2017-07-01

Comparing the AWL and AVL in Textbooks from an Intensive English Program

Michelle Morgan Hernandez
Brigham Young University

Comparing the AWL and AVL in Textbooks from an Intensive English Program

Michelle Morgan Hernandez

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Arts

Dee I. Gardner, Chair
Mark Edward Davies
K. James Hartshorn

Department of Linguistics and English Language

Brigham Young University

ABSTRACT

Comparing the AWL and AVL in Textbooks from an Intensive English Program

Michelle Morgan Hernandez
Department of Linguistics and English Language, BYU
Master of Arts

Academic vocabulary is an important determiner of academic success for both native and non-native speakers of English (Corson, 1997; Gardner, 2013; Hsueh-chao & Nation, 2000). In an attempt to address this need, Coxhead (2000) developed the Academic Word List (AWL)—a list of words common across a range of academic disciplines; however, Gardner & Davies (2014) identified potential limitations in the AWL and have more recently produced their own list of core academic vocabulary—the Academic Vocabulary List (AVL). This study compares the occurrences of the AWL and AVL word families in an intensive English program (IEP) corpus of 50 texts to determine which list has the best overall coverage, frequency, and range in the corpus. While the results show a strong presence of both lists in the IEP corpus, the AVL outperforms the AWL in every measure analyzed in the study. Suggestions for instruction and future research regarding these lists are provided.

Key Words: Academic Vocabulary, Academic Word List, AWL, Academic Vocabulary List, AVL, Intensive English Program, Corpus

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude for everyone who has supported me throughout my graduate experience, especially all those who have assisted with my thesis.

First, I would like to thank my family for all the support and inspiration that they have provided. In particular my parents, siblings, husband and son, without these individuals I would not have been able to complete my thesis.

Second, I would like to thank the faculty in the Linguistics and TESOL department at Brigham Young University for all their support. I would especially like to thank my chair, Dr. Dee Gardner for going beyond the usual to assist me in this process. I would also like to thank my other committee members for their support and guidance in my research.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

According to Gardner (2013), words are "the fuel of language," meaning that without sufficient vocabulary, communication cannot succeed. This is true not only for verbal communication, but also for information contained in a text. Specifically, text becomes incomprehensible when less than 95% of the words are unfamiliar to readers (Hsueh-chao & Nation, 2000; Nation, 2001; Nation, 2004), which often occurs in university settings for students with inadequate academic vocabulary (Corson, 1997; Gardner, 2013; Hsueh-chao & Nation, 2000). Many non-native English speakers (NNESs) and English language teachers struggle to overcome the daunting task of deciding which academic words are important to learn. Consider this example:

> Digital techniques and signals are used in nearly all new designs and system, yet analog issues and characteristics continue to be an inherent and unavoidable part of the system design and challenge. (Schweber, 2002, p. 3)

Which academic words should be learned to facilitate NNES reading comprehension, and in what order? Linguists have researched the challenge of developing academic reading skills (Biemiller, 2003; Klink & Loveland, 2015) and have created different academic word lists for NNESs (Campion & Elly, 1971; Coxhead, 2000; Gardner & Davies, 2014; Ghadessy, 1979; Praninskas, 1972). Among these lists is the Academic Word List (AWL) (Coxhead, 2000), which is accepted and used by many NNESs and teachers. In fact, a few authors of English as a second language (ESL) textbooks have included the AWL in indexes and glossaries, or have incorporated the words throughout the text to assist students with learning academic vocabulary (e.g., Beglar & Murray, 2009; Jeffries & Mikulecky, 2014). However, some linguists have

questioned whether the AWL accurately represents academic language (Chen & Ge, 2007; Gardner & Davies, 2014), due to its usage of word families and its disregard of words contained in the General Service List (GSL) (Gardner & Davies, 2014). In responding to these concerns, Gardner and Davies (2014) created the Academic Vocabulary List (AVL).

This study will compare the AWL and the AVL in a context familiar to NNESs: an intensive English program (IEP). The IEP used in this study is intended to prepare students for two settings: everyday life and university study. Consequently, the IEP is divided into two programs: Foundations, with the objective of preparation for everyday life; and Academic, with the objective of preparation for university study. The texts used in this IEP contain a variety of subjects and difficulty levels typically encountered in these contexts. Combined, the texts used in this IEP constitute the IEP corpus. This study's purpose is to compare and examine the AWL and the AVL quantitatively and qualitatively through data gathered from the IEP corpus to determine which list more accurately represents core academic language.

CHAPTER TWO

LITERATURE REVIEW

Importance of Vocabulary

Vocabulary is essential for both native speakers of English and NNESs. Gardner (2013) calls vocabulary "the fuel of language," and this is especially true in academic settings where academic vocabulary is crucial for success (Beck, Perfetti, & McKeown, 1982; Biemiller, 1999; Biemiller & Slonim, 2001; Chall, Jacobs, & Baldwin, 1990; Hart & Riley, 2003; Townsend, Filippini, Collins, & Biancarosa, 2012). Furthermore, academic vocabulary knowledge is a determining factor in the success of students on entrance exams such as the ACT, SAT, GRE, and GMAT (Gardner, 2013).

