Vocabulary Lists for EAP and Conversation Students

Mahmood Safari, Assistant Professor at Hazrat-e Masoumeh University, Qom, Iran

Mahmood.safari@gmail.com

Abstract

Despite the abundance of research investigating general and academic vocabularies and developing dozens of word lists, few studies have compared academic vocabulary with general service word lists such as conversation vocabulary. Many EAP researchers assume that university students need to know all the words in West's (1953) General Service List (GSL) as a prerequisite to academic words (e.g., Coxhead's, 2000) and teachers at language institutes recommend conversation students to learn words in Coxhead's Academic Word List (AWL) as a follow-up to the GSL. The present study compared the academic and conversation vocabularies by exploring frequency and coverage of words in academic and conversation corpora. The GSL and AWL words were investigated in a conversation corpus and an academic corpus, each containing around 12 million running words. The analysis revealed that 1200 GSL word families were highly frequent in both corpora and 645 GSL word families were highly frequent in the conversation corpus but of low frequency in the academic texts. Also, a new academic word list of 700 word families was developed, which proved to be much more rigorous than Coxhead's AWL. Further analysis indicated that the abovementioned 645 GSL words had a very low coverage of academic texts (0.7%), while they covered 4.05% of the conversation corpus. The new academic word list covered only 1.6% of the conversation corpus, whereas it had a high coverage of the academic texts (9.1%), much higher than that of the AWL (7.5%). The analysis of some other academic corpora revealed identical results.

Keywords: Academic vocabulary, AWL, Conversation vocabulary, Corpus analysis, GSL

Introduction

Vocabulary is the most important component of language learning and language use. McCarthy (1990) suggested that "it is the experience of most teachers that the single biggest component of any language course is vocabulary" (p. viii). The basic task of every language learner is to learn a great number of words in the target language. Vocabulary knowledge can also enhance other skills such as reading and academic success (Coxhead & Boutorwick, 2018; Uchihara & Hardada, 2018). According to Webb and Paribakht (2015) "Comprehension is likely to rise as the number of known words in a text increases" (p. 34). However, the number of words in every language, especially English, is very huge and beyond anybody's learning capacity. An educated English native speaker knows seventeen to twenty thousand word families, one third of English words (Goulden, Nation & Read, 1990; D'Anna, Zechmeister & Hall, 1991). Richards (2001) believes that "this is a much larger number of words that can be taught in a language course" (p. 5). Therefore, the words to be taught to second language learners must be selected prudently.

The interest in vocabulary research and instruction in early 20th century led to vocabulary control movement, which attempted to "use systematic criteria to select the most useful words for language learning" (Schmitt, 2000, p. 16). The most important product of vocabulary control movement was West's (1953) General Service List (GSL) of words, which contained 2000 word

families. The list has been the major source of vocabulary instruction and research since its development. And with the advent of English for Academic Purposes (EAP) in 1970s, there were further vocabulary research to figure out the crucial English words for academic purposes (Campion & Elley, 1971; Ghadessy, 1979; Lynn, 1973; Parninskas, 1972; Xue & Nation, 1984). These studies culminated in Coxheads's (2000) Academic Word List (AWL), a list of 570 academic word families worked out by exploring a 3.5-million-word corpus of academic texts.

Since the creation of the two lists, there have been some studies to evaluate them and figure out their coverage in other general and academic corpora (Chen & Ge, 2007; Cobb & Horst, 2004; Coxhead, 2000; Engels, 1968; Hirish & Nation, 1992; Hyland & Tse, 2007; Khani & Tazik, 2013; Moini & Islamizadeh, 2016; Nation & Hwang, 1995). However, few studies have delved into the two lists to find the GSL and AWL words which are most frequent in general and academic English. Most studies have evaluated the total coverage of the lists in their own corpora but they have not examined the frequency of every GSL and AWL word to find the truly general service and academic words. The present study attempted to investigate the frequency of GSL and AWL word families in some academic and conversation corpora in order to find the words which are highly frequent in academic and conversation English.

Moreover, many EAP practitioners and researchers assume that university students should know GSL words as a prerequisite to academic words (e.g., Coxhead, 2000). These researchers have developed their academic vocabulary on top of West's (1953) GSL. And conversation students at language institutes and English majors are proposed to study AWL words as a follow-up to the GSL words. However, many GSL words seem to be of very low frequency in most academic texts (e.g., *beak*, *chimney*, *deer*, *quarrel*, and *whistle*) and they are not worth to be invested on in EAP courses. And many academic words seem to be of low frequency in conversation English. The present study attempted to compare academic and conversation vocabularies in order to investigate whether university students need to know all GSL words and if conversation students are required to learn AWL words.

The researcher tried to find the GSL words which are highly frequent in and common to both academic and conversation texts, the GSL words which are highly frequent in conversation texts but of low frequency in academic texts and finally the coverage of these word lists in academic and conversation texts. To that end, the following research questions were put forth:

RQ1: Which GSL words are highly frequent in both academic and conversation texts?

RQ2: Which GSL words are exclusively highly frequent in conversation texts but less frequent in academic texts?

RQ3: What is the coverage of exclusively highly frequent conversation words in academic and conversation corpora?

RQ4: What is the coverage of academic words in conversation and academic texts?

Literature Review

West's (1953) General Service List (GSL) has been used since its creation by language teachers, materials developers and researchers as a list of basic words for general purposes such as language courses. However, the list has been criticized for its size (Engels, 1968), its age (Richards, 1974) and for not fully taking the concept of coverage into account (Nelson, 2000). Therefore, some scholars have suggested that the list needs to be revised (Hwang, 1989) and some researchers, working on larger and modern corpora, have developed new word lists for general purposes (Brezina & Gablasova, 2013; Browne, 2104).

Since 1970s, there have been a second group of studies attempting to develop a vocabulary list for academic purposes. Campion and Elley (1971) and Parninskas (1972)

developed their lists of academic words by analyzing corpora including texts from a range of disciplines. Lynn (1973) and Ghadessy (1979) developed their lists of academic vocabulary by gathering the words which university students learning English had written annotations above in their academic texts. Xue and Nations (1984) combined and edited these four word lists to create their University Word List (UWL), which consisted of 840 word families and covered around 8.5% of academic texts. The search for academic vocabulary climaxed with Coxhead's Academic Word List (AWL), which was developed by exploring a 3.5-million-word corpus of academic texts. It consisted of 570 word families and covered around 10% of the corpus. It has widely been used by materials developers, language teachers, researchers and learners (Coxhead, 2016).

