	541	therapist	535	parkinson	519
stimulation	502	physiotherapists	501	prevalence	494	neurologic	485
peak	482	mortality	472	Sensory	466	pulmonary	456
regression	455	height	448	cohort	445	video	445
disorders	444	vestibular	442	musculoskeletal	434	versus	426
lumbar	419	Aerobic	418	impairments	418	metabolic	413

Conclusion

In this study, we examined a 1.7 million-word corpus of research articles in physiotherapy. The aim was to identify frequently used words in physiotherapy research articles and develop a word list for physiotherapy students. We also compared the list with the two general and academic word lists (GSL and AWL) to explore their coverage. In total, we identified 1450 word families used with a reasonable frequently in the corpus. We called these words physiotherapy Academic Word List (PAWL). A comparison of PAWL with AWL showed that many of the AWL items were not used frequently in the subject area we examined. High frequent AWL items had a different frequency order from those in Coxhead's AWL, indicating that academic words are not used similarly across disciplines. In addition, there were many non-AWL content word families that occurred with high frequency in the corpus. This supports the idea of developing field-specific vocabulary lists.

Learners study more effectively by having access to a field-specific academic word list. This study investigated the frequency of the Coxhead's (2000) AWL in the compiled corpus of

physiotherapy research articles and develop a word list for Physiotherapy students. Altogether, 1450 word families with a reasonable frequently were identified in the corpus. These words were called physiotherapy Academic Word List (PAWL). A comparison of the words with AWL showed that many of the AWL items were not used frequently in the subject area examined in this study. In cases when some AWL items with high frequency occurred, their frequency order was different from those in Coxhead's AWL. The resulting field-specific academic word list, compared with the AWL, will cover a higher percentage of the PRAC and better reflect the lexical features of physiotherapy RAs. This finding indicates that academic words across different disciplines are not used similarly. It supports the idea of developing field-specific vocabulary lists which are obtained from the target genres and texts that students need to read and write in their own academic discipline (Hyland & Tse, 2007; Martinez, Beck & Panza, 2009; Wang et al., 2008).

Hyland and Tse recommend that "teachers help students develop a more restricted, discipline-based lexical repertoire" (2007, p. 235). Based on the findings of this study, students of physiotherapy are recommended to direct their attention to the first two subsets of AWL as well as the list of words provided through the analysis of this study. Field-specific word lists for students in different fields is recommended. Field-specific lists help students learn necessary words which are specifically important for their field of study. For teaching and learning general academic courses, these word lists are considered as one of the best efficient and practical methods. It would be of special significance for physiotherapy students/instructors and professionals in learning or using physiotherapy academic vocabulary in reading and writing.

This research is only a preliminary study on the medical academic vocabulary used in physiotherapy RAs. If possible, the PAWL needs to be rechecked in larger corpora or in other genres of physiotherapy, such as physiotherapy textbooks or spoken physiotherapy academic English. The occurrence of the words in different contexts was not examined. PAWL consisted of isolated words. Knowing words while in isolation does not mean knowing how to use or understand them in context. Furthermore, students need to know not only the meanings of words but also how they co-occur (collocations) with other words in specific domains. It is possible that each word may be associated with different words and that the frequency of these associations may vary in different contexts and disciplines. Thus, further research is needed to examine how words collocate with other words and how their specific meanings and functions vary in different disciplines We hope the availability of exercises and tests based on the PAWL will promote effective and efficient teaching and learning of physiotherapy academic vocabulary.

References

Ardasheva, Y., & Tretter, T. R. (2017). Developing science-specific, technical vocabulary of high school newcomer English learners. International Journal of Bilingual Education and Bilingualism, 20(3), 252-271.

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

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*(4), 502-514.

Chung, T. M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a foreign language*, *15*(2), 103.

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

Coxhead, A. (2017). *Vocabulary and English for Specific Purposes Research: Quantitative and Qualitative Perspectives*. Routledge.

Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. *Research perspectives on English for academic purposes*, 252-267.

Farrell, P. (1990). Vocabulary in ESP: A Lexical Analysis of the English of Electronics and a Study of Semi-Technical Vocabulary. CLCS Occasional Paper No. 25.

Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: a new approach. In *English Teaching Forum* (Vol. 17, No. 1, pp. 24-27).

Halliday, M. A. (2004). The language of science (The collected works of MAK Halliday Vol. 5). *London: Continuum*.

Hsu, W. (2013). Bridging the vocabulary gap for EFL medical undergraduates: The establishment of a medical word list. *Language Teaching Research*, *17*(4), 454-484.

Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, *9*(2), 116-127.

Jamalzadeh, M. (2017). A Corpus-based Study of Cohesive Conjunctions in Medical Research Articles Written by Iranian and Non-Iranian Authors. *Journal of Teaching English for Specific and Academic Purposes*, 669-686.

Kwary, D. A., & Artha, A. F. (2017). The academic article word list for social sciences. *MEXTESOL*, *41*(4), 1-11.

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, *22*, 42-53.

Li, Y., & Qian, D. D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System*, *38*(3), 402-411.

Lynn, R. W. (1973). Preparing word-lists: a suggested method. *RELC journal*, *4*(1), 25-28.

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.

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44-57.

Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, *39*, 26-44.

Nation, I. S. (2001). *Learning vocabulary in another language*. Ernst Klett Sprachen.

Nation, P., & Kyongho, H. (1995). Where would general service vocabulary stop and special purposes vocabulary begin?. *System*, *23*(1), 35-41.

Nation, P., & Waring, R. (1997). Vocabulary size, text coverage and word lists. *Vocabulary: Description, acquisition and pedagogy*, *14*, 6-19.Praninskas, J. (1972). *American university word list*. Longman Group Limited.

Nation, P. (2014). How Much Input Do You Need to Learn the Most Frequent 9,000 Words?. *Reading in a Foreign Language*, *26*(2), 1-16.

Omoniyi, T., Hyland, K., & Paltridge, B. (2011). Discourse and identity. K. Hyland & K. Paltridge, The continuum companion to discourse analysis, 260-276

Pauwels, P. (2017). Making and Using Word Lists for Language Learning and Testing. Nation, ISP (2016). *ITL-International Journal of Applied Linguistics*, *168*(1), 134-139.

Richards, J. C. (1974). Word lists: problems and prospects. *RELC journal*, *5*(2), 69-84.

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

Todd, R. W. (2017). An opaque engineering word list: Which words should a teacher focus on?. *English for Specific Purposes*, *45*, 31-39.

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.

Vongpumivitch, V., Huang, J. Y., & Chang, Y. C. (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*(1), 33-41.

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

West, M. P. (1953). *A General Service List of English Words. With Semantic Frequencies and Asupplementary Word-list for the Writing of Popular Science and Technology. Compiled and Edited by M. West.(Revised and Enlarged Edition.)*. Longmans, Green & Company..

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language learning and communication*, *3*(2), 215-229.