Journal of Modern Research in English Language Studies

Print ISSN: 2676-5357
Online ISSN: 2676-5985

A High Frequency Word List for Political Sciences

Reza Bagheri Nevisi[1]*, Mahmood Safari[2], Rasoul Mohammad Hosseinpur[3], Reyhaneh Sadat Mousakazemi[4]

[1*]Associate Professor of Applied Linguistics at University of Qom, Iran, *re.baghery@gmail.com*
[2]Assistant Professor of Applied linguistics at Hazrat-e Masoumeh University, Qom. mahmood.*safari@gmail.com*
[3] Associate Professor of Applied Linguistics at University of Qom, Iran *rmhosseinpur@gmail.com*
[4]MA Graduate at University of Qom. *mkreyhane@gmail.com*

Article Info

Article Type:

Research Article

Received:

22/12/2022

Accepted:

11/03/2023

ABSTRACT

Recent research favors specific academic wordlists over a general academic wordlist for preparing university students to read and publish academic papers in English. Although researchers have developed wordlists for various disciplines, some academic fields do not enjoy a well-developed technical wordlist. The present study aimed at developing and evaluating a specific academic wordlist for political sciences. A 3.5-million-word corpus of political sciences papers created and analyzed in order to develop the Politics Academic Word List (PAWL). The list included 2000 word families which were selected across and beyond the BNC/COCA wordlist based on frequency and range criteria. The word families enjoying an aggregate frequency of a hundred or more in the corpus and a minimum frequency of 10 in at least four of the seven sub-corpora were incorporated into the wordlist. The PAWL accounted for over 88% of the running words in the Politics Academic Corpus (PAC) and outperformed the list of General Service List (GSL) plus Academic Word List (AWL) words in coverage by 3 percent, despite containing 556 fewer word families. The study corroborates the value of a subject specific wordlist as a more fruitful source for academic vocabulary learning. Pedagogical implications and suggestions for further research are discussed.

Keywords: Political Sciences, Academic Word List, English for Specific Purposes

Cite this article: Bagheri Nevisi, R., Safari, M., Hosseinpur, R. M., & Mousakazemi, R. S. (2023). A High Frequency Word List for Political Sciences. *Journal of Modern Research in English Language Studies, 10*(4), 21-43.

DOI: 10.30479/jmrels.2023.18213.2161

1. Introduction

Most graduate learners and university professors are required to submit academic papers for graduation and promotion. They need to read and write scholarly papers in prestigious journals and the majority of respected high-ranking journals are international journals that are published in English. Therefore, graduate students and academicians need to be proficient enough in English to be able to publish their research findings in scholarly international journals. One of the primarily daunting dimensions of second language acquisition, especially for EAP language learners, is to learn the vocabulary of the second language. Unlike second language sounds and grammar structures, which are somehow limited and learnable in a specific period, word number in a language is too enormous to be acquired in a given time. On the other hand, EAP learners have their technical courses and have a limited time for second language learning. The precious time in EAP classes must be spent meticulously. An important aspect of teaching vocabulary in EAP is to decide on what lexical items we should concentrate on (Coxhead, 2000).

One of the predominant difficulties for EAP students is to learn academic words, the words which are used frequently across academic disciplines (Chen & Ge, 2007). Although students at tertiary levels might experience fewer challenges with technical words of their specific disciplines, since they are typically more commonly and globally applied or could be figured out and ascertained from shared word origins (Hutchinson & Waters, 1987), academic words present a grave difficulty for university students, especially graduate students who aim at reading and writing academic papers within their specific fields. Li and Pemberton (1994) reported that territory students had little problem with discipline-specific technical vocabulary in writing academic essays, but they found academic vocabulary problematic.

As a result, researchers have long been struggling to determine the vocabularies of paramount importance to be presented to EAP students and a number of vocabulary list have been made and offered to EAP students. University students have widely been recommended to master West's (1953) GSL and Coxhead's (2000) AWL to recognize and employ a large number of vocabularies that are frequently used in their specific domains. Some scholars are of the opinion that the two word lists can present a word coverage of more than 86% in every academic discipline (e.g., Coxhead, 2000). Adding the subject-specific technical words will bring about a large coverage of over 90% of running words, which is a rather acceptable coverage. Accordingly, many textbooks and instructional materials have been developed to instruct EAP learners from both lists.

However, other researchers have reported far lower coverage for the two lists over various academic corpora, have identified several highly frequent vocabularies that are nonexistent in the two lists and have found a number less frequent GSL and AWL words in their academic corpora. That is, they have challenged the value of the two lists as a source for academic vocabulary learning (e.g., Esfaniari & Moein, 2015; Safari, 2004; Yang, 2015). Many researchers have embarked on developing technical word lists for different academic fields and have shown that subject specific word lists are more favorable over a single academic word list. However, in spite of plenty of studies working on and creating vocabulary lists for various scholarly domains (e.g., Hsu, 2018; Lei & Liu, 2016; Martinez, et al., 2009), some academic fields have gone unnoticed and are still under-explored. This research intended to explore the scholarly texts of political sciences and develop a technical word list for it.

2. Literature Review

2.1. Academic Words

Words have been sub-classified into four subcategories: Highly frequent vocabularies, less frequent vocabularies, technical and academic ones. Highly frequent vocabularies are commonly used in basic language use such as every day conversation and common reading and writing. Less frequent vocabularies are rarely used in general and academic texts. They may occur once or twice in each text but the list of low frequency words includes a large number of words (Nation, 2001). There is specificity inherent in topics and sub-disciplines of technical vocabularies and they are rarely used or almost non-existent in other fields. Finally, academic words, otherwise known as semi-technical ones, can be regarded as less informal, less context-dependent vocabularies that are highly frequent and widely used across various academic fields and might be less typically found in general English courses (Farrel, 1990).

