

English Vocabulary for Equine Veterans: How Different from GSL and AWL Words

¹Mahmood Safari*

IJEAP-1903-1364

Abstract

ESP students are usually suggested to master general and academic word lists such as Wests' (1953) General Service List (GSL) and Coxhead's (2000) Academic Word List (AWL) to be able to read their academic texts. However, it seems that university students may not need to learn all the words in the two lists as some words in the lists are of less frequency in academic texts. Moreover, there are some nontechnical words which occur frequently in academic disciplines but are absent in the two lists. The present study attempted to identify words which frequently occur in equine veterinary academic texts. To that end, a corpus of over 3.6 million running words, containing equine veterinary journal articles, was developed and analyzed using some text analysis software (TextStat and TextAnalys). As Coxhead the frequency of 100 was set as the criterion for word selection. The results revealed that 1091 GSL word families and 116 AWL word families were of less frequency (i.e., occurred less than 100 times) in the EVC (Equine Veterinary Corpus) and there were 214 nontechnical word families which occurred frequently in the EVC but were absent in the GSL and AWL. The high frequency GSL and AWL words alongside the newly identified words constituted the Equine Veterinary Word List (EVWL), which covered the EVC 2.5% more than the combination of GSL and AWL although it consisted of 993 fewer words. The findings can benefit equine veterans, veterinary students, EAP teachers, materials developers and researchers.

Keywords: Academic Word List, General Service List, Equine Veterinary, Frequency, Coverage

1. Introduction

Vocabulary is one of the most important linguistic features, or even the most important one, in language learning and language use. "It would be impossible to learn a language without vocabulary-without words" (Rivers, 1968, p. 462). Learning and teaching vocabulary is one of the most important areas to which applied linguists have paid much attention since early times in applied linguistics. Second language learning encompasses learning a large number of words by EFL learners. Allen (1983) states that "Experienced teachers of English as a Second language know very well how important vocabulary is. They know students must learn thousands of words that speakers and writers of English use" (p. 1).

EFL learners need to acquire a sufficient number of words in order to be able to act properly in the second language. The number of words in every language, especially English, is very large and beyond the capacity of any language learner, even native speakers. Language learners cannot and do not need to master all the words in a language. Even native speakers of English know around one third of the words in their language. According to research (Goulden, Nation, & Read, 1990; D'Anna, Zechmeister, & Hall, 1991) educated native speakers of English know only 17000 to 20000 word families, which is one third of the 54000 word families in English. However, even this number of words seems to be too large to be presented to second language learners (Richards, 2001).

Therefore, language teaching practitioners need to select words which must be taught to language learners. According to Nation (2005), "The first decision to make when teaching a word is to decide whether the word is worth spending time on or not" (p. 49). The idea of word selection for presenting to language learners was recognized by scholars since long time ago. Vocabulary control movement in early

¹Assistant Professor (Corresponding Author), m.safari@hmu.ac.ir; Department of English Language Translation, Hazrat-e Masoumeh University, Qom, Iran

20th century assumed that there must be a limited number of words to be taught to language learners and the words must be chosen cautiously. Many researchers tried to find out the most important words to be presented to second language learners. West's (1953) word selection study was the most important research in mid-20th century, which developed a list of 2000 most important word families for general language use purposes. The word list was named the General Service List of words (GSL) and it has been widely used by language teachers, materials developers and researchers since then. However, the list has been criticized by many researchers for its age (Engels, 1968), size (Richards, 1974) and coverage (Nelson, 2000). The corpus which was analyzed by the researcher and his team included very old texts and the resulting word list does not include new words such as TV and computer. The second half of the word list has been shown to have a low coverage in nonfiction texts.

Some decades later in around 1970s, language teaching practitioners recognized that specific groups of language learners require different types of language and academic register has its own specific language. There were many studies to work out the most important words for academic purposes (Campion & Elley, 1971; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972; Xue & Nation, 1984). The most important research which developed a list of general academic words was Coxhead's (2000) study. Coxhead analyzed a corpus of 3.5 million running words, containing texts from different academic disciplines, and developed her Academic Word List (AWL). The list was placed on top of the GSL. That is, the researcher worked out the non-GSL words which occurred frequently in academic texts. The GSL covered around 76% of Coxhead's academic corpus and the AWL covered around 10% of the corpus. The researcher herself and some other researchers have proposed that university students should master the words in the GSL and AWL in order to be able to read their academic texts.

More recently, some studies have tried to develop vocabulary lists for specific academic disciplines (Chung, 2009; Lei & Liu, 2016; Martinez, Beck & Panza, 2009; Moini & Islamizadeh, 2016; Mudraya, 2006; Munoz, 2015; Ward, 2009; Wang, Liang, & Ge, 2008). They believed that a single general academic word list would not be a good source for students of different academic disciplines. The studies revealed that every academic discipline has its own specific word list, which includes fewer words than the combination of GSL and AWL word families.