As is the case with native speakers, academic vocabulary knowledge often determines the success of NNESs in universities; however, unlike native speakers, NNESs do not have the benefit of time and exposure to learn academic vocabulary. It is estimated that native English-speaking seventh grade students know between 4,760 (Dupuy, 1974) and 51,000 words (Smith, 1941), and these numbers increase by the time they begin university study. Universities not only require NNESs to have basic conversational skills and vocabulary knowledge, but knowledge of academic rhetoric. This can be a challenge since NNESs need to read a word in several different contexts multiple times (David, 2010), and understand 95-98% of the accompanying text to be able to comprehend and learn new vocabulary (Hsueh-chao & Nation, 2000; Nation, 2001; Nation, 2004). For a NNES to not only learn thousands of words through reading texts, but also to have multiple encounters with each new word may take years. Since many NNESs do not have sufficient time to learn the vocabulary necessary for university study in a traditional way, they should begin their study of academic language in a more focused manner.

Core Vocabulary

Many researchers have questioned what vocabulary NNESs need to learn (e.g., Gardner, 2013; Webb & Chang, 2012) and whether some vocabulary words are more valuable to know than others. Gardner (2013) posits that not all vocabulary is equal because some words are used in different circumstances than others and often have different purposes. These differences are often reflected in word lists that are created to facilitate vocabulary learning and teaching purposes (Campion & Elley, 1971; Coxhead, 2000; West, 1953; Xue & Nation, 1984).

West (1953) created the General Service List (GSL), which is composed of 2,000 core general word families. These word families were selected from a corpus of 2.5 million words mainly based on frequency. While the GSL is accepted as a representation of written English language, some researchers have questioned this acceptance and whether the GSL adequately represents written language, and if so, how extensively (Carter, 2012; Engels, 1968; Hirsh & Nation, 1992; Nation, 2001; Nation, 2004; Nation & Kyongho, 1995; Richards, 1974; Sutarsyah, Nation & Kennedy, 1995).

Although the GSL represents general core vocabulary, it does not address what constitutes an academic core, leaving NNESs to search for a list to assist them in acquiring the academic vocabulary imperative for university study. In response to this need, several academic vocabulary lists have been created (Campion & Elley, 1971; Coxhead, 2000; Gardner & Davies, 2014; Ghadessy, 1979; Lynn, 1973; Praninskas, 1972). One of the first extensive lists was the University Word List (UWL) (Xue & Nation, 1984). This list was created through the compilation of four academic word lists: Campion and Elley's (1971) academic word list, Praninskas's (1972) American University Word List, Lynn's (1973) academic word list, and Ghadessy's (1979) academic word list. These lists were all created by hand before computers

were capable of compiling linguistic corpora. However, since the creation of the UWL, technological advancements have been made; thus, creation of improved academic vocabulary lists has become possible.

The Academic Word List (AWL)

Coxhead (2000) noted that the UWL was created from several prior vocabulary lists that had their own problems, one of them being the small corpus sizes used to identify academic vocabulary. In response to this and other concerns, she created the Academic Word List (AWL).

Since the main criticism of the UWL was the small corpus base, Coxhead started her word list by researching the criteria for a corpus—which texts and additional materials should be included in a corpus? In addition, she investigated which words should be included from these texts. Coxhead wanted to create a corpus that represented the size and scope of academic texts (Coxhead, 2000; Sinclair, 1991) and would reflect the diversity of texts found in academic writing by using texts from various disciplines written by a variety of authors (Atkins, Clear, & Ostler, 1992; Biber, 1993; Coxhead, 2000; Sinclair, 1991). Furthermore, the corpus needed to have different categories of texts equally represented. To accomplish this, Coxhead included "28 subject areas organized into 7 general areas within each of four disciplines: arts, commerce, law, and science" (Coxhead, 2000, p. 216).

A key aspect of Coxhead's corpus is the quantity of words contained in it. The number of words in the UWL was limited by the gathering method (Coxhead, 2000). Collecting words and texts by hand takes time and does not allow a large corpus base to be examined. However, because of the development of computers, corpora can include a larger word count. Coxhead's corpus contains 3.5 million running words or tokens (p. 217). This number was the target, based on research stating that a corpus needs to have at least 3.5 million running words to have a word

family occur 100 times in the corpus (Coxhead, 2000; Francis & Kucera, 1982). Such a word count would allow around 25 word occurrences of a word family in each of the four disciplines: arts, commerce, law, and science. This ratio is important because the corpus needs to have a sufficiently large sample size to allow a reasonable frequency of academic words.

The final issue that Coxhead (2000) addressed during the development of the AWL was what to consider as a word. She examined research and concluded that if NNESs learned word roots with common affixes and prefixes, they would be able to learn the definition of words related to each other (Nagy, Anderson, Schommer, Scott, & Stallman, 1989; West, 1953; Zue & Nation, 1984). Coxhead decided to use words with this type of relationship, known as word families, as a basis for the AWL. One condition specified by Coxhead was that the prefixes and suffixes included in the AWL should only be those that are attached to a root that can stand on its own. For example, words such as *concept, conception*, and *conceptual* are considered a word family and, therefore, a word. On the other hand, words such as *specify* and *special* are not in the same family because there is no stand-alone root (Coxhead, 2000, p. 218).