Researches have mostly focused on highly frequent and academic words. Less frequent words are not worth learning due to their large number and low coverage and technical words are mostly known for students and researchers in every field. Technical words account for almost 5% of running vocabularies in scholarly materials (Nation, 2016). However, general English and academic words cover a remarkable percentage of running vocabularies in every scholarly text. West's (1953) GSL is reported to have an approximate 65% to75% vocabulary coverage of running words in tertiary materials (Khani & Tazik, 2013; Valipouri & Nassaji, 2013) and Coxhead's (2000) AWL enjoys an approximately 10 % vocabulary coverage in materials at tertiary levels (Li & Qian, 2010).

2.2. Vocabulary List Types

There have been three kinds of academic word lists: Field-specific vocabulary lists, general vocabulary lists and subject matter-specific vocabulary lists. GAL include highly frequent words in all academic disciplines. Coxhead's (2000) AWL incorporates the most important general vocabulary list for academic purposes, comprising of 570 word families and is reported to have an approximate 10% vocabulary coverage of running words in scholarly materials. There are two other general vocabulary list used for academic purposes: Academic Word List (Gardner & Davies, 2014) and New Academic Vocabulary List (Browne et al., 2013). Field-specialized vocabulary lists concern the words that are highly frequent in a specific discipline (i.e., a broad group of related academic areas) such as sciences. An example for such word lists is Coxhead and Hirsh's (2007) EAP Vocabulary list. Specificity in subject matters in vocabulary lists includes technical vocabularies that are highly frequent in specific academic subject. The focus is on common vocabulary in a specific subject matter like medicine (Hsu, 2013), engineering (Ward, 2009), chemistry (Valipouri & Nassaji, 2013), agriculture (Martinez, et al., 2009), nursing (Yang, 2015). Recently, the focus of academic word studies has shifted from developing general academic word lists to creating word lists for more specific academic domains.

Different academic word lists have taken different views on high frequency words. Some researchers have assumed that students are familiar with these words and have come up with lists of academic words beyond a list of general English words. Coxhead (2000) took West's (1953) GSL as a baseline and excluded GSL words from her list of academic vocabulary and Hsu (2013) developed her academic word list, selecting highly frequent academic vocabularies outside the British National Corpus (BNC) 3000-word list. However, some researchers have criticized the commonly held view that an academic word list must be developed on the basis of a list of general English words. Ward (2009) rejected the idea that students require an approximate 3000 word-family lexicon to embark upon reading efficaciously and that this word knowledge should be first founded upon a general vocabulary list and then be reinforced by an academic and/or technical vocabulary list. Ward suggested that there might be no necessity for learners to begin with a quite distinctive general word list and students with clearly defined objectives might be able to proceed with the most frequent vocabularies in their specialized fields. Hence, some researchers have investigated the frequency and range of GSL words in their corpora and have excluded low frequency GSL words from their academic word lists (Esfandiari & Moeini, 2015; Ward, 2009).

2.3. Empirical Studies on Academic Vocabulary

2.3.1 Early Academic Word Lists

Since early 1970s, there has been a plethora of research attempting to identify highly frequent vocabularies at tertiary levels and create academic vocabulary lists. Lynn (1973) and Ghadessy (1979) generated word lists for the annotations at tertiary level to enhance students' academic writings. Xue and Nation (1984) mixed and modified the four above-mentioned lists and created their University Word List, consisting of 840 word families and is reported to enjoy an 8.5%-word coverage of running words in academic materials whereas only a 1.7%-word coverage existed for fiction texts. Coxhead (2000) asserted that UWL did not possess steady and dependable choosing of fundamentals and was inundated with several flaws of previously done research. Working on a 3.5-million academic corpus, Coxhead created AWL, comprising of 570 word families and enjoying an approximate 10% vocabulary coverage of running words.

AWL was developed on top of West's (1953) GSL that consisted of 2000 highly frequent words in general English texts. The two lists (i.e., GSL and AWL) are reported to have an approximate 85%-word coverage of running words in academic materials. Therefore, they have been widely recommended for students' academic vocabulary learning. Nevertheless, criticisms of size limit, time limit and coverage restrictions have been leveled against the GSL and the AWL has also been blamed for its partiality, lack of timeliness and restricted size. The GSL was developed in 1953 by Michael West and the texts were even older. The second 1000 GSL words has very small coverage over general and academic texts and can be ignored. Coxhead's Academic Corpus was unfairly influenced by a number of academic fields (Coxhead, 2000; Hyland & Tse, 2007). Moreover, some textual information in Coxhead's Academic Corpus were taken from LOB Corpus and Brown Corpus, which dated back to 1970s. In addition, the size of the corpus was not sufficient either, each of the 28 sub-corpus containing 125 thousand running words.

2.3.2. Recent Academic Word Lists

Recent corpus-based studies on academic vocabulary focus mainly on more specific academic areas. There have been plenty of studies developing disciple-specific and subject-specific academic word lists for various academic fields. Some of these studies are reported below.

Drawing on Coxhead's (2000) AWL, Martinez et al. (2009) identified the academic vocabularies in an agricultural corpus of research papers. The quantitative analysis resulted in a very limited word list from AWL, 92 families. The results pointed to the needs to generate field-specific academic word lists.

Hsu (2013) developed a vocabulary list of medicine to reduce the divide between non-specialized and specialized words. The corpus consisted of 155 books with 31 medical themes gathered from e-book databases accounting for a total number of 15 million running words. Frequency and range of words beyond the most common 3,000-word word families were thoroughly investigated and 595 of the highly frequent word families were finally selected and constituted the MWL.