However, despite the plethora of research developing word lists for different academic disciplines such as Medicine (Chen & Ge, 2007; Lei & Liu, 2016; Wang, et al., 2008), Engineering (Mudraya, 2006; Ward, 2009), Applied Linguistics (Khani & Tazik, 2013; Vongpumivitch, Huang, & Chung, 2008), Finance and Economy (Li & Qian, 2010; Sutarsyah, 1993), Agriculture (Martinez, et al., 2009; Munoz, 2015), Chemistry (Valipouri & Nassaji, 2013), Social Sciences (Kwary & Artha, 2017) and Nursing (Yang 2015), there have been almost no studies, to the best knowledge of the author, investigating words frequently used in some academic disciplines such as veterinary field and subfields. Equine veterinary is an important sub-discipline of veterinary field as horses are domestic animals that man has long used for various purposes such as agriculture, entertainment and police work. Equine veterans and veterinary students need to increase their professional knowledge through reading scientific texts such as journal articles presenting most recent discoveries and innovations about horses and their problems. The majority of scientific writings, including veterinary texts, are written in English and most journals publish articles in English. Therefore, equine veterans need to have developed a large equine veterinary English lexicon so that they can read their professional texts easily and effectively. The present study aimed at developing a list of most important words of equine veterinary through a corpus study.

2. Literature Review

The main question regarding vocabulary instruction is "what words should be taught to language learners?" There have been many studies attempting to work out these words and many word lists have been developed. The vocabulary control movement in early 20th century was an attempt to limit the words to be taught in language courses. The movement "culminated in the appearance of Michel West's A

General Service List in 1936” (Seal, 1991, pp. 296-297). The list consists of 2000 word families, which are claimed to be frequently used in written and spoken language. However, the list has been criticized for some drawbacks. Engels (1968) claimed that the second 1000 GSL words are not truly general English words as they covered only around 4.7% of running words in non-fiction texts. He considered them ‘fallacious’ as “they cannot be called general service words” (p. 266). Richards (1974) questioned the age of the GSL as it was developed in early and mid-20th century and it does not include many words which have been created afterward. Words such as *television*, *computer*, *drug*, *battery*, and *internet*, which have been developed or gained currency since the first half of the 20th century are missing in the list. Also the GSL is criticized for not taking the concept of coverage into account (Nelson, 2000). Moreover, the employed corpus seems to have been replete with fiction texts. The list covered over 90% of the running words in a corpus of fiction texts (Hirsh & Nation, 1992) while it covered around 76% of nonfiction texts (Coxhead, 2000).

With the advent of English for Academic Purposes (EAP) in 1970s, there was a plethora of research trying to develop word lists for academic purposes (e.g., Cowan, 1974; Friel, 1978; Inman, 1978; Kirkham, 1978; Wingard, 1981). Some studies worked on academic corpora to develop their academic word lists and some other studies used other approaches to develop their lists. Campion and Elly (1971) and Praninskas (1972) developed their lists of academic vocabulary working on academic corpora containing texts from a range of academic disciplines. Lynn (1973) and Ghadessy (1979) collected words that university students had written annotations above them in their English textbooks. Xue and Nation (1984) combined and edited these four lists to create their University Word List (UWL). The list contained 840 word families and covered around 8.5% of academic texts. The search for academic vocabulary culminated in Coxhead’s (2000) Academic Word List (AWL). Coxhead developed her list examining an academic corpus of 3.5 million running words. The list consisted of 570 word families and was shown to cover around 10% of the academic corpus.

Due to the drawbacks of the GSL, there recently have been some studies to replace it with newer general service lists. Browne (2014) explored a subsection of Cambridge English Corpus (CEC), a corpus of 273 million running words, and created his own word list. The list was named A New General Service List and included 2801 headwords. However, this list, too, was criticized for some drawbacks. A major criticism against the employed corpus was that it left out academic and newspaper texts from the CEC corpus and the existence of the learner English was unwarranted as learner English is most probably different from native speaker English. Also, the resulting list is too large to be presented in EAP courses or learned by EAP students. Brezina and Gablasova (2013) developed their New General Service List comparing four general English corpora of varying sizes (from 1 million to 12 billion running words). The researchers came up with a list of 2122 words common to the four corpora to which they added 372 new words. The researchers compared corpora of varying sizes and ages and one of the corpora (EnTenTen12) was developed by extensive web scrawling without an appropriate genre classification. It was not made sure that different genres have proportionate size in the corpus.

Regarding academic vocabulary, some researchers have criticized a single monolithic academic vocabulary and have favored academic vocabulary lists for each academic discipline, arguing that words have different meanings and functions in different disciplines (e.g., Hyland & Tse, 2007). They assumed that every academic discipline has its own vocabulary list and tried to develop lists for more specific academic disciplines (e.g., Chung, 2009; Lei & Liu, 2016; Martinez et al., 2009; Moini & Islamizadeh, 2016; Mudraya, 2006; Munoz, 2015; Valipouri & Nassaji, 2013; Ward, 2009; Wang et al., 2008). Ward analyzed a corpus of engineering texts and developed his “basic engineering English word list”. Munoz, analyzing a corpus of 700 agriculture semi-popularization articles, developed a highly specific word list, which covered 6% of the running words in the corpus. Valipouri and Nassaji (2013) developed a chemistry academic wordlist of 1400 word families, which included only 327 AWL word families.