By following these criteria, Coxhead created the AWL. The corpus used to create this list contained 414 texts with 3,513,330 running words from four disciplines and 28 subject areas (see below).

Table 1
Subject Areas Used in Corpus

Arts	Commerce	Law	Science
Education	Accounting	Constitutional	Biology
History	Economics	Criminal	Chemistry
Linguistics	Finance	Family Medicolegal	Computer science
Philosophy	Industrial relations	International	Geography
Politics	Management	Pure commercial	Geology
Sociology	Marketing	Quasi-commercial	Mathematics
	Public policy	Rights and remedies	Physics

(Coxhead, 2000)

The words in the corpus that are found in the General Service List (GSL) were not considered in the count for high-frequency academic words (Coxhead, 2000; Coxhead, 2011; Gardner & Davies, 2014). Coxhead included several additional criteria that words needed to fulfill to be a part of the AWL, including (1) specialized occurrence, meaning that they are not one of the first 2,000 most frequent occurring words, (2) range, meaning that a word family needs to occur 10 times in each of the four disciplines and in 15 or more of the subject areas, and (3) frequency, meaning that word families need to occur 100 times in the academic corpus (Coxhead, 2000).

Based on these criteria, the AWL consists of 570 word families and covered 10% of the academic corpus that Coxhead created. Furthermore, the list covered 8.5% of a second academic corpus used to reinsure the word list's validity (Coxhead, 2000). Coxhead compiled this second corpus using the same criteria as the first corpus, and it was composed of 678,000 tokens from materials that were either gathered too late to be used in the original corpus or were part of a subject area that already had enough material in the original corpus.

Since its creation, the AWL has been used in several textbooks for ESL courses, such as *Advanced Reading Power* (Mikulecky & Jeffries, 2007) and *Contemporary Topics* (Kisslinger, 2009). In addition, the list has been used in various online educational websites (e.g., AWL Exercises Homepage, 2016; Cole, 2010).

Criticisms of the AWL

Although the AWL has played an extensive role in the world of vocabulary acquisition and instruction in ESL settings (Coxhead, 2011) and in primary and secondary schools in the USA (Baumann & Graves, 2010; Hiebert & Lubliner, 2008; Nagy & Townsend, 2012; Thomas, 2013), researchers have questioned if the AWL adequately represents core vocabulary in

academic texts (Chen & Ge, 2007; Gardner & Davies, 2014). Gardner and Davies (2014) identified two specific areas of AWL criticism: the AWL's usage of word families to determine the frequency of words and the relationship between the AWL and the GSL.

The first criticism that Gardner and Davies (2014) addressed was the usage of word families. As noted previously, word families are groups of words that contain the same root with inflectional and derivational affixes added. For example, the word family *analyse* (British spelling) consists of *analysed*, *analyser*, *analysers*, *analyses*, *analysing*, *analysis*, *analyst*, *analysts*, *analytic*, *analytical*, *analytically*, *analyze*, *analyzed*, *analyzes*, and *analyzing*. Since the creation of the UWL, word families have been a common method of organizing words in a vocabulary list (Coxhead, 2000). Gardner and Davies (2014), however, suggested that it is difficult for ESL learners to associate all the words in a word family.

The first criticism is that many words change meaning from one word form to another in a word family (Nagy & Townsend, 2012). In addition, these words can change meaning between different disciplines (Hyland & Tse, 2007). An example is the word *major* (Hyland & Tse, 2007), which has drastically varying definitions. In the military, major is a rank; in education, a subject area to focus study; in sports, the highest league to play; and in music, a distinctive scale or chord to be played. These changes can be confusing and difficult for learners to grasp. Gardner and Davies (2014) indicate that part of this meaning problem is also caused because word families ignore the different parts of speech that can change the meaning of a word. An example of this is the word family *institute*: -*institute* (verb) to set up or establish, and *institute* (noun) a building or society. These definitions might appear similar to native speakers, but for a non-native English speaker, these meanings could be difficult to associate with each other without explicit instruction.

Table 2
Inflectional and Derivational Morphemes

Inflectional		Derivational	
Morpheme	Word	Morpheme	Word
-s	Group	-tion	Relate (verb)
	Groups		Relation (noun)

In addition, other research has suggested that derivational knowledge (affixation that changes the part of speech and/or meaning of the base word) comes later than inflectional knowledge (affixation that does not change part of speech but indicates grammatical function—for example, plurality) (Gardner, 2007; Nippold & Sun, 2008). Consequently, some words in a word family are more difficult for English learners to identify as being part of the same word family. Table 2 provides an example of a simple inflectional change (*group* to *groups*) versus a more complex derivational change (*relate*, the verb, to *relation*, the noun).