However, recently there have been some criticisms against a monolithic academic vocabulary list and a trend to search for vocabulary lists for more specific academic disciplines (Chung, 2009; Esfandiari & Moein, 2015; Lei & Liu, 2016; Martinez, Beck & Panza, 2009; Moini & Islamizadeh, 2016; Mudraya, 2006; Munzo, 2015; Ward, 2009; Wang, Liang & Ge, 2008). Moreover, more recently there have been some studies trying to work out the most frequent technical words for some academic disciplines (Coxhead & Demecheleer, 2018; Hsu, 2018: Tongpoon-Patanasorn, 2018). Even, there have been much more specific studies. Abdollahpour and Gholami (2018) explored frequency, functions and structures of lexical bundles in medical abstracts.

Nevertheless, the GSL and AWL are still considered as the major general and academic vocabulary lists and many textbooks, graded readers and other instructional materials are developed based on them. Some studies have tried to evaluate the GSL (Engles, 1968; Hirsh, 1992; Hwang, 1989; Nation, 2004; Sutarsyah, 1993) and the AWL (Cobb & Horst, 2004; Li & Qian, 2010; Martinez et al, 2009; Vongpumivitch, Huang & Chang, 2008) by exploring them in their corpora. Nation and Hwang (1995) investigated the coverage of the GSL in the Lancaster-Oslo-Bergen (LOB) Corpus and indicated that the list covered 82.3% of the corpus. Coxhead (2000) explored the GSL words in her academic corpus and came up with a coverage of 76% for the list. The GSL covered 72.48% of a corpus of linguistic research articles in a study by Moini and Islamizadeh (2016). Hirish and Nation (1992) indicated that the GSL list covered 90-92% of a corpus of fiction texts. Also, some studies have investigated the coverage of the AWL in new corpora. Chen and Ge (2007) explored AWL in a medical corpus of articles and indicated that the list covered 10.07% of their 190-thousand-word corpus. The AWL covered 10.06% of a 3.3 million-word corpus in a study by Hyland and Tse (2007). Other studies have come up with identical coverage for the AWL in their corpora (Cobb & Horst, 2004; Khani & Tazik, 2013; Li & Qian, 2010; Vongpumivitch, et al, 2008).

Despite all these studies, which tried to develop or evaluate general service and academic vocabulary lists, there is a paucity of research attempting to work out general English words which are common in academic texts and compare academic and general English vocabularies.

Method

Employed Corpora

The corpora which were explored in the present research included a corpus of conversation English, a general academic corpus and several more specific academic corpora. Table 1 displays the information about the corpora and their size. The corpora were of sufficient size and included texts of various relevant genres and fields to let most frequents words appear in the corpora.

Table 1. *The Size of the Corpora Employed in the Study*

Corpus	Size
Conversation corpus	11,711,325
General academic corpus	12,202,487
Social sciences	3,736,103
Basic sciences	3,723,537
Medicine	3,639,318
Technology	3,723,063

The conversation corpus consisted of over 11.7 million running words and was composed of movie scripts downloaded from movie script websites on the internet. The websites included Daily Script (www.dailyscript.com), Simply Scripts (www.simplyscripts.com) and Screenplays for You (https://sfy.ru). The websites provide movie scripts in different genres produced by various movie directors. The conversation corpus was composed of 518 movie scripts, mainly produced since 1980s to the recent time and were of different genres ranging from action to romance. Every movie script consisted of 200,000 to 250,000 running words. The downloaded movie scripts were in HTML, PDF, TEXT, and WORD formats. However, the scripts in PDF and WORD formats were converted into TEXT format, as the employed software worked only with TEXT and HTML formats.

The general academic corpus was composed of academic texts across a range of academic disciplines, amounting to over 12 million running words. The academic corpus included the following sub-corpora: law, economy, education, philosophy, politics, psychology, and sociology (Social Sciences); math, physics, and chemistry (Basic Sciences); computer, electronics, mechanics, chemical engineering, and metallurgy (Technology); biology and medicine (Medical Sciences); and finally, agriculture, and geography (Earth Sciences). The sub-corpora were of almost the same size, each containing around 650, 000 tokens. The corpora included journal articles published in scholarly journals available on the internet. First, the journals for each academic sub-discipline were identified and then journal articles were downloaded from the sights. The articles in other formats were converted into TXT in order to be processed by the employed software programs (i.e., TexstStat and TextAnalys). The references, appendices and information about the authors in the articles were deleted from the articles to include only article main scripts.

The specific academic corpora included social sciences, basic sciences, medicine, and technology corpora. These corpora were intended to further check the coverage of the AWL words, the newly developed academic words list (called Academic Vocabulary) and the GSL words which were of low frequency in academic texts. The size of these corpora was over 3.5 million running words. The sub-corpora included journal articles in the related fields which were downloaded from scholarly journals available on the internet. The journal articles in other formats were converted to TEXT format in order to be analysable by the employed software. And the references, appendices and information about the authors were deleted.

Text Analysis Software

To analyze the corpora some text analysis programs were employed. The first software program was TextStat 1.5, which analyzes a corpus of any size and lists the words in the corpus alongside information regarding their frequency and ratio. The words are listed in the first column and the frequency and ratio of each word are presented in the second and third columns.

The analysis output is available in MS Word and the researcher can save the file for further study. The other employed text analysis software was TextAnalys. The software adds up the frequencies of the member words of a word family to calculate the aggregate frequency of a word family. Also, it adds up the frequencies of the word families in a word list to calculate the total frequency of a word list in a corpus. Moreover, the software lists all the words outside a specified list according to their frequency and the researcher can identify high frequency words which are absent in the list. It lists the word families in the order of their aggregate frequency (i.e., the sum of the frequencies of the word members) and the more frequent and less frequent word families are easily distinguished. Table 2 displays an example output file of the TextAnalys software. The software was useful in finding GSL and AWL word families which were highly frequent or of less frequency in the corpora. The conversation and academic corpora were analyzed through the software and the frequency of the GSL and AWL words were worked out in the two corpora. Then the frequencies of the members of each word family were added up to figure out high frequency and low frequency GSL and AWL words.

Table 2. *An Example Output File of the TextAnalys Software*

the:156838 --> 156838
this:12453 , these:8426 --> 20879
study:3738 , studies:4241 , studying:339 , studied:429 --> 8747
do:2855 , does:1328 , did:924 , doing, 341, done:358 --> 5806
.
.
beak:0 , beaks:0 --> 0
cottage:0 , cottages:0 --> 0
Total : 192270
---
cognitive:4239
american:3124

Selection Criteria: Frequency, Range and Coverage

The present study employed some specific criteria for selecting words to be included in the intended vocabulary lists, which included frequency, range and coverage. The first criterion was frequency. Richards and Schmidt (2010) define frequency as "the number of occurrences of a linguistic item in a text or corpus". Range refers to "a measure of the distribution of linguistic items throughout a sample, which are generally expressed as a measure of the number of texts or samples in which a linguistic item occurs" (Richards & Schmidt, 2010, p. 479). The coverage of a word list in a corpus is the percentage of the tokens of the corpus which is accounted by the word list. It is calculated by dividing the aggregate frequency of the words in a list by the total number of the tokens in a corpus.