Muñoz (2015) probed into a corpus of 700 agricultural research papers in English. The researchers employed a mixed-methods approach and the results pointed to high lexical variability in the corpus and low word range. Academic vocabularies covered only 6% compared to the 10-12% coverage reported for academic papers but the coverage was higher compared to newspapers 4% coverage.

Yang (2015) delved into the most commonly employed nursing academic words within different nursing subfields. A 1,006,934-word corpus including 252 English scholarly papers of nursing was collected. NAWL including the most commonly employed nursing vocabularies was extracted. The list entailed 676 word families accounting for nearly 13.64% of the nursing corpus. The results demonstrated that it is important to produce domain-specific vocabulary list at tertiary levels for EFL nursing learners to solidify their overall academic language proficiency.

Tongpoon-Patanasorn (2018) created a technical word list in the finance field through taking up a hybrid method. The list included 979 finance-related words that were sub-classified into 569 word families. Such vocabularies were listed in both GSL (413) and AWL (291) words.

Safari (2019) investigated highly common vocabularies in equine veterinary academic scholarly papers. A 3.6 million corpus of running words was examined by some text analysis software. The findings suggested that 1091 general and 116 academic word families were less frequent in the Equine Veterinary Corpus (EVC). Furthermore, 214 non-technical word families were commonly used in the EVC but were non-existent in both. The resulting list enjoyed a 2.5% higher coverage than the list of GSL and AWL words together, despite containing 993 fewer words.

Heidari, et al (2020), working on a 3.45-million-word corpus of Pharmacy, determined the highly frequent words in pharmacy discipline and developed their Pharmacy Academic Word List (PAWL), which consisted of 750 word families and covered 17.69% of the tokens. The results corroborated the need to gather field-specific vocabulary lists to cope with the requirements of EFL/ESL practitioners and postgraduates over different domains.

Jamalzadeh & Chalak (2019) developed their list of physiology academic word list through examining a 1.7 million-word corpus. 1450 commonly-used word families were extracted and made up the Physiotherapy vocabulary list. The results indicated that AWL might not be completely beneficial to such students due to its low vocabulary coverage and restricted application of commonly used physiotherapy vocabularies.

Despite the development of these academic vocabulary lists in above-mentioned fields, no list has specifically addressed political science. Such word lists can be helpful to both graduate students of political science and their instructors. This research aims to develop such a list to serve as a reference point for ESP educators and material developers in English for Political Purposes, to provide EAP practitioners with more evidence to generate domain-specific or specialized vocabulary lists and to expedite learners' political science's learning of academic vocabulary. Such specialized word list aimed primarily at students of political science can be instructed and thoroughly examined similar to lexical items from GSL and AWL. Accordingly, the researcher formulated the following research questions to achieve the above-stated objectives of the research:

Q1: What high frequency words, across and beyond BNC/COCA 25,000-word list, make a politics academic word list (PAWL)?

Q2: Which GSL and AWL words are not frequently applied in political science papers? What none-GSL and none-AWL words are the most frequently used ones in political science texts?

Q3: How does the coverage of PAWL over a politics academic corpus (PAC) compare to that of the list of GSL and AWL words together?

3. Methodology

3.1. Politics Academic Corpus (PAC)

To ascertain highly frequent vocabularies in political science texts and develop a *politics academic word list* (*PAWL*), the researchers were required to develop a corpus of political science texts. The corpus was composed of research articles from seven political science subfields (*comparative politics*, *international relations*, *political economy*, *political methodology*, *political theory*, *public administration*, *public policy*). Two university professors of political sciences were consulted for the main sub-disciplines of political sciences and they agreed on the sub-disciplines in the above-mentioned list as the major areas in political sciences. Furthermore, the websites of some universities offering political sciences majors at MA and PhD levels and some political sciences journals were investigated to confirm the list. The

incorporated papers were published in some leading political science research journals during the years between 2010 and 2020. The corpus, which was named *Politics Academic Corpus* (*PAC*), comprised over 3.5 million running words and included seven sub-corpora of almost the same size, each relating to one political science subfield. Table 1 displays the information on the corpus size and each sub-corpus, and the number of research articles in the corpus.

Table 1.
Size of the Sub-corpora, and Number of Research Articles

Sub-corpora	Size	Number of Articles
Comparative politics	500,790	89
International relations	499,323	105
Political economy	501,038	83
Political methodology	499,846	72
Political theory	500,688	108
Public administration	500,191	69
Public policy	498,986	98
Total	3,500,862	624

3.2. Analysis Software

Some text analysis software was required to delve into the corpus to discover the most frequent words in and evenly spread across various political science subfields. The major software employed in the study was Range, which was developed by Heatly, et al. (2002) and is available, free of charge, at https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/vocabulary-analysis-programs. The researchers used the version of the software that has the 25,000 word families from the BNC/COCA as its base word lists. The software analyzes a corpus and lists all the words. BNC/COCA words and word families, alongside their frequency, are presented at the top of output file and the other words are listed below. The words are listed according to the size of their frequency, higher frequency words appearing at the top of the lists. The word families are listed according to their aggregate frequency. Therefore, researchers can identify less and more frequent words and word families conveniently. Moreover, the software can analyze several corpora, or sub-corpora, at the same time and, hence, compare the frequency of each word in different corpora. Checking the frequency of each word in all the analyzed corpora, researchers can identify words that are frequently used in most corpora and evenly spread in various fields or subfields.

The second software utilized in the study was the Excel program, which was used to compare the words in the newly developed list, Politics Academic Word List (PAWL), and the GSL-AWL (i.e., the list of GSL and AWL words together), in order to find words that are common or specific to the lists. The software was also used to compute the PAWL's aggregate frequency and its coverage over the political science corpus.