The second criticism mentioned by Gardner and Davies (2014) is the relationship between the AWL and the GSL. According to Coxhead (2000) "the AWL was built on top of the GSL" (Gardner & Davies, 2014, p. 4). This decision was based on the assumption that the GSL, containing general high-frequency words of English, should precede any academic word list in terms of what learners need to know to function in the language (Coxhead, 2000). However, there are three important considerations regarding this relationship between the GSL and AWL: first, the GSL was created many years ago and is therefore considered outdated by some; second, by excluding the GSL from consideration in the AWL, many potentially important core academic words might be omitted; lastly, as Gardner and Davies (2014) demonstrated, the AWL itself contains many words found to be high frequency in large modern corpora like the *Corpus of Contemporary American English* (COCA) (Davies, 2012). In fact, they show that 451 of the 570 AWL word families are in the 4,000 most frequent words of COCA, with 41% appearing in

the 2,000 most frequent words of COCA. Despite these shortcomings, Gardner and Davies (2014) acknowledge that the AWL has done much good in the English teaching world and was certainly an improvement over prior lists in establishing a core academic vocabulary.

Academic Vocabulary List (AVL)

The criticisms of the AWL outlined in the previous section were considered as the Academic Vocabulary List (AVL) was being created (Gardner & Davies, 2014), and the AVL was based on the following standards:

1. The new list must initially be determined by using lemmas, not word families. Subsequent groupings of the list into families may be warranted for certain instructional and research purposes.
2. The new list must be based on a large and representative corpus of academic English, covering many important academic disciplines.
3. The new list must be statistically derived (using both frequency and dispersion statistics) from a large and balanced corpus consisting of both academic and non-academic materials. The corpus must be large enough and the statistics powerful enough to be able to separate academic core words (those that appear in the vast majority of the various academic disciplines) from general high-frequency words (those that appear with roughly equal and high frequency across all major registers of the larger corpus, including the academic register), as well as from academic technical words (those that appear in a narrow range of academic disciplines).
4. The academic materials in the larger corpus, as well as the non-academic materials to which it will be compared, must represent contemporary English, not dated materials from 20 to 100 years ago. Otherwise, the validity of the new list could be questioned.

> 5. The new list must be tested against both academic and non-academic corpora, or corpus-derived lists, to determine its validity and reliability as a list of core academic words. (p. 312)

These criteria guided the researchers in the creation of the AVL. The first criterion was met by using lemmas (base form with inflectional affixes) instead of word families to group and count words. Words in the COCA corpus are already tagged for grammatical parts of speech by the CLAWS tagger, making this lemmatization process much easier. The second and fourth criteria were met by utilizing a 120 million-word academic subcorpus within the larger 425 million-word COCA—a modern and up-to-date corpus of English. The academic subcorpus consisted of nine major academic disciplines: education; humanities; history; social science; law and political science; science and technology; medicine and health, business and finance; and philosophy, religion and psychology.

The third criterion was fulfilled through several statistical tests which separated academic words from other words in the corpus (requiring academic words to appear 50% more often in academic materials than in the general materials), then separating core academic words from technical academic words by requiring that core academic words have a range of 20% of their expected frequency in at least seven of the academic disciplines. This range requirement helped eliminate technical and area-specific vocabulary from the list of core academic vocabulary. Furthermore, the researchers required the words to meet a dispersion value of at least 0.80, meaning that the words had to have fairly equal distributions across the academic disciplines, and to meet a "discipline measure" (p. 316) that a word could not occur more than three times the expected frequency in any one discipline. These four statistical tests controlled for general

high-frequency, technical, and discipline-specific words appearing in the *Academic Vocabulary List* (AVL).

The fifth and final criterion was met by testing the AVL against academic and non-academic materials and comparing coverage of the AVL to coverage of the AWL in two large academic corpora (COCA academic—120+ million words; BNC—32+ million words). In order to perform these tests, it was necessary to convert the AVL lemmas to word families so that direct comparisons with word-family based lists could be made. The first case study demonstrated that the AVL consistently covered a much higher percentage of words in academic materials (COCA, 13.8%; BNC, 13.7%) than in newspapers (COCA, 8.0%; BNC, 7.0%) and in fiction (COCA, 3.4%; BNC, 3.4%). The second case study demonstrated that the AVL consistently covered more words than the AWL in COCA academic (AVL, 13.8%; AWL 7.2%) and in BNC academic (AVL, 13.7%; and AWL, 6.9%). These data from the case studies demonstrated a significant difference in coverage between the AVL and the AWL. However, the results are from research involving large mega-corpora, and lack a certain degree of practical validity when it comes to the academic needs of NNESs in actual instructional settings.

To address this issue of practical validity, Newman (2016) compared the AWL and the AVL in the Academic Textbook Corpus (ATC). This corpus consisted of nine texts from 8th grade to introductory university courses in the disciplines of American history, mathematics, and physical sciences. Newman's study indicated that the AVL covered more of the examined texts than the AWL. The purpose of this study is to address the same issues involving AWL and AVL vocabulary in another practical setting—an Intensive English Program (IEP).

Current Study

Following Newman (2016), the current study compares coverage of the AWL and AVL in an actual instructional environment—an IEP curriculum at a major university. The following are the research questions posed in the current study:

1. To what extent do the AWL and AVL represent the vocabulary used in an actual IEP textbook-based corpus?
2. What are the quantitative and qualitative similarities (shared word families) between the AWL and the AVL word families found in the IEP corpus?
3. What are the quantitative and qualitative differences (unique word families) between the AWL and the AVL word families found in the IEP corpus?