The frequency criterion was set on 100 times of occurrence. The word families which occurred 100 times or more in the employed corpora were selected to be included in the intended lists. Regarding the range of GSL words, the words were supposed to occur in both conversation and academic corpora with a frequency of 100 times or more to be included in the list of GSL words common to conversation and EAP English. The coverage of the word lists over the conversation and academic corpora was calculated by dividing the total frequency of the word

members of the word lists by the total number of tokens in the corpora. The coverage of the lists is presented as percentages of the running words covered by a word list.

Results

Results of Corpus Analysis

To compare academic and conversation vocabularies and find high frequency words in conversational English and university texts, the researcher investigated the GSL and AWL word families in the academic and conversation corpora.

First, to find the GSL words which are highly frequent in and common to conversation and academic English, the researcher worked out the frequency of the GSL word families in the two corpora. The frequencies of the members of each word family were added up to reveal the aggregate frequency of each word family. The word families which had a total frequency of 100 or above in each corpus were identified and considered as the common core vocabulary for conversation and academic English. There were 1196 GSL word families which met the criterion and occurred 100 or more times in each corpus. Six GSL word families had a frequency of 99 in the academic corpus and a frequency of over 100 times in the conversation corpus, but in order to have a round number of words common to conversation and academic English, these word families were included in the list. Therefore, 1200 GSL word families were considered as common core vocabulary for the two areas of language uses, that is, conversation and academic English. The shared vocabulary included structure words such as *to, of, with, however* and general English words commonly employed in most language use situations such as *accept, damage, large, prevent*. Thus, the answer to the first research question (Which GSL words are highly frequent in both conversation and academic texts?) is 1200 GSL word families are highly frequent in and common to academic and conversation English. The base words of these word families are presented in Appendix A.

In order to find the answer to the second research question and work out the words which are commonly used in conversational English but are of less frequency in academic English, the GSL words were investigated in the two corpora. The GSL word families whose total frequency was below 100 times in the general academic corpus were identified and recorded. Eight hundred and six GSL word families occurred less than 100 times in the academic corpus. Then the frequency of these word families was figured out in the conversation corpus. The majority of these words had a frequency of 100 or above in the conversation corpus. In fact, 645 GSL word families occurred 100 times or more in the conversation corpus but less than 100 times in the general academic corpus. These word families were considered as exclusively conversation vocabulary. The base words of these word families are presented in Appendix B. They mainly include words which are used in specific language use areas such as romance and household. Table 3 displays some examples of the GSL words which were highly frequent in the conversation corpus but of low frequency in the academic corpus. As it is evident, most of them are not expected to occur frequently in academic texts. The words *kitchen, bottle, pocket, shirt* and *knife* are commonly used in conversation but less commonly used in academic texts. To sum up, 1200 GSL word families were highly frequent in and common to conversation and academic English, 645 GSL word families occurred frequently in conversation English but were of less frequency in academic texts and 155 GSL word families were of less frequency in both academic and conversational English. Therefore, the answer to the second research question (Which GSL words are highly frequent in conversation texts but less frequent in academic texts?) was the 645 GSL word families which occurred 100 times or more in the conversation corpus but was of less frequency in the academic corpus. They were considered as exclusively conversation vocabulary.

Table 3. *Frequency of some GSL Words in Conversation and Academic Corpora*

Word	Conversation corpus	Academic corpus
Kiss	3712	32
Kitchen	3370	32
Desk	3623	27
Shoulder	2759	82
Tear	2278	47
Pocket	1779	82
Bottle	1667	50
Lady	1663	52
Shirt	1567	3
Knife	1363	53

Subsequently, to compute the coverage of exclusively conversation vocabulary in academic and conversation English and answer the third research question (What is the coverage of exclusively highly frequent conversation words in academic and conversation corpora?), the researcher investigated the words in the two corpora. The exclusively conversation vocabulary covered 4.05% of the running words in the conversation corpus, while it covered only 0.7% of the academic corpus. That is, out of every 100 words in the conversation corpus around 4 four words were from the list of exclusively conversation vocabulary and in the academic corpus out of every 100 words, less than one word was from the list of exclusively conversation vocabulary. Then the coverage of the word list was worked out in some more specific academic corpora. Table 4 displays the coverage of the list in the additional academic corpora. As the table shows, the exclusively conversation vocabulary covered .44% to 1.95% of the more specific corpora. On the average, it covered less than one percent (i.e., .98%) of the more specific corpora.

Table 4. *Coverage of the Exclusively Conversation Vocabulary in Academic Corpora*

Word List	Basic Sciences	Social Sciences	Biology	Medicine	Technology
Exclusively Conversation Vocabulary	0.44%	1.95%	1.12%	0.87%	0.54%

Finally, to answer the fourth research question (What is the coverage of academic words in conversation and academic texts?), the researcher investigated the academic words in the conversation and academic corpora. First, as Coxhead's AWL has been criticized for being biased against some disciplines such as biology and medicine and favoring certain academic fields such as economy and law, a new academic vocabulary list (to be called Academic Vocabulary) was developed. The AWL word families were investigated in the academic corpus to find the AWL words which were less frequently used in academic texts and were erroneously placed in the AWL list. Research has revealed that AWL is not a truly general academic word list and some AWL words are included in the list due to the large size of sub-corpora belonging to specific fields. Moreover, the academic corpus was analyzed to find the words which were highly

frequent in the academic texts but were absent in the AWL list due to the absence or small size of the sub-corpora belonging to specific academic disciplines such as biology. The investigation of the AWL words in the academic corpus revealed that 70 AWL word families were less frequent in the corpus (i.e., they occurred less than 100 times in the general academic corpus) and they were excluded from the new academic word list.