3.3. Word Selection Criteria

Three criteria were set to decide upon the words to be incorporated in the political science word list. They were *frequency*, *range* and *word family*. The corpus which was used in the present study was almost the same size (i.e., 3.5 million running words) as Coxhead's (2000) Academic Corpus, so the researchers decided to set the frequency criterion at 100 occurrences. That is, each word family needed to have an aggregate frequency of 100 or above to be incorporated in the politics word list. As for the range criterion, each word family was required to have a minimum frequency of 10 in at least five of the seven sub-corpora. It was identical to Coxhead's (2000) range criterion, a minimum frequency of 10 in each of the four corpus sections and occurring at least in 14 out of 28 sub-disciplines. Many studies have employed similar range and frequency criteria (e.g., Esfandiari & Moein, 2015; Valipouri & Nassaji, 2013; Wang et al, 2008; Yang, 2015). Finally, different units of counting, such as word type, lemma and word family, are used in corpus-based word list studies, among which word family has been the most frequently employed unit (Dang, 2020). The authors used word family as the word counting unit in the present study, as the coverage of a list consisting of word families over a corpus is much higher than that of a list composed of the same number of lemmas and word types. Moreover, university students are mainly familiar with the most important English suffixes, which are used to develop word family members in the BNC/COCA word list.

3.4. Data Collection and Analysis

Several steps were taken in order to develop the political science word list and evaluate it. First, the researchers had to develop a corpus that was a representative sample of political science academic texts. To that end, they initially consulted some political science university professors and scientific internet sites, such as webpages of universities offering political science courses and seven major politics subfields were selected as a sample representing political science. The selected subfields were *comparative politics*, *international relations*, *political economy*, *political methodology*, *political theory*, *public administration*, *public policy*. Then, some leading scholarly research journals publishing political science articles were identified. The articles that were published in the years between 2010 and 2020 were downloaded to be incorporated in the corpus. The articles were in

different computer formats such as PDF, HTML, WORD, but as the employed software processes only TEXT files, all the files were converted to TEXT format. Moreover, the *affiliations*, *references*, *appendices*, *acknowledgements* and *biodata* were removed from the research articles so that the corpus only enjoyed pure political science scholarly research papers. The cleaned manuscripts of the research articles were added together to develop 7 subcorpora of almost equal size (around 500,000 running words), which constituted the *Politics Academic Corpus* (PAC).

Subsequently, the *Range* software was run over the corpus and most common words and evenly spread ones in the seven subcorpora were identified. The BNC/COCA word families that enjoyed a total frequency of 100 in the PAC and a minimum frequency of 10 in five or more subcorpora were worked out to be incorporated in the political science word list (i.e., PAWL). Checking the frequency and range of the word families, the researchers noticed some words which just failed to meet the criteria; that is, their frequency and range were just below the set criteria. Some of these words seemed to be important general English or political science words (e.g., *absorb*, *alien*, *damage*, *enemy*, *logistics*, *paragraph*, *voluntary*). Therefore, the words with 85-99 overall occurrences and had a frequency of 10 in four sub-corpora were further investigated to choose the acceptable words to be incorporated in the PAWL. The researchers used objective (the words' range and frequency values) and subjective evaluations in selecting the right words. For subjective evaluations, three PhD students of political sciences were consulted in selecting words which just failed to meet the frequency and range criteria (i.e., the words whose total frequency was 90 to 99 or the ones which occurred 10 or more times in 4 sub-disciplines) but seemed to be important in political sciences. In addition, the researchers identified the words which were beyond the BNC/COCA word list but met both criteria (i.e., 100 occurrences or more in the total corpus and at least 10 occurrences in five or more sub-corpora) and incorporated them in the PAWL.

4. Results and Discussion

4.1. Developing the PAWL

There were 1858 BNC/COCA word families, which had an aggregate 100 occurrences or more in the corpus and a minimum frequency of 10 in at least five sub-corpora. Most of these words are general English words and can be found in general service lists of words, such as BNC/COCA word list. Over 90% of these words were from the first three BNC/COCA base word lists, BNC/COCA list consisting of 34 base word lists. Table 2 displays the baseword lists which contributed the greatest number of words to the PAWL, alongside the number of word families they contributed.

Table 2.

BNC/COCA Base Word Lists Contributing Most Words to PAWL

Base word List	Number of words in PAWL
Base word list 1	623
Base word list 2	439
Base word list 3	616
Base word list 4	186
Base word list 31	91
Base word list 5	75
Base word list 6	39

However, several highly frequent vocabularies were associated with politics. For instance, among the top 100 highly frequent vocabularies in the corpus, there were 25 lexical words, five of which are especially related to political science (*public*, *state*, *power*, *world*, *party*) and some partly associated with politics (e.g., *system*, *control*). And in the second top 100 words, there were 69 lexical words, 10 of which are especially or partly associated with politics and expected to be found in political science texts (*country*, *local*, *war*, *press*, *history*, *rule*, *rights*, *market*, *view*, *tax*). Table 3 demonstrates the top fifty lexical words in the political science corpus. As it is evident, all of the words are general service vocabularies and some are general English vocabularies related to and highly expected in political science texts.