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). Coxhead and Demecheleer explored written and spoken corpora of plumbing and developed a plumbing technical word list, which covered 30% of the written corpus and 11% of the spoken corpus. Hsu worked out the most frequent word families in English-medium traditional Chinese medicine textbooks and Tongpoon-Patanasorn developed a list of frequent technical words for finance.

However, despite the plethora of studies creating word lists for some academic fields and sub-fields, no vocabulary lists have been developed for many university disciplines. Veterinary and equine veterinary are the field and sub-field for which no vocabulary list has ever been identified. The students of such disciplines are mainly advised to master the words of the GSL and AWL in order to be able to read their academic texts. The lists, together, consist of 2570 word families and are supposed to cover around 86% of academic texts (Coxhead, 2000). However, this number of words seems to be huge for university students to master. Moreover, many GSL and AWL words seem to be of less frequency in many academic disciplines and there seems to be many words which are highly frequent in academic disciplines but absent in the GSL and AWL. The present study attempted to work out high frequency words in equine veterinary academic texts and develop a word list for equine veterinary sub-discipline. To that end, the GSL and AWL word families were checked in a corpus of 3.6 million running words in order to identify the GSL and AWL words which are of high frequency in equine veterinary texts. Moreover, the corpus was explored to find the words which are of high frequency in equine veterinary texts but are absent in the GSL and AWL lists. The following research questions were put forth in order to develop the EVWL.

Research Question One: Which GSL and AWL word families are highly frequent in equine veterinary texts?

Research Question Two: Which words are highly frequent in equine veterinary texts but absent in the GSL and AWL?

Research Question Three: How do the size and coverage of an equine veterinary word list compare to those of a list containing GSL and AWL words?

3. Method

3.1. Corpus

In order to find the most important words for equine veterinary sub-discipline, the researcher was required to explore a large corpus of equine veterinary texts. Thus, a corpus of over 3.6 million running words was developed. The corpus contained 3,688,758 tokens and was composed of 936 journal articles published in scholarly journals of equine veterinary. The journals were ISI-indexed and Scopus-indexed journals. They included *Equine Veterinary Journal*, *Equine Veterinary Education* and *Journal of Equine Veterinary Science*. The articles which constituted the Equine Veterinary Corpus (EVC) were published since 2010 to 2019 in the above-mentioned journals. The reference sections of the articles were deleted from the texts in order to have a corpus of only journal article passages. The corpus was a representative sample of equine veterinary journal articles as the size of the corpus was large enough (i.e., over 3.6 million running words) and it consisted of articles from various areas and topics within equine veterinary as it included articles published over ten years in the mentioned journals. The corpus was analyzed using some text analysis software.

3.2. Text Analysis Software

The researcher used some text analysis software to analyze the corpus. The first software was TextStat 1.5, which analyzes a corpus of any size and lists the words in the corpus together with information about their frequency and ratio. It lists the words in the first column and the frequency and ratio of each word in the second and third columns. The software provides the analysis output in MS Word and Excel files, which can be saved for further study by the researcher. The second 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 of 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, so the researcher can identify high frequency words which are outside 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 1 displays an example output of the second software.

Table 1: 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

3.3. Data Collection Procedure

Initially the researcher searched for scholarly journals of equine veterinary on the internet and some ISI- and Scopus-indexed journals were identified. The research journals which were identified included *Equine Veterinary Journal*, *Equine Veterinary Education* and *Journal of Equine Veterinary Science*. Then a great number of articles were downloaded from the journals. The reference section of the articles was deleted in order to have clean texts including only the main passages of the articles. The texts were loaded into the first text analysis software (TextStat 1.5), which analyzed the corpus and listed the words in the corpus alongside their frequency. The words were listed according to their frequency size from the most frequent word (i.e., the) to the least frequent ones (i.e., the words that occurred just once in the corpus).

Then the developed list was loaded into the second employed text analysis software (TextAnalys). The word families in the GSL and AWL were placed as the database in the software in order to compute their frequency and coverage. The software added up the frequencies of the member words of a word family in order to calculate the aggregate frequency of each word family. And also it added up the frequencies of all word families in the word lists to compute the total frequency of each word list. The software listed the word families according to their frequency size from the most frequent to the least frequent. Therefore, the word families whose aggregate frequency was more than 100 were easily identified. These GSL and AWL word families were recoded as high frequency general and academic words. Also, the GSL and AWL word families whose aggregate frequency was less than 100 were identified and recorded as less frequency words.