Furthermore, the analysis of the academic corpus revealed 292 general academic words which were highly frequent in the corpus and occurred more than 100 times in the general academic corpus but were absent in Coxhead's (2000) list. The second criterion to select the word families to be added to the new academic word list was range, i.e., the occurrence of the words with a frequency of ten times or more in at least 15 out of the 19 sub-corpora in the general academic corpora. There were 198 word families which occurred more than 100 times in the corpus and in at least 15 sub-corpora. Two words occurred in less than 15 sub-corpora but were added to the list in order to have a round number of words in the list. In addition, the words were checked in technical dictionaries to verify that they were not technical words of any specific academic discipline. Therefore, the new academic vocabulary list included 700 academic words. Table 5 displays some example words from the excluded 70 AWL words and some instances of the added 200 academic words. As the list indicates, the excluded AWL words are mainly technical words of law and economy and they were included in the AWL due to the large size of law and economy sub-corpora in Coxhead's corpus. The list of added academic words reveals that some academic words are absent in the AWL due to underrepresentation of some academic disciplines such as technology and medicine. The base words of the newly developed Academic Vocabulary including 700 academic word families (i.e., the 500 high frequency AWL word families and 200 newly identified academic word families) are presented in Appendix C.

Table 5. *Example Words from the Excluded AWL Words and Added Academic Words*

Excluded AWL words	Added academic words
amendment	absorb
clause	accomplish
currency	blend
deduce	deposit
estate	fluid
integrity	infect
levy	launch
ordination	pollute
subordination	resist
subsidy	verify

Then the coverage of the newly developed academic vocabulary list (i.e., Academic Vocabulary) and Coxherad's AWL in the conversation and academic corpora was calculated. Table 6 displays the coverage of the two word lists over the conversation and academic corpora. As the table indicates, the Academic Vocabulary covered only 1.6% of the running words in the conversation corpus but 9.1% of the academic corpus. Coxhead's AWL, too, had a love coverage of 1.2% in the conversation corpus but high coverage of 7.5% in the academic corpus.

Table 6. *Coverage of the Academic Word Lists in Conversation and Academic Corpora*

Word Lists	Conversation Corpus	Academic Corpus
Academic Vocabulary	1.6%	9.1%
Coxhead's AWL	1.2%	7.5%

Then the coverage of the two lists were checked in the more specific academic corpora, whose information is presented in Table 7. On the average, the Academic Vocabulary covered 12.76% of the more specific academic corpora and the AWL covered 9.84% of the corpora. The Academic Vocabulary had a higher coverage than the AWL on the average and in each academic discipline and this suggests that it is a better choice to be employed in the development of academic materials and to be presented to university students.

Table 7. *Coverage of the Academic Word Lists in the more Specific Academic Corpora*

Word Lists	Basic Sciences	Social Sciences	Biology	Medicine	Technology
Academic Vocabulary	11.7%	10%	14.2%	13.8%	14.0%
Coxhead's AWL	8.3%	8.9%	10.4%	9.4%	11.9%

Discussion

The analysis of the word lists and corpora indicated that there are around 1200 GSL word families which are commonly used in both academic and conversation English. In fact, the results indicated that the two areas of language use share 60% of their most frequent words (i.e., 1200 words out the most frequent 2000 words). Moreover, the study prepared a list of GSL words which occur frequently in academic texts and EAP students commonly encounter them in their educational texts. University students will need to learn these words as a prerequisite to academic words as they need to know their meanings to be able to read their academic texts. This finding is in line with studies which revealed that around half of GSL words are highly frequent in non-fiction and scientific texts. Nation and Hwang (1995) compared the GSL with the list of words extracted from two more modern corpora, namely the Brown Corpus and the LOB Corpus, and revealed that the two lists shared 1331 word families. In a study by Engels (1968), the second 1000 GSL words covered only 4.7% of non-fiction texts.

The second finding of the present study was that there were around 800 GSL word families which are of low frequency in academic texts and university students are not expected to encounter them frequently in their academic texts. In fact, it was shown that not all GSL words are commonly used in academic texts and EAP students do not need to learn all GSL words. Also, the results revealed that there are 645 GSL word families which are highly frequent in conversation English but of low frequency in academic texts. Therefore, according to the findings, 1845 GSL words are highly frequent in conversation texts, while only 1200 GSL words are commonly used in academic texts. The above-mentioned list of 645 GSL words covered 4.05% of the tokens in the conversation corpus but only 1.6% of the academic corpus. The list had low coverage in more specific academic corpora too. The findings are in line with the studies

which revealed that the second half of the GSL word families do not occur frequently in scientific and non-fiction texts (Engels, 1968; Moini & Islamizadeh, 2016; Nation, 2004; Nation & Hwang, 1995). In Moini & Islamizadeh's study, 1342 GSL word families did not have the required frequency to be included in their list of academic words for linguistics. Nation (2004) indicated that the GSL covered 89.6% of the running words in a spoken corpus, while it covered only 75.5% of an academic corpus.

The focus of EAP courses and materials must be general words which are truly common in academic texts. Many researchers have developed their academic vocabulary on top of the GSL or other lists of common general English words (Coxhead, 2000) but the results of the present study revealed that many general English words are not frequently used in academic texts. There also have been some researchers who created their academic word lists without taking a general service list for granted (Mudraya, 2006; Martinez et al, 2009; Valipoori & Nassaji, 2013; Ward, 2009; Yang, 2014). Ward (1999) criticized starting the development of academic word lists with a base general service list and created his engineering word list without such general English vocabulary. The present study revealed that not all GSL words are highly frequent in academic texts and it suggests that only highly frequent general English words must be presented to EAP students as they do not have sufficient time to learn so many general English words which might not have a beneficial return for their great endeavour.

The analysis of the academic word lists in conversation and academic corpora revealed that academic words have a high coverage in academic texts but a low coverage in conversation texts. This indicates that there are some academic words which are frequently used in academic texts but less frequently in non-academic texts. The finding suggests that only EAP students and not students of conversation courses are required to master academic words as these words are not expected to be frequently encountered in everyday conversation. This finding confirms previous studies which indicated that academic words are frequently used in academic texts but are rarely used in other text types. Xue and Nation's (1984) University Word List covered 8.5% of academic texts, while it covered only 1.7% of fiction texts. Coxhead's Academic Word List covered 9% to 12% of academic texts (Chen & Ge, 2007; Coxhead's, 2000; Hyland & Tse, 2007; Khani & Tazik, 2013; Li & Qian, 2010; Valipour & Nassaji, 2013; Wang et al, 2008), but it covered only 1.7% of a corpus of fiction texts (Coxhead, 2000), and 3.9% of tokens in newspaper texts (Coxhead & Nation, 2001). Gardner and Davies' (2013) Academic Vocabulary List (AVL) covered over 13.7% of academic texts in British National Corpus (BNC) and Corpus of Contemporary American English (COCA), while it covered only 3.4% of fiction texts in both corpora. The study by Paribakht and Webb (2016) indicated that AWL coverage in the passages of 12 versions of an English proficiency test used for admission purposes at Canadian universities was consistently present and substantial. Academic words have even been shown to be less frequently used in spoken academic texts. The AWL covered only 2.4% of the running words in a corpus of spoken academic English, consisting of presentations on technical topics given by a group of Swedish students (Hinks, 2003). Thompson (2006) found that 340 AWL word families occurred less than once in every two university lectures at undergraduate and post graduate levels. These studies indicate that academic vocabulary is specific to academic texts and must be focused only in EAP courses and instructional materials.