CHAPTER THREE

METHODS

This chapter describes the processes used to compare the AWL and the AVL in an authentic IEP context. This includes the processes of compiling the selected corpus and comparing the two lists, as well as the reasoning behind each step taken.

Creation of the Corpus

To answer the proposed research questions, an IEP corpus was created from the textbooks being used in the IEP at the time of this study. The IEP in this study consisted of two course options: Foundations and Academic. These two course options were further divided into levels (see Table 3), each with specific objectives corresponding to the overarching program goals: prepare NNESs to navigate daily life (Foundations) or prepare NNESs for university study (Academic).

The IEP corpus was created using all the textbooks that were part of the regular curriculum of this IEP. The corpus consisted of 50 texts (see Appendix A) totaling 1,639,182 running words. The nature of the Foundations course texts is to provide general simplified contexts for language development. This course contained 18 of the 50 texts. The nature of the Academic course texts is to prepare for university study and consisted of the remaining 32 texts.

Table 3
Courses and Levels of the Intensive English Program with ACTFL equivalent

Foundations		Academic	
		Academic Preparation	Intermediate Low/Mid
Foundations A	Novice Mid	Academic A	Intermediate Mid/High
Foundations B	Novice High	Academic B	Advanced Low
Foundations C	Intermediate Low	University Preparation	Advanced Mid

Converting texts. Texts that were not already in Portable Document Format (PDF) were scanned. All texts were then converted into .txt documents using the Adobe Acrobat program in order to be in the proper format for analytical software. All .txt documents were edited for any spelling errors that occurred during transcription and were saved on a computer to create the IEP corpus.

AWL and AVL Word Lists

The word lists that were used as a basis for this study were the 570 word families of the AWL and the top 570 word families of the AVL. The AVL was originally created using lemmas (base form with inflectional affixes) and not word families (base form with derivational and inflectional affixes). To accurately compare the two lists, however, the AVL needed to be in the form of word families. The AWL could not be converted from word families to lemmas because the frequency data needed to convert to lemmas is not available and would change the AWL. As described previously, the word families of the AVL were created by Gardner and Davies (2014). This was done with the aid of Paul Nation's 20,000+ word families by merging the word families with the database. This alternative AVL format was created primarily for comparison purposes with other lists based on word families, which is why it was used in the current study to compare with the AWL.

One important limitation is that using the AVL in the form of word families causes overinflated frequencies for academic word families containing highly frequent general words. An example is the word *use*, which has both academic and general word forms in its word family. The general word form of *use* is *use* (a verb), which is highly frequent, whereas the primary academic form of the word is *use* (a noun). Since the verb form of the word is a frequent word across genres (e.g., fiction and newspapers, as well as academic), the frequency count for this

word family is most likely exaggerated. For example, if the word family *use* occurs in the corpus 5,000 times but the verb *use* frequency count is 3,500, then the noun *use* (academic usage) may have only occurred 1,500 times. Therefore, the academic usage of the word family *use* is inflated. However, it is important to note that this inflation of academic usage can apply to word family counts in the AWL as well. This important limitation of the study will also be addressed in other sections of this thesis.

Program and Data Collection Procedures

The Range program (Heatley, Nation, & Coxhead, 2002) was used to collect data for the quantitative and qualitative comparisons of the AVL and the AWL. This program was designed to compute the frequency of word lists in any given text. The program originally came formatted with three base-word lists: (1) the first 1,000 word families in the General Service List (GSL), (2) the second 1,000 word families in the GSL, and (3) the 570 word families in the AWL. To analyze the AWL, only base word list 3 was used; the other two lists were simply not selected when using the Range program. To analyze the AVL, the base word lists in the Range program were modified to contain these two base lists: (1) the top 570 word families of the AVL, and (2) the remaining 1,421 word families of the AVL. Table 4 summarizes the base word lists analyzed in the Range program for the AWL and the AVL word lists.

The Range program outputs the following information for the provided word families: the number of types (different words), the number of tokens (times each type occurs), word families

Table 4
Base Word Lists Used in the Range Program

List	Words used in Range Program
Academic Word List	570 AWL word families
Academic Vocabulary List	570 top AVL word families
	1,421 remaining AVL word families

(base words with inflectionally and derivationally related forms), range (the number of texts the word family appears in), word family frequency (combined number of tokens for base word and each base word with inflectional and derivational affixes), and coverage percentages, of the word lists. In addition, the program has the ability to order the word family output according to word family frequency, word family range, and word family alphabetical.

Data was collected from the Range program by inputting, in a two-step process, the three base word lists outlined previously in Table 4. First, the Range program ordered the lists' words found in the IEP corpus according to word family frequency; second, the Range program ordered the lists' words according to word family range. This two-step process allowed the word families of the AWL and the AVL to be compared quantitatively by investigating their similarities and differences. The qualitative comparison was conducted through an examination of the characteristics of the shared and unique word families.

CHAPTER FOUR

RESULTS AND DISCUSSION

Academic Core Vocabulary in the IEP

This study investigated the similarities and differences between the AWL and AVL word families found in the IEP corpus of BYU's ELC curriculum. The research questions stated previously stimulated an investigation into the overall coverage, frequency, and range of the AWL and the AVL in the IEP. An analysis of the results revealed differences between the lists in these three areas.