Table 3

Top Fifty Lexical Words in the PAWL

1- Public	*1- World*	*1- Level*	*1-Country*	*1-Group*
2- State	*2-Party*	*2- Table*	*2-Order*	*2-Part*
3- See	*3-Science*	*3- Number*	*3-Set*	*3-Rights*
4- New	*4-System*	*4- Control*	*4-Make*	*4-Human*
5- Two	*5-Important*	*5- Particular*	*5Question*	*5-Market*
6- Use	*6-Support*	*6- Way*	*6-War*	*6-Like*
7- Power	*7-Change*	*7- Interest*	*7-General*	*7-Possible*
8- Different	*8-Case*	*8- People*	*8-Figure*	*8-Rate*
9- First	*9-Study*	*9- Second*	*9-Point*	*9-Need*
10- Time	*10-Work*	*10- Local*	*10-Subject*	*10- Problem*

In the second step, the words that met the frequency and range criteria but were not present in the BNC/COCA vocabulary lists were recognized by checking word frequency and range beyond the BNC/COCA list. Nineteen non-BNC/COCA word families were found to have 100 occurrences or more

in the corpus and at least ten occurrences in five sub-corpora. Some examples are: *cross-national*, *institutionalism*, *nuance*, *socio-economic*, *subfield*.

Checking the frequency and range of the words to be incorporated in the PAWL, the researchers noticed some words whose frequency and range were just below the set criteria but seemed to be important words in political science texts. Thus, the researchers scrutinized the words which just failed to identify the ones which were worth to be incorporated in the PAWL. The word families with 85-99 occurrences and those having a least frequency of 10 in four of the subcorpora were further evaluated, objectively and subjectively, in order to select the more important ones. The researchers selected 122 word families from among these words. They were 76 word families that enjoyed an adequate dispersion (i.e., range) across the subcorpora but occurred just less than 100 times in the PAC, and 46 word families with 100 occurrences, but a least frequency of 10 in four subcorpora, their frequency in the other three sub-corpora being around 10. Table 4 represents some examples for the word families whose frequency and range were just below the set criteria. The first six word families had less than 100 occurrences in the PAC but had an adequate dispersion across the subfields, and the next four word families had an adequate aggregate frequency but failed to have a minimum frequency of 10 in five or more sub-corpora. As it is shown, the frequency and range of these word families were very close to the set criterion values. Moreover, some of the evaluated words seemed to be important words in political science and were selected to be included in the list. Some examples are *cabinet*, *export*, *manifesto*, *newspaper*, *partiality*, *personnel*, and *senate*.

Table 4.

Word Families Just Failing to Meet the Frequency and Range Criteria

Word	Range	Frequency	Sub1	Sub2	Sub3	Sub4	Sub 5	Sub6	Sub7
Word	Range	Frequency	Sub1	Sub2	Sub3	Sub4	Sub 5	Sub6	Sub7
	File	7 99	19	18	4	19	20	16	3
	Host	799	13	19	8	16	11	12	20
	Decisive	7 99	10	16	22	20	13	8	10
	Couple	798	8	20	12	17	15	9	17
	Absorb	797	12	18	13	11	29	8	6
	Solid	797	20	13	11	17	14	15	7
	Deliberate	7	320	10	17	258	10	9	9	7
	Resemble	7	105	12	8	17	32	9	9	18
	Periphery	7	264	9	9	9	101	97	13	26
	Refine	7	104	12	16	13	36	9	9	9

Putting all these words together, the researchers came up with 2000 word families which were highly frequent and evenly spread in various political science subfields and formed PAWL. The list is highly advisable for political science students, as it contains the most important words in their academic texts. The list is considerably smaller than GSL and AWL together (2570 word families) and is expected to be more beneficial for political science students.

4.2. Words common/specific to the PAWL and GSL-AWL

After developing the PAWL, the researchers compared and contrasted vocabularies in the PAWL and GSL-AWL lists to find out which words are common to the two lists, how many and which GSL and AWL vocabularies were less frequent in political sciences and thus absent in the PAWL, and what vocabularies were frequently used in political science texts but non-existing in both. Table 5 displays the number of words which were common to the two lists and those which were specific to each. As shown in the table, 1367 GSL-AWL word families were shared by the PAWL, indicating that only 53% of the GSL and AWL words were frequently used in political science articles. Moreover, 1189 GSL-AWL word families were not commonly used in political sciences; that is, around 47% of GSL-AWL words were less frequent in politics. Moreover, only 213 vocabularies in the second 1000 GSL words were present in the PAWL. This revealed that around 80% of the second 1000 GSL word list was not commonly used in political sciences. However, the AWL and first 1000 GSL word list have more word families in common with the PAWL.

Table 5.

Number of Words Common/Specific to the PAWL and GSL-AWL

Word List	Number of word families	Number of words absent in PAWL	Number of words present in PAWL
General Service List	1986	1049	937
First 1000 words	998	274	724
Second 1000 words	988	775	213
Academic Word List	570	140	430
Total	2556	1189	1367

Around 72.5% of the first 1000 GSL words and 75% of the AWL words were present in the PAWL. Finally, there existed 633 word families that were frequently-used in the political science corpus but non-present in both general and academic lists. Table 6 displays some example words for the less frequent

general and academic words in the PAC and the most frequently used vocabularies in political sciences but non-existent in the GSL-AWL. As the example words indicate, low frequency GSL and AWL words are expected to occur less frequently in political sciences while high frequency non-GSL-AWL words are highly associated with political science topics and are highly expected in politics texts

Table 6.