Conclusions

The results suggest that courses and instructional materials aiming at teaching English for academic purposes must not invest on many GSL words as they are rarely used in academic texts. Instead, they can focus on academic words which can more effectively benefit EAP students.

Moreover, most EAP courses are short and cannot present too many words and university students do not have sufficient time to learn a huge number of words. Unlike what some vocabulary researchers assume, university students do not need to acquire all GSL words. Students must focus their attention on general words that occur most frequently in academic texts. The present study worked out the 1200 GSL words which occur commonly in academic disciples. Together with the 700 academic word families, it makes a list of 1900 word families which occur most frequently in academic texts. The list is much shorter than the combination of GSL plus AWL (2570 word families) but has a higher coverage of academic texts. These words must be explicitly taught to EAP students as incidental vocabulary learning is slow and unpredictable (Paribakht & Wesche, 1997).

The findings of the present study can be beneficial for language teachers, materials developers and learners. EAP teachers and materials developers can take advantage of the provided list and present them in their EAP textbooks and materials. Students can review the list to learn the general and academic English words that they do not know but are frequently used in academic texts. On the other hand, the results revealed that most academic words occur less frequently in conversation texts. The academic word list covered only 1.8% of the running words in the conversation texts. This suggests that conversation courses must not focus on academic words as they are not commonly used in everyday conversation. Focusing on more everyday conversation words will benefit students in conversation courses more. The results of the present study revealed that the majority of GSL words were highly frequent in conversation texts and all GSL words are suggested to be presented at conversation courses. Even a newer conversation vocabulary can be worked out by analysing conversation texts like movie scripts, recorded everyday conversations and available spoken corpora.

Like all research studies the present research had some limitations, which interested researchers can remove through further research. First, the present study was limited to movie scripts, further research can investigate other conversation texts such as transcribed spoken dialogues. Moreover, researchers can explore and identify most frequent multiword units in academic and conversation English. Interested researchers can also identify most frequent grammatical structures in academic and conversation English. Finally other linguistic features, such as discourse markers, can be investigated in academic and conversation corpora.

References

Abdollahpour, Z. & Gholami, J. (2018). Building blocks of medical abstracts:Frequency, functions and structures of lexical bundles. *Asian ESP Journal, 14*(1).

Brezina, V. & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics, 36*(1), 1-22.

Browne, Ch. (2014). A new General Service List: The better mousetrap we've been looking for? *Vocabulary Learning and Instruction, 3*(2), 1-10.

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

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*, 502-514.

Chung, T. (2009). The newspaper word list: A specialized vocabulary for reading newspapers. *JALT Journal, 31*(2), 159-182.

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). Amsterdam, the Netherlands: John Benjamins.

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

Coxhead, A. (2016). Reflection on Coxhead (2000), "A new academic word list".*TESOL Quarterly, 50*(1), 181-185.

Coxhead, A. & Boutorwick (2018). Longitudinal vocabulary development in an EMI International school context: Learners and texts in EAL, Maths, and Science. *TESOL Quarterly, 52*(3).

Coxhead, A. & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes, 51*, 84-97.

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

D'Anna, C.A., Zechmeister E.B. & Hall. J.W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior, 23*, 109-122.

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

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

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics, 35*, 305–327.

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

Goulden, R., Nation, P. & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics, 11*, 341-363.

Hincks, R. (2003). Pronouncing the Academic Word List: Features of L2 student oral presentations. *Proceedings of the 15th International Congress of Phonetics Sciences*. Universitat Auto`noma de Barcelona, Barcelona. Retrieved from http://www. speech.kth. se/ctt/publications/papers03/ icphs03_1545.pdf.

Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*, 689-696.

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.

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), 195-214.

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*, 402-411.

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*, 183-198.

McCarthy, M. (1990). *Vocabulary*. Oxford University Press.

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.

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

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. (2004). A study of the most frequent word families in the British National Corpus. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 3-13). Amsterdam: John Benjamins.

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

Nelson, M. (2000). *A corpus-based study of business English and business English teaching materials*. Unpublished PhD Thesis. University of Manchester, Manchester. Retrieved May, 24, 2003, from http://www.kielikanava.com/thesis.html

Paribakht, T. S. & Webb, S. (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *English for Academic Purposes, 21*, 121-132.

Paribakht, T.S. & Wesche, M. (1997). Vocabulary enhancement activities and reading for meaning in second language vocabulary development. In J. Coady and T. Huckin (eds.)*Second language vocabulary acquisition: A rationale for pedagogy* (pp. 174-200). Cambridge: Cambridge University Press.

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

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

Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.

Richards, J.C. & Schmidt, R. (2010). *Longman dictionary of language teaching and applied linguistics*. London: Longman.

Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge UniversityPress.

Seal, B.D. (1991). Vocabulary learning and teaching. In M. Celce-Murcia (ED.), *Teaching English as a second and foreign language* (pp. 296-311). Boston, Massachusetts: Heinle & Heinle Publishers.

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

Thompson, P. (2006). A corpus perspective on the lexis of lectures, with a focus on economics lectures. In K. Hyland, & M. Bondi (Eds.), *Academic discourse across disciplines* (pp. 253-270). New York: Peter Lang.

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

Uchihara, T. & Hardada, T. (2018). Roles of vocabulary knowledge for success in English-medium instruction: Self-perceptions and academic outcomes of Japanese undergraduates. *TESOL Quarterly, 52*(3), 564-587.

Valipoori, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *English for Academic Purposes, 12*, 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 wordlist. *English for Specific Purposes, 27*(4), 442–458.

Ward, J. (1999). How large a vocabulary do EAP engineering students need? *Reading in a Foreign Language, 12*(2), 309-324.

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

Webb, S., & Paribakht, T.S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes*, *38*, 34-43.