The software also listed the words outside the database GSL and AWL word families based on their frequency. The words which occurred more than 100 times in the corpus but were absent in the GSL and AWL word lists were identified and recorded. Since, the study worked on a single sub-discipline of veterinary, the range criterion was not adopted. Then the researcher reviewed the words from the top of the list and omitted the words which were proper nouns, numbers and junk words. Then the newly identified high frequency words were checked in a technical dictionary of veterinary, *Black's Veterinary*

Dictionary (Boden, 2005), and the technical words of equine veterinary were identified and excluded from the new list. The next step was to put words into word families. The different forms of a word were put in one line to form a word family, that is, the plural, third person singular, past form, participle form and different derivations of every word were put together to form a word family. The researcher added the different forms of every word and developed a list of word families for high frequency words which were absent in the GSL and AWL. The high frequency GSL and AWL word families alongside the newly developed list of word families constituted the Equine Veterinary Word List. Then the total frequencies of the word lists were computed in order to compare the coverage of the GSL plus AWL words to that of the EVWL.

4. Results of Corpus Analysis

The aim of the present research was to identify the most frequent words for equine veterinary as a sub-discipline of veterinary discipline. Hence, the researcher analyzed a corpus of equine veterinary journal articles. The GSL and AWL word families were examined in the corpus in order to find the GSL and AWL words which are highly frequent in equine veterinary. The following word lists were identified by analyzing the corpus:

List 1: GSL and AWL word families which are highly frequent in equine veterinary texts,

List 2: GSL and AWL word families which are of low frequency in equine veterinary texts,

List 3: Highly frequent general and academic word families in equine veterinary texts which are absent in the GSL and AWL.

The analysis of the corpus revealed that 909 GSL word families and 454 AWL word families met the criterion and occurred more than 100 times in the EVC and were considered as high frequency general and academic words in equine veterinary. However, 1091 GSL word families occurred less than 100 times in the corpus and were considered as low frequency general words in equine veterinary sub-discipline. And 116 AWL word families had an aggregate frequency below 100 and were regarded as low frequency words. Therefore, more than half of the GSL words were not frequently used in the equine veterinary texts and over 20% of AWL word families were of less frequency in the equine veterinary texts. Table 2 displays the information about the high and low frequency GSL and AWL word families.

Table 2: High and Low Frequency GSL and AWL Words in the Corpus

Word lists	High frequency	Low frequency
GSL word families	909	1091
AWL word families	454	116

The second step was to work out words which are highly frequent in equine veterinary texts but absent in GSL and AWL lists. TextAnalysis software listed the words outside the GSL and AWL lists according to their frequency and the researcher was able to identify high frequency equine veterinary words which are absent in the GSL and AWL. The non-GSL and non-AWL words which occurred more than 100 times in the corpus were worked out and recorded for analysis. The researcher reviewed these words and excluded proper nouns (i.e., names of people, places, etc.), numbers, and junk words in order to come up with real words. There were 326 such words which occurred more than 100 times in the corpus. Then the researcher checked these words in a technical dictionary of veterinary, *Black's Veterinary Dictionary* (Boden, 2005), to identify technical words of equine veterinary. There were 28 technical words of veterinary which occurred more than 100 times in the equine veterinary corpus (Appendix B). They were excluded from the list of high frequency equine veterinary words absent in the GSL and AWL. Finally, the researcher developed a list of word families by placing the words of the same family together. The list consisted of 214 word families (Appendix A).

The combination of these 214 word families and high frequency GSL and AWL word families created a list of 1577 words, which was much smaller than the combination of GSL and AWL words (2570 word families). To evaluate the coverage of the new list in the equine veterinary corpus and compare it to the coverage of the list of GSL plus AWL words, the researcher divided the total frequency (the sum of the frequencies of all the member words) of the lists by the total number of the tokens in the corpus. Table 3 displays information on the size, total frequency and coverage of the list of high frequency words which were absent in the GSL and AWL and those of the list of GSL and AWL word families which were of low frequency in the equine veterinary corpus. The coverage of the former was much larger than that of the latter. As the table indicates, the 214 word families covered 4.1 % of the corpus while the list of low frequency GSL and AWL words covered only 1.6% of the corpus. This indicates that teaching these 214 word families will benefit veterinary students more than teaching them the 1207 low frequency GSL and AWL words. The former is smaller and covers larger number of running words in equine veterinary texts.

Table 3: Size, Total Frequency and Coverage of the Word Lists

Word Lists	Size	Total Frequency	Coverage
LF GSL-AWL	1207	59,020	1.6%
NHFW	214	151,239	4.1 %

LF GSL-AWL = Low frequency GSL and AWL words; NHFW = New high frequency words list

Over 410 GSL words occurred less than 10 times in the corpus and 65 GSL words did not occur at all. And 10 AWL words occurred less than ten times in the corpus. Table 4 shows some of the GSL and AWL words which were of very low frequency in the equine veterinary corpus. As the table indicates, many GSL words are not expected to occur frequently in veterinary texts and most other academic disciplines. They are general English words that occur rarely in scientific texts (e.g., *chalk*, *chimney*, *sock*, *sacred*). They may even be of less frequency in general English and everyday conversation. Expecting university students to know or learn these words is not logical. It would be much more effective to present students with vocabulary lists which do not contain such words.