Example Low Frequency GSL-AWL and High Frequency Non-GSL-AWL Words

Word families	Example words
Low Frequency GSL Words	
First 1000 words	*bread, corn, egg, flower, glad, lip, mouth, salt, temple, vessel*
Second 1000 words	*ash, button, dip, hurray, leaf, mud, pig, shower, wax, yard*
Low Frequency AWL	*adjacent, bulk, chemical, clause, commence, grade, injure, insert*
High Frequency Non-GSL-AWL	*agenda, crisis, democracy, download, engage, inflation, reform, scholar*

4.3. Coverage of the PAWL and GSL-AWL lists

The last step in the study was to compare the coverage of PAWL with that of the GSL-AWL word lists. The coverage of each list was computed (i.e., 3.5 million). Table 7 reveals the coverage of the lists over the Politics Academic Corpus (PAC). As it is shown, the PAWL accounted for 88.05% of the running words, while the GSL-AWL's coverage was 85.12%. The PAWL's coverage of the PAC was around 3 percent higher than that of such lists, though the previous one contained 556 fewer word families. Therefore, replacing the 1189 general and academic word families that were less frequent in political sciences and absent in the PAWL with the 663 non-GSL-AWL word families that were the most commonly-applied in political sciences brought about an increase in the coverage of politics by almost 3% and a decrease of 556 word families in the number of words in the political science word list. Moreover, the second 1000 general vocabularies accounted for just 4.63% of the PAC, which is a rather small coverage, and the first 1000 GSL words accounted for 67.67% of the running words in the PAC. As structure words had more than 51% coverage of texts (Kucera & Francis, 1967, as cited in Bowen et al, 1985), leaving out the coverage of structure words, the lexical words of the first 1000 general vocabularies would approximately cover 16% of the running vocabularies in the political science corpus that was not an enormous coverage. However, the AWL word families constituted 12.82% of the running words in the Politics Academic Corpus that was much more than that of Coxhead's (2000) Academic Corpus (i.e., 10%).

Table 7.

Coverage of the PAWL and GSL-AWL over the PAC

Word List	Number of word families	Coverage of the PAC
PAWL	2000	88.05%
GSL-AWL	2556	85.12%
GSL	1986	72.30%
1st 1000 words	998	67.67%
2nd 1000 words	988	4.63%
AWL	570	12.82%

4.4. Discussion

The research intended to develop and evaluate a word list for political sciences, therefore; the researchers analyzed and developed a corpus of political sciences texts.

4.4.1. The Politics Academic Word List (PAWL)

The first research question was aimed at finding words across and beyond BNC/COCA wordlist which are highly frequent and evenly spread in political sciences texts to develop a specialized vocabulary list for political sciences. The analysis of the corpus and its seven subcorpora revealed that 2000 words, namely 1981 families across the BNC/COCA baseword lists and 19 word families beyond the BNC/COCA, had the necessary commonness and range to be incorporated in the intended world list. The analysis further revealed that although most of the words were from top BNC/COCA base word lists, many of the words in the top BNC/COCA base word lists were not highly frequent in political sciences texts. Moreover, there were many words in the middle or bottom BNC/COCA base word lists and some beyond BNC/COCA which frequently occurred in the politics corpus. This suggests that there is a political sciences technical wordlist, which is different from such lists. Thus, the students of political sciences do not need to master all the words in BNC/COCA list or even the words in the top base word lists to be to effectively study scholarly works within their specific field. Learning all BNC/COCA words or even the top BNC/COCA base word lists will take a very long time and is not worth the effort. In addition, there are some words beyond BNC/COCA, which students need to learn for an effective L2 communication within their specific field. The developed wordlist could help students and researchers of political sciences to avoid wasting their time learning words that are of less value and utility for them and help them master

the words that are truly common in political sciences and useful for reading and writing academic manuscript.

Results are congruent with previous research that were indicative of the fact that not all BNC/COCA words were commonly applied in scholarly works of various university disciplines (e.g., Coxhead, et al, 2016; Hsu, 2018; Lu & Coxhead, 2020). Hsu (2018) divided BNC/COCA words into highly frequent, mid frequent and less frequent words in Traditional Chinese Medicine (TCM) texts and identified 605 highly frequent BNC/COCA word families to develop a TCM word list. Coxhead et al indicated that many words in BNC/COCA list do not occur frequently in carpentry academic texts.

4.4.2. The PAWL versus the GSL-AWL List

The second research question tried to juxtapose the word families in the PAWL and the GSL-AWL lists to ascertain how many and which GSL and AWL words were commonly used in political sciences texts and what words were frequently-used in political sciences but non-present in both. The aim was to find commonalities and discrepancies between these two lists and a specific academic word list (i.e., the PAWL). University students have widely been recommended to gain mastery over such words to grasp and apprehend scholarly information within their specific disciplines. This research question attempted to investigate the validity of this recommendation by comparing the GSL-AWL words to those of a specific technical word list. The analysis of the PAC indicated that 1367 GSL and AWL word families (937 GSL and 430 AWL word families) were highly frequent and evenly spread in political sciences and common to the PAWL and GSL-AWL lists. However, 1189 GSL-AWL words (1049 GSL and 140 AWL word families) failed to meet the needed frequency and range in the political sciences corpus. In fact, almost half (i.e., 47%) of the GSL-AWL words were shown to be less frequently used in political sciences.

Around 53 percent of the GSL word families are of less frequently used in the political sciences. This implies that the GSL is not an appropriate prerequisite for vocabulary learning by politics students. Learning over 1000 GSL word families which are not commonly used political sciences texts would save practitioners' time and energy interested in enhancing their academic reading and writing. A well-developed political sciences technical word list would simply exclude these word families. The finding is congruent with many reseach that have challenged the value of the GSL as a prerequisite for academic vocabulary learning (e.g., Esfandiari & Moein, 2015; Hwang, 1989; Moini & Islamzadeh, 2006; Safari, 2018; Sutarsyah, 1993). Engels (1968) demonstrated that the second 1000 GSL words were of little importance for vocabulary learning due to its low coverage. The second 1000 GSL words had also a low coverage (4.27%) over the academic corpus

developed by Moini and Islamzadeh, who proposed that to master the second 1000 general vocabularies prior to academic ones is not required. The researchers excluded 1342 GSL words from their linguistics word list as they did not have the required frequency in the corpus of linguistic articles. Finally, only 740 GSL words were commonly used in the Food Science and Technology Corpus developed by Esfandiari and Moin (2015), suggesting that only one-third of the GSL word list is worth learning by students of food science and technology.