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

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

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

Appendices

Appendix A: Base words of the 1200 word families common to conversation and academic English

bad
balance
ball
band
bank
bar
bargain
base
basin
bay
be
beam
bean
bear
because
become
bed
before
begin
behave
behind
being
believe
belong
below
bend
beside
best
better
between
beyond
big
bind
birth
bit
bite
black
blade
blind
block
blood
blue
amount
an
ancient
and
angle
animal
another
answer
anxiety
any
apart
appear
apply
appoint
approve
April
arch
argue
arise
arm
army
around
arrange
arrive
arrow
art
article
artificial
as
ask
association
at
attack
attempt
attend
attract
august
average
avoid
away
axe
back
a
able
about
above
absence
accept
accident
accord
account
across
act
actual
add
address
admit
adopt
advance
advantage
advertise
advice
affair
afford
after
again
against
age
agent
ago
agree
agriculture
aim
air
all
allow
almost
alone
along
already
also
although
always
among

content
continue
control
convenience
cook
copper
copy
corner
correct
cost
council
count
country
course
court
cover
crack
crime
critic
crop
cross
cultivate
cure
current
curve
custom
cut
damage
danger
dark
date
day
dead
deal
decay
December
decide
declare
decrease
deep
defeat
defend
degree
delay
deliver
demand
department
chalk
chance
change
character
charge
check
chief
child
choose
church
circle
city
class
clay
clear
cliff
clock
close
cloud
coal
coarse
coast
coffee
cold
collect
college
colony
combine
come
command
commerce
committee
common
company
compare
compete
complete
complicated
compose
concern
condition
confidence
confuse
connect
conscious
consider
contain
board
body
bone
book
border
born
both
bottom
bound
boundary
box
boy
brain
branch
break
bridge
bring
broad
brown
build
bundle
burn
bus
business
but
buy
by
calculate
call
camera
camp
can
canal
cape
capital
car
care
carry
case
cause
cave
cent
center
century
certain
chain
chair

fail
fair
faith
fall
false
familiar
family
fan
far
farm
fashion
fast
fat
father
fault
fear
February
feed
feel
fellow
female
few
field
figure
fill
film
find
fine
finish
fire
firm
first
fish
fit
five
fix
flame
flat
flood
floor
flow
fly
fold
follow
food
foot
for
eight
either
elastic
elect
electricity
else
employ
empty
encourage
end
engine
English
enjoy
enough
enter
entire
equal
escape
especial
essence
even
event
ever
every
exact
examination
example
excellent
except
excess
excite
exercise
exist
expect
expense
experience
experiment
explain
explode
explore
express
extend
extra
extreme
eye
face
fact
depend
describe
desire
destroy
detail
determine
develop
die
difference
difficult
dip
direct
discipline
discover
discuss
disease
distance
distinguish
district
disturb
divide
do
dollar
door
dot
double
doubt
down
draw
drive
drop
dry
due
during
each
early
earn
earth
ease
east
easy
eat
edge
educate
effect
effort
egg

in
inch
include
increase
indeed
independent
industry
influence
inform
insect
inside
instead
instrument
insure
intend
interest
interfere
international
into
introduce
invent
iron
island
it
its
January
jaw
join
joint
judge
July
June
just
keep
key
kill
kind
king
know
lack
lake
land
language
large
last
late
latter
hall
hand
handle
happen
hard
harm
harvest
have
he
head
health
hear
heart
heat
heavy
height
help
here
high
hill
hire
his
history
hold
hole
home
hope
horizon
hospital
host
hot
hour
house
how
human
hundred
hunt
I
ice
idea
ideal
if
imagine
immediate
important
impossible
improve
force
foreign
forest
form
formal
former
forward
four
frame
free
frequency
frequent
fresh
friend
from
front
fruit
full
further
future
gain
game
gap
gas
gate
gather
gay
general
get
give
glass
go
god
gold
good
govern
gradual
grain
grass
great
green
grind
ground
group
grow
guide
half

motion
motor
mountain
mouse
mouth
move
much
mud
multiply
must
my
name
narrow
nation
native
nature
near
necessary
need
neglect
neither
net
never
new
next
night
nine
nineteen
no
noise
none
nor
north
not
note
nothing
notice
November
now
number
obey
object
observe
occasion
ocean
October
of
map
march
mark
market
mass
master
match
material
matter
may
mean
measure
mechanic
medicine
meet
melt
member
memory
mention
mere
message
metal
middle
might
mild
mile
mill
mind
mine
minister
minute
miss
mistake
mix
model
moderate
modern
moment
money
month
moon
moral
more
moreover
morning
most
mother
law
lay
lead
leaf
learn
least
leave
left
length
less
lesson
let
letter
level
library
lie
life
light
like
likely
limit
line
liquid
list
literature
little
live
load
local
lock
log
long
look
loose
lose
loss
lot
love
low
lung
machine
main
make
male
man
manage
many

probable
problem
produce
profession
profit
progress
promise
pronounce
proof
proper
property
propose
protect
prove
provide
public
pull
pump
punish
pure
purpose
push
put
qualify
quality
quantity
question
quick
race
radio
rain
raise
rank
rapid
rare
rat
rate
rather
ray
reach
read
ready
real
reason
receive
recent
recognize
pay
people
per
perfect
perform
perhaps
permanent
permit
person
photograph
pick
picture
piece
pin
pipe
place
plain
plan
plant
plate
play
please
point
political
pool
poor
popular
population
position
possess
possible
post
poverty
power
practical
practice
prefer
prepare
present
preserve
president
press
prevent
price
print
prison
private
off
offer
office
often
oil
old
on
once
one
only
onto
open
operation
opinion
opportunity
opposite
or
order
ordinary
organ
organization
organize
origin
other
otherwise
our
out
over
overcome
owe
own
pack
page
pain
pair
paper
parent
park
part
particular
party
pass
passage
past
path
patient
pattern

she
sheet
shell
ship
shock
shoot
shore
short
should
show
shower
side
sign
signal
silver
simple
since
single
sit
situation
six
size
skill
skin
slide
slight
slip
slope
slow
small
smooth
so
society
soft
soil
solid
solve
some
son
soon
sort
sound
south
space
speak
special
speed
round
row
rule
run
safe
sale
salt
same
sample
sand
satisfy
save
say
scale
scarce
scatter
scene
school
science
screen
sea
search
season
second
secret
see
seed
seem
self
sell
send
sense
sentence
separate
September
serious
serve
set
settle
seven
several
severe
shall
shallow
shape
share
sharp
recommend
record
red
reduce
refer
reflect
refuse
regard
regular
relation
relieve
religion
remain
remark
remember
repair
repeat
replace
report
represent
reproduce
republic
request
reserve
resist
respect
responsible
rest
result
return
review
reward
rice
rich
right
ring
rise
risk
river
road
rock
rod
roll
roof
room
root
rough