Coverage. The coverage of these lists in the IEP was best understood through a comparison of the two lists and their differences (see Table 5). First, both lists demonstrated an increase in coverage from the Foundations Program to the Academic Program of the IEP. This suggested that both lists represented the academic language present in the IEP. However, the AVL consistently had higher levels of coverage (tokens, types, and families) than the AWL. Specifically note the coverage of tokens in the IEP, where the AWL had coverage of 4.01% and the AVL had coverage of 10.20%. This 6.19% difference in coverage is noteworthy, particularly since the IEP textbook corpus consists of only 1,639,182 words. Furthermore, all 570 of the top AVL word families appeared in the IEP corpus, whereas only 566 AWL word families are represented.

Table 5
AWL and AVL coverage in the IEP

	Academic Word List			Academic Vocabulary List		
Program	Tokens	Types	Families	Tokens	Types	Families
Foundations	1.94%	5.52%	481	7.21%%	7.92%	512
Academic	5.07%	6.18%	561	11.73%	7.37%	570
IEP	4.01%	5.23%	566	10.20%	6.24%	570

Table 6
Average Word Family Frequency

	Academic Word List	Academic Vocabulary List
Foundations	22	78
Academic	98	223
IEP	116	293

Frequency. The average word family frequency is higher for the AVL than the AWL (see Table 6). The average word family frequency was calculated by totaling the frequency of all word families and dividing that number by the total number of different word families from each list. The average word family frequency for the AWL in the IEP was 116, whereas for the AVL it was 293. This means that, on average, there are 177 more occurrences of AVL families than AWL word families in this IEP corpus, which is a substantial difference.

Range. The average range for each word family was higher for the AVL than the AWL in this IEP context (see Table 7). To calculate the average range, the total range for each word family (50 possible texts) was added and divided by the number of word families found in the corpus. Similar to the overall coverage and average frequency, the average word family range was higher in each program and in the total IEP corpus for the AVL (21) when compared to the AWL (14).

To summarize, the AVL consistently outperforms the AWL in terms of coverage, frequency, and range of core academic vocabulary in the IEP corpus. The following sections

Table 7
Average Word Family Range

	Academic Word List	Academic Vocabulary List
Foundations	3	5
Academic	11	16
IEP	14	21

compare the actual words from the two lists that are found in the IEP corpus to determine if qualitative differences exist as well.

Words in the AWL and the AVL

The first qualitative investigation conducted examined the words that appeared in the IEP from both the AWL and the AVL (see Table 8). The words that were considered had a frequency of 500+ or a range of 29+. These shared word families have high frequencies (number of occurrences), ranging from 702 on the high end to 110 on the low end. The average frequency for these word families was 305. The fact that they occur in both lists suggested that they should be considered core academic vocabulary that is important for pre-university NNESs to learn.

The fact that only 24 qualifying word families (500+ frequency or 29+ range) in the IEP corpus are found in both the AWL and AVL (4.2% of a possible 570) also suggests that the contents of the two lists are quite different. However, in addition to these shared qualifying word families, there were an additional 237 shared word families found in the IEP. These word families had an average frequency of 108. These 237 lower frequency and range word families combined with the 24 in Table 8 bring the shared total to 261, or only 46% of a possible 570 word families—another indication that the two lists are quite different.

Table 8
AWL and AVL Word Families Frequency Found in the IEP

Word	Freq.	Word	Freq.	Word	Freq.
Affect	245	Focus	290	Process	412
Author	341	Identify	281	Publish	282
Available	288	Involve	225	Rely	110
Benefit	160	Link	134	Region	113
Contact	195	Locate	198	Require	330
Create	660	Occur	323	Research	702
Design	359	Percent	434	Similar	503
Goal	173	Period	357	Survive	211

Words Families with High Ranges in the AWL and the AVL

Several word families unique to the AWL or the AVL with a range of 29 texts or more are listed in Table 9. The range of 29 was selected since it was greater than 50% of the texts and provided adequate word families for comparison. It is clear that the AVL with 121 word families has many more of these high range unique word families than the AWL (13). The average range of the unique AVL word families on this list (35.9) is also higher than the AWL (34.4), with the highest words in the AVL (*give, help*, and *need*) appearing in 49 of 50 texts and the highest word in the AWL (*found*) appearing in 45 of 50 texts.

At this juncture, it is crucial to reiterate that this study is a "word family" comparison, and that words like *give* often made the AVL as lemmas in other forms that are more academic in nature. For instance, *give* (the word family in this study) was actually *given*, an adjective, in the original lemma-based AVL (e.g., *In any given circumstance*). When *given* was converted to a word family to make it comparable with AWL families, it was subsumed under the liberal family "give," which makes no distinctions between noun forms of *give* (*It didn't have any give to it*; *I take that as a given*), verb forms (*I gave it to him; It was given to her*), adjectives (*In any given circumstance*), or any other pertinent part of speech. Thus, as they also do on the AWL, such families tend to exaggerate "academic" coverage. In the case of *give*, for example, the family frequency is 2,420 in the IEP corpus, but the more academic usage of *given* (adjective) only occurs 252 times. This is one of the primary criticisms of lists based on word families (e.g., Gardner & Davies, 2014; Nagy & Townsend, 2012). However, some word families are primarily academic throughout. For example, the *experience* word family occurs 515 times in the IEP corpus, with three academically-salient members accounting for most of that total: *experience*, a noun (361); *experience*, a verb (126); and *experienced*, an adjective (19).