Also, some AWL words are not expected to be frequently used in veterinary texts (e.g., *append*, *commodity*, *deduce*, *levy*, *subsidy*) and university students majoring in veterinary do not need to learn them. One would easily recognize that the presence of these words in the AWL was due to the fact that some academic disciplines like economy and law were overrepresented in Coxhead's corpus (i.e., these disciplines contributed more texts to the corpus), while some other disciplines like medicine and biology were underrepresented in the corpus.

Moreover, there were shown to be many high frequency words which university students in general and veterinary students in particular need to know but the words are absent in the GSL and AWL lists. Some of the words which were highly frequent in the equine veterinary corpus but absent in the two lists include *adolescent*, *climate*, *coach*, *deficit*, *disorder*, *fatigue*, *impair*, *physician*, *substantial*, *urban*, *victim*. As one would easily recognize, these words are very common words and are expected to be known by university students for effective reading. The absence of these words in the AWL indicates that the employed corpus was not a truly general academic corpus and was biased against some academic disciplines such as veterinary.

Table 4: Low Frequency GSL and AWL Words in Equine Veterinary

GSL words	AWL words
barber	append
chalk	clause
chimney	commodity
cupboard	commence
ditch	deduce
hut	estate
sacred	fee
sock	levy
whistle	revenue
widow	subsidy

5. Discussion

The analysis of the equine veterinary corpus and the investigation of the GSL and AWL word families in the corpus indicated that a great number of GSL and AWL words are of less frequency in veterinary texts. Thus, the answer to the first research question is: 1363 GSL and AWL word families (909 GSL and 454 AWL word families) are highly frequent in equine veterinary texts and the remaining 1207 GSL and AWL word families are not frequently used in equine veterinary texts. Most of the low frequency GSL words are general English words that mainly occur in everyday conversations and less in academic texts and the less frequent AWL words are mainly specific to some academic disciplines such as economy and law.

The results are in line with many previous research findings. In the present study, over half of the GSL words were shown to be of low frequency in equine veterinary texts. In several previous studies the second 1000 GSL word families were shown to have very low coverage of nonfiction texts (e.g., Engles, 1968; Hwang, 1989; Sutarsyah, 1993). Engels questioned the importance of the second 1000 word families as it covered about 4.7% of the running words in non-fiction texts. He stated “they cannot be called general service words” (p. 266). Also, the second 1000 GSL word families covered only 4.27% of the running words in Moini and Islamizadeh’s (2016) academic corpus and they concluded that “it is not necessary to learn the 2nd 1000 GSL words before AWL words” (p. 79). Around 90 GSL words never occurred in Moini and Islamizadeh’s corpus and 1342 GSL words did not have the required frequency to be included in their Linguistics Word List. Nation & Hwang (1995) revealed that over 450 GSL words were not present in the lists of high frequency words derived from the analysis of the Brown Corpus and Lancaster-Oslo-Bergen Corpus. Mudraya (2006) exploring a corpus of engineering texts developed a list of 1200 words, which did not include many GSL word families.

Research has also shown that many AWL words are of low frequency in some academic disciplines. Over 270 AWL words were less frequent in Chen and Ge’s (2007) corpus of medical texts and they concluded that AWL “is far from complete in representing the academic words frequently used in medical RAs [research articles]” (p. 502). More than 170 AWL words were not among Yang’s (2015) nursing academic word list, which included 676 word families. In Moini and Islamizadeh’s (2016) study, 189 AWL word families occurred less frequently in their linguistics corpus and were excluded from their list. Valipouri and Nassaji (2013) developed a chemistry academic wordlist of 1400 word families, which included only 327 AWL word families. Over 25% of the AWL words did not overlap with academic words found in Khani and Tazik’s (2013) corpus of applied linguistics.

Furthermore, the present research indicated that there are a fairly large number of general and academic words which are commonly used in equine veterinary texts but are absent in the GSL and AWL. The study revealed 214 word families which were highly frequent in the equine veterinary corpus but not

included in the two lists. These new words are general and academic words which are expected to occur in academic texts but due to the inadequacy of the corpora used by West (1953) and Coxhead (2000), they were not included in the lists. The 214 high frequency words covered 4.1% of the equine veterinary texts, while the 1207 low frequency GSL and AWL word families covered only 1.6% of the corpus.