those
though
thought
thousand
threaten
three
through
thus
tide
tie
tight
time
tin
tip
title
to
tobacco
today
together
too
tool
tooth
top
total
touch
toward
town
track
trade
train
translate
trap
travel
treat
tree
trial
trouble
true
trust
try
tube
turn
twenty
two
type
under
understand
sudden
suffer
sugar
suggest
suit
summer
sun
supply
support
suppose
sure
surface
surprise
surround
suspect
sweet
swim
system
table
tail
take
talk
taste
tax
teach
tell
temperature
ten
tend
term
test
than
thank
that
the
their
then
there
therefore
these
they
thick 2
thin
thing
think
this
thorough
spend
spin
spirit
split
spot
spread
spring
square
staff
stage
stain
stand
standard
star
start
state
station
stay
steady
steam
steel
stem
step
stiff
still
stock
stone
stop
store
storm
story
straight
stream
street
strength
stretch
strict
strike
string
strong
struggle
study
subject
substance
succeed
success
such

wind
window
winter
wire
wise
wish
with
within
without
woman
wood
word
work
world
worth
would
wrap
write
wrong
year
yellow
yes
yet
yield
you
young
zero
waste
water
wave
way
we
weak
wealth
weapon
weather
week
weigh
well
west
what
wheat
when
where
whether
which
while
white
who
whole
whose
why
wide
wild
will
win
union
universe
university
unless
until 2
up
upon
upper
upward
use
usual
valley
value
various
very
vessel
view
village
violent
visit
vote
voyage
wage
wait
walk
wall
want
war
wash

Appendix B: Base words of 645 word families exclusively highly frequent in conversation English

autumn
avenue
awake
awkward
baby
bag
bake
barber
bare
bargain
barrel
basket
bath
amongst
amuse
anger
angry
annoy
apologize
applaud
apple
arrest
ash
aside
astonish
audience
absolutely
accuse
admire
adventure
afraid
afternoon
ahead
airplane
alike
alive
aloud
altogether
ambition

crowd
crown
cruel
crush
cry
cup
curious
curl
curtain
cushion
damp
dance
dare
daughter
deaf
dear
debt
deceive
deed
deer
delicate
delight
descend
desert
deserve
desk
despair
devil
diamond
dictionary
dig
dinner
dirt
disappoint
dish
dismiss
ditch
dive
doctor
dog
donkey
dozen
drag
dream
dress
drink
drown

carriage
cart
castle
cat
catch
cattle
caution
ceremony
cheap
cheat
cheese
chest
chicken
chimney
Christmas
clean
clerk
clever
climb
cloth
club
coat
coin
collar
color
comb
comfort
companion
complain
confess
conquer
conscience
conversation
cool
cork
corn
cottage
cotton
cough
courage
cousin
cow
coward
crash
cream
creature
creep

battle
beast
beat
beauty
beg
bell
belt
beneath
berry
bicycle
bill
bird
bitter
blame
bless
blow
boast
boat
boil
bold
borrow
bottle
bow
bowl
brass
brave
bread
breath
bribe
brick
bright
brother
brush
bucket
bunch
burst
bury
bush
busy
butter
button
cage
cake
calm
cap
captain
card

heap
heaven
hesitate
hide
hit
holiday
hollow
holy
honest
honor
hook
horse
hotel
humble
hunger
hurry
hurt
husband
hut
idle
ill
imitate
immense
ink
inquire
instant
insult
interrupt
invite
inward
joke
journey
joy
juice
jump
kick
kiss
kitchen
knee
knife
knock
knot
ladder
lady
lamp
laugh
lazy
forbid
forget
fork
fortune
forty
fourteen
freeze
frequent
Friday
fright
fry
fun
funeral
fur
furnish
gallon
garage
garden
generous
gentle
girl
glory
goat
grace
grand
grateful
grave
grease
greed
grey
guard
guess
guest
guilty
gun
habit
hair
hammer
hang
happy
harbor
hardly
haste
hat
hate
hay
heal
drum
duck
dull
dust
duty
eager
ear
earnest
eighteen
eighty
elephant
eleven
empire
enemy
entertain
entrance
envelope
envy
evening
evil
excuse
extraordinary
factory
fade
faint
famous
fancy
fate
favor
feast
feather
fence
fever
fierce
fifteen
fifty
fight
finger
flag
flash
flavor
flesh
float
flour
flower
fond
fool

pretend
pretty
pride
priest
prize
procession
program
prompt
proud
pupil
purple
puzzle
quarrel
quart
quarter
queen
quiet
rabbit
rail
rake
raw
refresh
regret
rejoice
remind
rent
reply
reputation
rescue
resign
restaurant
retire
revenge
ribbon
rid
ride
ripe
rival
roar
roast
rob
rope
rot
royal
rub
rubber
rude
noble
nonsense
noon
nose
nowhere
nurse
nut
oar
offend
omit
orange
ought
ounce
pad
paint
pale
pan
parcel
paste
pause
peace
pearl
peculiar
pen
pencil
pet
pig
pigeon
pile
pinch
pink
plaster
plenty
pocket
poet
poison
polish
polite
postpone
pot
pound
pour
powder
praise
pray
preach
precious
lean
leather
leg
lend
liberty
lid
lift
limb
lip
listen
loaf
loan
lonely
lord
loud
loyal
luck
lump
lunch
mad
mail
manners
marry
mat
meal
meanwhile
meat
merchant
mercy
milk
miserable
modest
Monday
monkey
murder
music
mystery
nail
neat
neck
needle
neighbor
nephew
nest
nice
niece
ninety

sympathy
tailor
tall
tap
tea
tear
telegraph
telephone
temper
temple
tempt
tender
tent
terrible
thief
thirst
thirteen
thirty
thorn
thread
throat
throw
thumb
Thursday
ticket
tidy
tire
toe
tomorrow
ton
tongue
tough
tour
tower
toy
tray
treasure
tribe
trick
trip
trunk
Tuesday
tune
twelve
twist
ugly
umbrella
slave
sleep
smell
smile
smoke
snake
snow
soap
soldier
sore
sorrow
soul
soup
sour
spade
spare
spell
spill
spit
spite
splendid
spoil
spoon
sport
stair
stamp
steal
steep
steer
stick
stir
stomach
stove
strange
straw
strip
stripe
stuff
stupid
suck
Sunday
swallow
sweat
sweep
swell
swing
sword
ruin
rush
rust
sacred
sacrifice
sad
saddle
sail
sake
salary
Saturday
sauce
saucer
scent
scrape
scratch
screw
seat
secretary
seize
seventeen
seventy
shade
shadow
shake
shame
sheep
shelf
shelter
shield
shine
shoe
shop
shoulder
shout
shut
sick
sight
silence
silk
sincere
sing
sir
sister
sixteen
sixty
sky

witness
wonder
wool
worm
worry
worse
worship
worst
wound
wreck
wrist
yard
yesterday
youth
wax
wear
weave
Wednesday
weed
welcome
wet
wheel
whip
widow
wife
wine
wing
wipe
upright
upset
urge
vain
veil
verse
victory
voice
voyage
wake
wander
warm
warn
watch