As for the AWL, 140 word families did not occur frequently and evenly enough in the corpus of political sciences to be included in the PAWL. This indicates that students of political sciences do not need to master all the words in Coxhead's AWL to be able to read and/or write manuscripts in the discipline. Almost one-fourth of the AWL words are not commonly used in political sciences texts. The students can invest their time and energy in learning other more relevant and important words instead of learning these words. This finding corroborates the previous research findings, which indicated that many AWL vocabularies are not commonly used in various university fields (e.g., Chen & Ge, 2007; Moinin & Islamzedeh, 2016; Safari, 2018; Valipouri & Nassaji, 2013; Yang, 2015). Chen and Ge (2007) revealed that over 270 AWL word families are not frequently used in medical texts, and stated that AWL is incomplete in presenting the most commonly used lexical items in medicinal sciences. Furthermore, 170 AWL words were absent in Yang's list of high frequency nursing words and 189 AWL word families were excluded from Moini and Islamzedeh's linguistics word list. Finally, Safari (2019) identified 116 AWL word families that were of low frequency in veterinary research articles.

Furthermore, the analysis of the PAC revealed 633 lexical items that were frequently used and evenly spread in the corpus but non-existing in both AWL and GSL. This suggests that mastering GSL-AWL words would not adequately prepare political sciences students to read and write academic manuscripts within their specific discipline, as there are a large number of words beyond GSL-AWL commonly employed in political sciences texts. Thus, the two lists would not make an advisable sufficient source for political sciences students' vocabulary learning. In fact, one-third of the words that learners require to go through scholarly works (i.e., 633 word families out of the 2000) is not catered by the GSL and AWL lists. The students need to master these words for effective performance in their academic field. A word list including these words will certainly cover a larger proportion of words in political sciences texts and will benefit the students far more than the GSL and AWL together. The words are some general or common academic words which were excluded from the GSL and AW due to the inadequate corpora used by the two researchers. The finding confirms the research results which

introduced lists of highly frequent non-GSL/non-AWL words in various academic disciplines. Vongpumivitch et al (2008) identified 128 sub-technical lexical items that were commonly used in applied linguistics but not present in the AWL. There were 278 non-AWL high frequency academic words in Yang's (2015) nursing academic corpus. Moreover, Jamalzadeh and Chalak (2019) came up with 406 non-technical words outside GSL-AWL which are commonly applied in physiotherapy scholarly papers. Other studies have reported similar lists of high frequency non-GSL, non-AWL words in various academic disciplines (e.g., Esfandiari & Moein, 2015; Moini & Islamzadeh, 2016; Safari, 2018). These findings suggest that the GSL is not an adequate prerequisite for academic vocabulary learning and the AWL is not a genuine general academic list. The presence of *estate*, l*evy*, *subsidy* and the absence of *characteristic, cognitive*, *laboratory* in the AWL corroborates the list's bias for and against some academic disciplines. The list covered only 6% of the running words in agriculture, biology and medicine corpora (Cobb & Horst, 2004; Hyland & Tse, 2007; Munoz, 2015).

4.4.3. Coverage and size of the PAWL versus GSL-AWL Lists

Regarding the third research question, which compared the size and coverage of the PAWL to those of the GSL-AWL list, the results of the analyses indicated that the PAWL was much smaller than the GSL-AWL, including 556 fewer word families, while its coverage over the political sciences corpus was 3% larger than that of the GSL-AWL list. Thus, the PAWL is preferable over the combination of GSL and AWL words, as a reference point for political sciences learners. The students spend shorter time for mastering these words but gain a better knowledge of political sciences vocabulary. The finding confirms the results of previous relevant research studies (e.g., Esfanidiari & Moein, 2015; Jamalzadeh & Chalak, 2019; Safari, 2018; Wang, et al, 2008; Yang, 2015). The coverage of Safari's (2018) psychology word list, containing 1587 word families, over his psychology corpus was 2.2% larger than that of the GSL-AWL list, which included 2570 word families. Esfandiari and Moein's academic vocabulary list contained 13% of the words in food science and technology corpora, while the AWL's coverage was only 8.8%. The coverage of Yang's nursing academic word list over her corpus was larger than that of the AWL by more than 3 percent. Jamalzadeh and Chalak's list of 406 non-GSL-AWL words covered 24.66% of their Physiography corpus, while AWL, containing 570 word families, covered 11.51% of the running words.

5. Conclusion

Most EFL/ESL students find it challenging to learn academic vocabulary, specifically if the goal is to attain an acceptable literacy level in the second or foreign language (Cobb & Horst, 2004). Some post-graduates are obliged to write their theses in English without taking into account their specialization. Furthermore, some academic institutions even go further than that and force their graduates to publish papers in journals (Yang, 2015). Accordingly, students of political sciences who are offered their pertinent courses and texts in English might find it a daunting task to comprehend and write well for academic purposes in the target language since they are not familiarized with such technical vocabularies in English

A political academic word list might play a pivotal part in establishing vocabulary-learning objectives, boosting students' autonomy, and assisting course designers to design better pedagogical materials, choose the right passages and create appropriate instructional tasks. The word families discovered in the corpus can be included in EAP courses and instructors could direct students' attention to such specialized vocabularies and ask students to apply them in their academic writings. Materials developers can also specially scheme and design English books to instruct academic political words commonly used in political research papers. EFL and ESL learners can properly integrate such specialized words into their lexicons and expand their receptive and productive word knowledge (Yang, 2015). Moreover, university students can develop their academic vocabulary through corpus-based activities, as these activities have proven to be effective in learning different language features (Oveidi, et al, 2022). Finally, the present research, like almost all studies, had some limitations and further research can shed more light on the issue. This study worked on political sciences journal articles and did not include texts of other types and lengths such as textbooks and reports, further studies can identify high frequency words in such corpora. Moreover, the study concentrated on written texts and did not include spoken data, so researchers can work spoken texts in political sciences such as speeches.