Thus the answer to the second research question is: 214 word families are highly frequent in equine veterinary texts but absent in the GSL and AWL. The finding is in line with most previous study results. Vongpumivitch et al. (2008) indicated that 95 AWL word families were of low frequency in applied linguistics texts and there were 128 highly frequent words in their corpus but absent in the AWL. In Moini and Islamizadeh's (2016) study, 224 words were highly frequent in their academic corpus but absent in the AWL. In Yang's (2015) nursing academic word list, there were 278 word families which were absent in the AWL. There have been some claims that the AWL is biased for and against some academic disciplines and it is not a truly general academic word list. The existence of such technical words as *amendment*, *levy*, *estate*, *subsidy* and the absence of such general academic words as *adolescent*, *coach*, *deficit*, *fatigue*, *urban* corroborates the fact that Coxhead's academic corpus and her AWL were biased for some academic disciplines like economy and law and against some university fields such as medical sciences. The AWL covered 11% of the social sciences texts and only around 6% of the agriculture, biology and medicine corpora (Cobb & Horst, 2004; Hyland & Tse, 2007; Munoz, 2015).

Regarding the third research question, the results of the present study revealed that the size of the equine veterinary word list (1577 word families) is much smaller than that of the combination of the GSL and AWL words (2570 word families) and the coverage of the former is larger than that of the latter. The coverage of the equine veterinary word list was 2.5% larger than that of the list of GSL and AWL words, although the former included 993 fewer words. In Moini and Islamizadeh's (2016) study, the 224 non-AWL/non GSL high frequency word families had a higher coverage of their academic corpus than the 1531 low frequency GSL and AWL words. Valipouri and Nassaji (2013) identified 390 word families which frequently occurred in their chemistry corpus and covered 7% of the running words.

6. Conclusion and Pedagogical Implications

The present study aimed at identifying general and academic words which are highly frequent in equine veterinary texts by exploring a 3.6-million-word corpus of equine veterinary journal articles. The researcher investigated the GSL and AWL word families in the EVC and identified the ones which are most frequently used in equine veterinary texts. Around half of the GSL word families were highly frequent in the EVC and 454 AWL word families occurred with high frequency in the corpus. That is, half of the GSL words and over 20% of the AWL words were of low frequency in equine veterinary journal articles. Moreover, there were 214 word families which were highly frequent in the EVC but absent in the GSL and AWL. Therefore, as it was expected there are many GSL and AWL words that equine veterans and veterinary students do not need to master and there are many non-GSL and non-AWL words that should be learned by veterans. It was shown that the combination of GSL and AWL word families is not the best source for veterans and veterinary students' vocabulary learning. The high frequency GSL and AWL word families and the newly identified 214 high frequency word families constituted the Equine Veterinary Word List (EVWL), which included 1577 word families. The list contained 993 word families fewer than the combination of the GSL and AWL. Moreover, further analysis of the word lists indicated that the EVWL covered the equine veterinary corpus 2.5% more than the combination of GSL and AWL. So, EVWL contained fewer words and enjoyed higher coverage of equine veterinary texts than the combination of GSL and AWL words. The findings suggest that the newly developed list (i.e., EVWL) is a much better source for veterinary students and veterans and will benefit them more.

The findings of the present research can be used by equine veterans, veterinary students, EAP teachers, materials developers and researchers. According to Nation (2016), word lists can be used for

many instructional purposes, ranging from material development for language teaching to designing graded reading books, analysis of text vocabulary load, and development of language vocabulary tests. Materials developers can use the list in developing EAP materials. Most EAP materials are developed intuitively without considering the frequency of words in academic texts. If materials developers have access to most frequent linguistic features, such as vocabulary, and grammar, their materials will be more effective and beneficial for EAP students. Materials for veterinary students can be based on a list of most important words in the discipline and present the students with the words that they will meet most often in their academic texts. Moreover, materials developers can develop supplementary vocabulary materials to increase students' knowledge of common words used in equine veterinary texts and improve their reading ability. As most EAP courses are short and cannot cover all important academic words in a discipline, students need to have access to supplementary materials in order to improve their academic reading and writing.

EAP teachers can use the list for developing their own materials if there are not appropriate materials available on the market. They can also develop classroom activities and tasks to improve their students' competence in common words of equine veterinary. Teachers can divide the equine veterinary word list into several sub-lists and develop handouts to instruct the words or ask their students to study the words in each sub-list for a classroom quiz.

The list can also be beneficial for equine veterans and veterinary students. They can design a plan to study and master the words in the list during a specific time period. They can look up the words in a dictionary and write down the words, their meaning and example sentences for them on flash cards or vocabulary notebooks. Also, they can check the words in veterinary corpora through concordance programs in order to see the words in context.