Appendix C: Base words of the 700 word families highly frequent in Academic English

abandon
absorb
abstract
academy
accelerate
access
accommodate
accompany
accomplish
accumulate
accurate
achieve
acid
acknowledge
acquire
adapt
adequate
adhesive
adjacent
adjust
administrate
affect
agenda
aggregate
aid
align
allocate
alloy
alter
alternative
assist
assume
assure
attach
attain
attitude
attribute
author
authority
automate
available
award
aware
barrier
benefit
bias
blend
bomb
bond
brief
budget
bulk
bureau
cable
capable
capacity
capture
career
cast
category
coincide
collaborate
collapse
collide
column
combust
comment
commerce
commission
commit
commodity
communicate
community
compatible
compensate
compile
complement
complex
component
composite
compound
comprehensive
comprise
compute
concentrate
concept
conclude
concrete
conduct
confer

analogy
analyze
annual
anticipate
apparatus
apparent
append
appendix
appreciate
approach
appropriate
approximate
arbitrary
architecture
area
array
aspect
assemble
assess
assign
contract
contrast
contribute
convene
convention
convert
convey
cooperate
coordinate
core
corporate
correlate
correspond
corrosion
couple
create
credit
criteria
crucial
crystal
cube
culture
curriculum
cycle
data
decade
decline
cell
challenge
chamber
channel
chapter
chart
chemical
chip
choice
circumstance
cite
civil
clarify
classic
click
client
climate
cluster
code
coherent
distort
distribute
diverse
document
domain
domestic
dominate
draft
drain
drama
drill
drug
duration
dynamic
economy
edit
element
elevate
eliminate
emerge
emit
emphasis
empirical
enable
encounter
energy
enforce
confine
confirm
conflict
conform
congress
consent
consequent
conserve
considerable
consist
consolidate
constant
constitute
constrain
construct
consult
consume
contact
contaminate
context
expand
expert
explicit
exploit
export
expose
external
extract
fabric
facilitate
factor
faculty
feature
federal
fertile
fiber
file
filter
final
finance
finite
flexible
flight
fluctuate
fluid
focus
format

defect
define
definite
demonstrate
denote
dense
deposit
derive
design
despite
detect
deviate
device
devote
dilute
dimension
diminish
discrete
disperse
displace
display
dispose
dissertation
dissolve
distinct
guideline
handbook
hardware
harmony
hazard
hence
hierarchy
highlight
homogeneity
hybrid
hypothesis
identical
identify
ignorant
illustrate
image
impact
implement
implicate
implicit
imply
import
enhance
enormous
ensure
entity
environment
equate
equilibrium
equip
equivalent
erode
error
establish
estimate
evaluate
eventual
evident
evolve
exceed
exchange
exclude
execute
exert
exhaust
exhibit
exit
interact
intermediate
internal
interpret
interval
intrinsic
invert
invest
investigate
involve
isolate
issue
item
job
journal
justify
keyword
label
labor
laboratory
laser
launch
formula
found
foundation
fraction
fracture
framework
frontier
fuel
function
fund
fundamental
furnace
furthermore
generate
generation
globe
glossary
goal
goods
grade
graduate
grant
gravitate
grid
guarantee
mesh
meter
method
migrate
military
minimal
minimum
minor
mirror
mission
mitigate
mobile
mode
modify
moist
molecule
monitor
motive
mutual
navigate
navy
negate

impose
impress
incentive
incidence
incline
income
incorporate
index
indicate
individual
induce
infer
infrastructure
inherent
inhibit
initial
initiate
injure
innovate
input
insert
insight
inspect
instance
institute
instruct
integral
integrate
intelligent
intense
perceive
percent
period
perspective
phase
phenomenon
philosophy
physical
planet
plastic
plot
plus
polar
policy
pollute
port
portion
layer
layout
lecture
legal
legislate
liable
license
link
locate
logic
longitude
loop
magnetic
magnitude
maintain
major
manipulate
manual
margin
marine
matrix
mature
maximize
mechanism
media
medical
medium
membrane
mental
menu
radical
random
range
ratio
rational
react
recall
recover
refine
regime
region
register
regress
regulate
reinforce
reject
relax
network
neural
neutral
nevertheless
normal
notion
novel
nuclear
objective
obtain
obvious
occupy
occur
offset
online
onset
optic
optimal
option
orient
outcome
output
overall
overlap
panel
parallel
parameter
participate
partner
passive
schedule
scheme
scholar
scope
score
section
sector
secure
sediment
seek
segment
select
sequence
series
session
shear
shift

pose
positive
potential
precede
precise
predict
preliminary
previous
primary
prime
principal
principle
prior
priority
probe
proceed
process
professional
profile
project
promote
propagate
proportion
prospect
protocol
proxy
psychology
publication
publish
pulse
purchase
pursue
qualitative
quote
radiate
sufficient
suit
sum
summary
superior
supervise
supplement
suppress
survey
survive
suspend
sustain
release
relevant
rely
remote
remove
render
require
research
reservoir
reside
residual
resolve
resonate
resource
respond
restore
restrain
restrict
retain
reveal
revenue
reverse
revise
revolution
rigid
robust
role
rotate
route
routine
rural
satellite
saturate
scan
scenario
ultimate
undergo
underlie
undertake
uniform
unique
update
urban
utilize
valid
vapor
vary
significant
similar
simulate
simultaneous
singular
site
sketch
so-called
software
solar
sole
solvent
somewhat
source
species
specific
specify
spectrum
sphere
stable
statistic
status
stimulate
strain
strategy
stratum
stress
structure
style
submit
subsequent
subset
substantial
substitute
successor

switch
symbol
symmetry
symposium
synthesize
target
task
team
technical
technique
technology
temporary
tense
terminate
text
theme
theory
thereby
thermal
thesis
threshold
thrust
topic
trace
tradition
traffic
transact
transfer
transform
transient
transit
transmit
transport
trend
truck
tunnel
turbine
turbulent
twice
typical
vehicle
verify
version
versus
vertical
via
vibrate
violate
virtual
visible
vision
visual
void
volume
web
whereas
x-ray
zone