Table 9
AVL and AWL Word Families Found in the IEP and Ranges

Academic Word List		Academic Vocabulary List							
Word	Rng	Word	Rng	Word	Rng	Word	Rng	Word	Rng
Found	45	Give	48	Wide	39	Common	34	Author*	30
Final	41	Help	48	General	38	Direct	34	Basic	30
Job	41	Need	48	However	38	Increase	34	Compare	30
Area	40	Use	47	Inform	38	Introduce	34	Connect	30
Chapter	36	Find	47	Language	38	Publish*	34	Discuss	30
Create*	36	Change	46	Present	38	Relate	34	Goal*	30
Publish*	34	Mean	46	Develop	37	Science	34	Limit	30
Similar*	34	High	46	Manage	37	Similar*	34	Modern	30
Percent*	33	Part	46	University	37	Success	34	Occur*	30
Process*	33	Important	46	Century	36	Support	34	Rate	30
Edit	32	Grow	46	Create*	36	System	34	Require*	30
Identify*	32	Move	46	Describe	36	Active	33	Shape	30
Remove	32	IE	45	Discover	36	Consider	33	Suggest	30
Research*	32	Large	45	Experience	36	Active	33	Term	30
Adult	31	Both	44	Nature	36	Consider	33	Test	30
Involve*	31	Group	44	Provide	36	Contain	33	Various	30
Author*	30	State	44	Subject	36	Difference	33	Accept	29
Energy	30	Act	43	View	36	Percent*	33	Account	29
Goal*	30	Follow	43	Above	35	Practice	33	Africa	29
Major	30	Strong	43	Actual	35	Process*	33	Apply	29
Occur*	30	Difficult	42	Condition	35	Identify*	32	Argue	29
Require*	30	Example	42	Europe	35	Level	32	Combine	29
Respond	30	Understand	42	Explain	35	Prefer	32	Contact*	29
Team	30	Add	41	Govern	35	Produce	32	Content	29
Contact*	29	Interest	41	Member	35	Research*	32	Current	29
Credit	29	Plan	41	Necessary	35	Effect	31	Degree	29
Period*	29	Study	41	Note	35	Gain	31	Exist	29
Rely*	29	Table	41	Protect	35	Germany	31	Figure	29
Survive*	29	Continue	40	Report	35	Human	31	Likely	29
		Include	40	Result	35	Involve*	31	Period*	29
		Center	39	Type	35	Organize	31	Rely*	29
		Form	39	Value	35	Particular	31	Social	29
		History	39	Base	34	Product	31	Survive*	29
		Low	39	Collect	34	Tool	31	Variety	29
		Whole	39						

** indicates words that appear in both lists (16 shared; 13 unique AWL; 121 unique AVL)*

Unique AWL and AVL Word Families in the IEP Corpus (Frequency Analysis)

A frequency of 500+ was chosen to compare unique AWL and AVL word families in the IEP corpus (see Table 10). Several important conclusions can be drawn from this comparison.

Table 10
Most Frequent AWL and AVL Word Families and Corresponding Frequencies

Academic Word List		Academic Vocabulary List					
Family	Frq	Family	Frq	Family	Frq	Family	Frq
Chapter	2729	IE	21424	Add	980	Create*	654
Chart	2497	Use	7238	Active	975	Test	657
Clause	1538	Example	2922	High	890	Explain	657
Partner	1339	Part	2503	Include	887	Introduce	656
Topic	1222	Give	2420	However	886	Common	653
Paragraph	1192	Mean	2418	Develop	884	Describe	621
Job	1184	Help	2334	Science	881	Effect	620
Economy*	837	Find	2273	Interest	861	Product	619
Found	787	Need	2172	Economy*	837	Social	612
Culture*	778	Group	1880	Human	833	Govern	610
Computer	751	Practice	1880	Grow	817	Section*	602
Lecture	740	State	1865	Culture*	778	Difficult	583
Research*	702	Study	1669	Both	747	Passage	581
Define*	677	Form	1649	Large	737	Mental*	568
Create*	654	Inform	1557	Subject	722	History	567
Section*	602	Follow	1531	Act	717	Type	557
Mental*	568	Change	1522	Future	714	Source*	551
Final	567	Present	1397	University	710	Depress	547
Area	566	Important	1335	Compare	709	Continue	536
Source*	551	Understand	1255	General	704	Review	531
Similar*	502	Express	1134	Research*	702	Nature	529
		Discuss	1077	Plan	698	Report	527
		Language	1070	Support	696	Result	525
		Note	1064	Organize	694	Experience	515
		Move	1021	Define*	677	Similar*	502
		Unit	998	Progress	666	Enjoy	502

** indicates words that appear in both lists (12 unique AWL; 69 unique AVL; 9 shared)*

First, similar to the word families with the highest ranges, the AVL had a much higher number of word families with high frequencies (500+). There are 69 unique AVL word families that met the threshold and only 12 unique AWL families, with the average frequency of the unique word families also showing marked disparities between the two lists (AVL—1,438 vs. AWL—1,238).