The study can also benefit EAP researchers. They can use the methodology, data analysis and the findings of the study in their own research. Interested researchers can explore other related topics and areas. Researchers can develop word lists for other academic disciplines such as law, sociology, physics etc. Most general and academic word lists include most frequent single words, ignoring the fact that many words in the texts are parts of some multiword units and students need to master the meanings and functions of these units. Interested researchers can focus on most important multiword units such as verb phrases and idioms and identify most important multiword units in veterinary or other academic disciplines. Moreover, interested researchers can work on other linguistic features like grammar structures. They can find out the most frequent grammar structures in veterinary or other academic disciplines. Most university students have problem with grammar structures and the list of most important grammar structures will help EAP learners enormously in increasing their linguistic competence and comprehension of their academic texts. Finally, the present researcher worked on written texts, further research can investigate most important words in spoken texts, like university lectures, in academic disciplines.

Acknowledgement: TextAnalys software was developed in the project on the development of English materials for the university students of non-English majors, under Dr. Hossein Farhady's supervision, sponsored by SAMT.

References

- Allen, V. F. (1983). *Techniques in teaching vocabulary*. Oxford: Oxford University Press.
- Boden, A. (2005). *Black's veterinary dictionary*. London: A & C Black Publishers Limited.
- Brezina, V. & Gablasova, D. (2013). Is there a core general vocabulary? Introducing the New General Service List. *Applied Linguistics*, 36(1), 1-22.
- Browne, C. (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.
- Cowan, J. R. (1974). Lexical and syntactic research for the design of EFL reading materials. *TESOL Quarterly*, 8(4), 389-399.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Coxhead, A. & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, 51, 84-97.
- D'Anna, C. A., Zechmeister, E. B. & Hall, J.W. (1991). Toward a meaningful definition of vocabulary size. *Journal of Reading Behavior*, 23(1), 109-122.
- Engels, L., K. (1968). The fallacy of word counts. *International Review of Applied Linguistics*, 6(3), 213-231.
- Friel, M. (1978). A verb frequency count in legal English. *ESPMENA Bulletin*, 10, 18-21.
- Ghadessy, P. (1979). Frequency counts, word lists, and materials preparation: A new approach. *English Teaching Forum*, 17(1), 24-27.
- Goulden, R., Nation, P., & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11(4), 341-363.
- Hirsh, D., & Nation, P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language*, 8(2), 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.

- Inman, M. (1978). Lexical analysis of scientific and technical prose. In M.T., Trimble, L., Trimble, & K., Drobnic (eds.), *English for science and technology: A discourse approach* (pp. 242-249). Oregon: ELI, Oregon State University.
- Khani, R., & Tazik, K. (2013). Towards the development of an academic word list for applied linguistics research articles. *RELC Journal*, 44(2), 195-214.
- Kirkham, J. (1978). How common are 'common core' words? *ESPMENA Bulletin*, 10, 15-18.
- Kwary, D. A. & Artha, A.F. (2017). The Academic Article Word List for Social Sciences. *MEXTESOL Journal*, 41(4) 1-11.
- Lei, L. & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes*, 22, 42-53.
- Li, Y., Qian, D.D. (2010). Profiling the academic word list (AWL) in a financial corpus. *System*, 38(3), 402-411.
- Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*, 4(1), 25-32.
- 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.
- 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, I. S. P. (2005). Teaching vocabulary. *Asian EFL Journal*, 7(3), 47-54.
- Nation, I. S. P. (2016). *Making and using word lists for language learning and testing*. Amsterdam: John Benjamins Publishing Company.
- Nation, I. S. 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>
- 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.
- Rivers, M. W. (1968). *Teaching foreign language skills*. Chicago: The University of Chicago Press.
- 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.
- 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. *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 wordlist. *English for Specific Purposes, 27*(4), 442-458.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*(3), 170-182.
- West, M. (1953). *A general service list of English words*. London: Longman.
- Wingard, P. (1981). Some verb forms and functions in six medical texts. In L., Selinker, E., Tarone, & V. Hanzeli (eds.), *English for academic and technical purposes: Studies in honor of Louis Trimble*. (pp. 53-64). Rowley, Mass: Newbury House.
- 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*(1), 27-38.

Appendix A: High frequency non-technical words in EVC which are absent in GSL and AWL