References

Bowen, J. D., Madsen, H., and Hilferty, A. (1985). *TESOL techniques and procedures.* Newbury House.

Browne, C., Culligan, B., & Phillips, J. (2013). New academic word list. Retrieved from http://www.newgeneralservicelist.org/nawl-new-academic-word-list.

Chen, Q., & Ge, G. (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.

Cobb, T., & Horst, M. (2004). Is there room for an AWL in French? In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 15-38). John Benjamin

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

Coxhead, A., Demecheleer, M., & McLaughlin, E. (2016). The technical vocabulary of Carpentry: Loads, lists and bearings. *TESOLANZ Journal*, *24*, 38-71.

Coxhead, A., & Hirsch, D. (2007). A pilot science-specific word list. *Revue Française de Linguistique Appliquée*, *12*(2), 65-78.

Dang, T. N. Y. (2020). Corpus-based word lists in second language vocabulary research, learning and teaching. In S. Webb (ed.), *The Routledge handbook of vocabulary stu*dies (pp. 288-304). Routledge.

Engels, L., K. (1968). The fallacy of word counts. *International Review of Applied Linguistics, 6*(3), 213-231.

Esfandiari, R., & Moein, G. (2015). A corpus-driven food science and technology academic word list. *Issues in Language Teaching*, *4*(2), 157-131.

Farrell, P. (1990). *Vocabulary in ESP: A lexical analysis of the English for electronics and a study of semi-technical vocabulary*. CLCS Occasional Paper No. 25. Trinity College, Dublin. Centre for Language and Communication Studies.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305-327.

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

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range computer program. http://www.vuw.ac.nz/lals/staff/Paul_Nation.

Heidari, F., Jalilifar, A., & Salimi, A. (2020). Developing a corpus-based word list in pharmacy research articles: A focus on academic culture. *International Journal of Society, Culture, & Language, 8*(1), 1-16.

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.

Hsu, W. (2014). Measuring the vocabulary load of engineering textbooks. *English for Specific Purposes, 33*, 54-65.

Hsu, W. (2018). The most frequent BNC/COCA mid- and low-frequency word families in English- medium traditional Chinese medicine (TCM) textbooks. *English for Specific Purposes, 51*, 98-110.

Hutchinson, T., & Waters, A. (1987). *English for specific purposes: A learner-centered approach.* Cambridge University Press.

Hwang, K. (1989). *Reading newspapers for the improvement of vocabulary and reading skills*. Unpublished MA thesis, Victoria University of Wellington, New Zealand.

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

Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC Journal*, *44*(2), 209-232.

Jamalzadeh, M. & Chalak, A. (2019). A corpus-based study of academic vocabulary in physiotherapy research articles. *Language Teaching Research Quarterly*, *9*, 69-82.

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, L., & Pemberton, R. (1994). An investigation of students' knowledge of academic and subtechnical vocabulary. In L. Flowerdew & A. K. k. Tongs (Eds.) *Entering texts* (pp. 183-196). The Hong Kong University of Science and Technology.

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

Lu, Y., & Coxhead, A. (2020). A corpus-based study of vocabulary in the new concept English text book series. *RELC Journal, 53*(3), 597-611.

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

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

Mohammad Hosseinpur, R., Bagheri Nevisi, R., & Afzali Shahri, M. (2018). An Investigation of the reasons for committing plagiarism in the Iranian academic community: Editor-in-chiefs, professors, and students' perspective. *Journal of Foreign Language Research*, *7*(2), 555-588.

Moini, R., & Islamizadeh, Z. (2016). Do we need discipline-specific word lists? Linguistic academic word list (LAWL). *Journal of Teaching Language Skills, 35*(3), 65-90.

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

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

Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. John Benjamins Publishing Company.

Oveidi, M., Sepehri, M., & Shafiee, S. (2022). Differential Effects of Input/Output Tasks on Learning English Collocations by Iranian EFL Learners through the Corpus-Based Instruction. *Journal of Modern Research in English Language Studies*, *9*(4), 143-168.

Safari, M. (2004). *Frequency of words and their technicality in EST*. Unpublished MA thesis. IUST University, Iran.

Safari, M. (2018). Do university students need to master the GSL and AWL words: A psychology word list. *Journal of Modern Research in English Language Studies*, *5*(2), 101-122.

Safari, M. (2019). English vocabulary for equine veterans: How different from GSL and AWL words. *Iranian Journal of English for Academic Purposes*, *8*(2), 51-65.

Sutarsyah, C. (1993). *The vocabulary of economics and academic English*. Unpublished MA thesis. Victoria University of Wellington, New Zealand.

Tongpoon-Patanasorn, A. (2018). Developing a frequent technical words list for finance: A hybrid approach. *English for Specific Purposes, 51*, 45-54.

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.

Vongpumivitich, V., Huang, J., & Chung, Y. (2008). Frequency analysis of the words in the Academic Word List (AWL) and non-AWL content words in applied linguistics papers. *English for Specific Purposes, 28*(1), 33-41.

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*. Longman.

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

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