Abuse, abused, abuses, abusing, acculturation, acculturate, acculturates, acculturated, acute, acuter, acutest, addict, addicted, addicts, addiction, addictions, adherence, adherent, adherences, adhere, adheres, adhered, administered, administer, administers, administered, administering, adolescence, adolescent, adolescents, adverse, adversity, affirmative, aggression, aggressive, aggressions, alcohol, alcoholism, allergy, allergic, allergies, amnesia, amnesiac, antecedent, antecedents, appraise, appraises, appraised, appraising, appraisingly, appraisal, appraisals, archive, archives arouse, arouses, aroused, arousing, arousal athlete, athletes, athletic attribute, attributes, attributed, attributing, attribution, attributions auditory autonomy, autonomous aviation axis, axes background, backgrounds barrier, barriers bereaved, bereavement biofeedback biology, biological bore, boring, bored, bores, boredom breast, breasts bulletin, bulletins bully, bullies, bullying, bullied burden, burdens burnout, burnouts cancer, cancers, cancerous career, careers caregiver, caregivers, caregiving cell, cells chronic client, clients climate, climatic, climates clinic, clinical, clinician, clinicians, clinics coach, coaches, coaching, coached cognition, cognitions, cognitive collaborate, collaborates, collaborated, collaborating, collaboration, collaborative competence, competencies, competency, competent comply, complies, complied, complying, compliance comprehend, comprehends, comprehended, comprehending, comprehension, comprehensions concrete conscientiousness, conscientious continuum, continuums cope, coped, copes, coping cord, cords coronary, coronaries correlate, correlated, correlates, correlation, correlations counsel, counseling, counselling, counselor, counselors crisis cue, cues, cued, cuing curriculum, curriculums, curricular deficit, deficits demographic, demography dense, denser, densest, density, densities deprive, deprives, deprived, deprivation, deprivations diabetes, diabetic diagnose, diagnosed, diagnoses, diagnosis, diagnostic diet, diets discourse, discourses disorder, disorders disposition, dispositions, dispositional distress, distresses, distressed, distressing drug, drugs dysfunction, dysfunctions, dysfunctional ecology, ecologies, ecological efficacy elevate, elevates, elevated, elevating, elevator, elevators elicit, elicits, elicited, eliciting emergency, emergent, emergencies emotion, emotions, emotional encyclopedia, encyclopedias, encyclopedic engage, engages, engaging, engaged epidemiology episode, episodes era, eras esteem, esteems, esteemed et al. etc. executive, executives, execution extrinsic eyewitness, eyewitnesses fatigue, fatigues feedback, feedbacks feminist, feminists fluid, fluids forensic foster, fosters, fostered, fostering genetic, genetics glossary, glossaries graduate, graduates, graduated, graduating handbook, handbooks harass, harasses, harassed, harassing, harassment healthcare homosexuality hostility, hostilities humor, humorous, humor, humorous hygiene hyperactive, hyperactivity hypertension hypnosis, hypotheses immune, immunity impair, impairing, impairs, impaired, impairment, impairments indigenous infant, infants infraction, infractions infection, infectious, infections insane, insanity insomnia intellect, intellects, intellectual interdependence, interdependent interdisciplinary internet interview, interviewing, interviews inventory, inventories laboratory, laboratories, laboratorial leisure, lineup, lineups literate, illiterate, literacy locus, loci longitudinal maladaptive marital maternal medication, medications mentor, mentors, mentoring mood, moods, moody morbid, morbidity, morbidities mortality, mortalities multiple muscle, muscles nausea nervous neural, neurology, neurological, neuropsychological, neuroscience normative obese, obesity onset, onsets optimal, optimum, optimism oral overview, overviews, overviewed, overviewing panic pathway, pathways peer, peers periphery, peripheral personnel physician, physicians physiology, physiological placebo, placebos population, populations pregnant, pregnancy, pregnancies prescribe, prescribes, prescribed, prescription prevalence, prevalent private, privacy profile, profiles prognosis, prognostic prominent, prominence prosocial psychiatric, psychiatry, psychoanalysis, psychoanalytic, psychodynamic, psychometric, psychopathology, psychophysiological, psychosocial, psychosomatic, psychotherapy quantitative questionnaire, questionnaires radiotherapy rape, rapes, raped, raping recipient, recipients recruitment, recruit, recruits, recruited, rehabilitate, rehabilitates, rehabilitated, rehabilitation, rehabilitations relapse, relapses, relapsed, relapsing respire, respire, respired, respiring, respiration, respirator, respirators, respiratory retrieve, retrieves, retrieved, retrieving, retrieval, retrievals routine, routines salient, salience scholar, scholars score, scores simultaneous, simultaneously sociocultural, socioeconomic, sociology span, spans, spanning spatial species spinal stereotype, stereotypes stigma, stigmas stimulate, stimulates, stimulating, stimulated, stimulation, stimulus, stimuli strain, strains substantial suicide superior, superiority surgery, surgeries, surgical, surgeon, surgeons susceptible, susceptibility symptom, symptoms testimony therapy, therapeutic, therapies, therapist, therapists tolerate, tolerates, tolerated, tolerating, tolerance traffic, traffics trait, traits treatment, treatments turnover urban verbal, nonverbal versus victim, victims virus, viruses vocation, vocations, vocational vulnerability, vulnerable wellbeing willing, willingness withdraw, withdrawal workplace worldwide

Appendix B: High frequency technical words in equine veterinary corpus

acne, acnes, acned antenatal artery, arteries arthritis, arthritic asthma autonomic cardiac cardiovascular cerebral chemotherapy cortex cortisol dementiaepilepsy lancet, lancets lesion, lesions myocardial nicotine oncology, oncological pathology pediatric, pediatrics, pediatrician, pediatrician sprostateplenumpsoriasis somatic syndrome, syndromes transplant, transplants, transplanted, transplanting, transplantation, transplantations trauma, traumas, traumatic, posttraumatic