Second, the top five most frequent word families demonstrated a notable difference in frequencies between the two lists. The five most frequent word families in the AWL were *chapter* (2,729), *chart* (2,497), *clause* (1,538), *partner* (1,339), and *topic* (1,222). In contrast, the five most frequent AVL word families had substantially higher frequencies were *IE* (21,424), *use* (7,238), *example* (2,922), *part* (2,503), and *give* (2,420), with all five of the top AVL word families having higher frequencies than the highest frequency AWL word family (*chapter*).

Third, nine word families with 500+ frequencies appeared in both the AWL and the AVL (see words with asterisks in Table 10). These word families demonstrated similarities and differences between the lists. Of note is the fact that these shared word families have much higher list rankings in the AWL than in the AVL. For instance, *economy* was ranked as the eighth most frequent word family in the AWL, but as the 35th most frequent in the AVL. Likewise, the word family *culture* is ranked tenth in the AWL and 38th in the AVL. These rankings provide additional support that the AVL and AWL lists are quite different, with the AVL list containing many higher frequency words than the AWL.

Other AVL Word Families in the IEP Corpus

Because only the top 570 AVL word families were used in this study to compare with the 570 word families of the AWL, this last section examined the impact of AVL families in the IEP corpus beyond the top 570. To clarify, Gardner and Davies (2014) consolidated their list of 3,015 AVL lemmas into 1,991 total AVL word families. This discussion deals with the families between 571 and 1,991. To limit the discussion to the most impactful of these word families, it was decided to examine only those with a range of at least 29 and frequencies above 500. There were 38 AVL word families that met these requirements (see

Table *11*).

Table 11
AVL Word Families Ranges and Frequency not in the top 570

Family	Ran	Freq.	Family	Ran	Freq.	Family	Ran	Freq.
Word	44	5190	Name	42	1185	Skill	33	728
Work	47	4207	Problem	42	1116	Order	42	711
Read	46	4172	Operate	43	1084	Nation	34	701
Question	40	3350	Cause	39	1016	True	42	698
Know	48	3346	Choose	41	1001	Circle	30	665
Think	48	3314	Great	44	985	Lead	43	656
First	49	1996	Simple	37	910	Reason	38	578
Complete	38	1876	Better	46	890	Short	43	561
Class	34	1862	Number	42	862	Situation	29	550
Correct	29	1652	Main	37	858	Stand	41	541
Last	45	1534	Open	44	808	Point	39	530
Learn	47	1461	Able	42	820	Second	42	520
Family	48	1222	Program	31	752			

The word family with the highest frequency was *word* occurring 5,190 times with a range of 44. This was the third most frequent word in the entire AVL. The impact of this word may be due to the texts used in the IEP, which are written for NNESs and contain explanations of grammar and vocabulary. However, like many of the word families compared in this study (both AVL and AWL), it is also possible that non-academic senses of *word* are inflating the numbers. Like the word family *give*, discussed above, *word* first made the AVL as two different lemmas with strongly academic senses—*word*, the verb, as in *How should we <u>word</u> this*, and *wording*, the noun, as in *Change the <u>wording</u>*. When it was made into a word family for comparison purposes, these two lemmas and all others containing *word* as a root were consolidated, making a large family of *word*, the verb, noun, etc. with all possible uses and meanings, some of which are strongly academic, and some of which are simply general English. This same phenomenon can be useful in understanding the high frequencies of *work* (4,207) and *read* (4,172), and many other families on both the AVL and AWL examined in this study. This is certainly a limitation

of any study utilizing word families as the construct of word (Gardner, 2007). Despite this limitation, it is clear that the expanded AVL (word families beyond the top 570) have a substantial impact in the IEP corpus, and may warrant further examination.

CHAPTER FIVE

CONCLUSION

The AWL and the AVL were both created with the purpose of representing the most common (core) academic vocabulary in the English language, particularly as it occurs in university settings and the texts used in those environments. The primary objective of the current study was to investigate whether these lists are also representative of texts used in a pre-university (IEP) setting, and how the two lists compared in this regard. This was viewed as a practical examination of the two lists, similar to Newman's (2016) study of K-12 textbooks in history, science, and math. The findings indicate that although both lists were significantly covered in the IEP corpus, the AVL had a greater breadth and depth of coverage of the texts specific to this IEP, suggesting that this may be the case in similar programs and contexts. This was demonstrated in terms of general coverage, number of different word families, average frequencies of those families, and average ranges of those families.

A second objective of the study regarded similar AWL and AVL words contained in the IEP corpus. Since both lists were created for the purpose of representing academic vocabulary found in university settings, it was assumed that many of the words would be shared between the two lists. However, this did not turn out to be the case, as only 22 of 570 possible word families of the IEP corpus were found in both the AWL and AVL with high frequency (500+) and/or range (29+). Of course, these 22 would be prime candidates for NNES vocabulary study.

A third purpose of this research was to investigate the differences between the AWL and AVL in accounting for academic core words in the IEP. Differences were observed by comparing the word families that met certain criteria. One difference was the number of words that met the range criterion of 29 texts or more. The AWL had 13 unique word families meeting

this criterion, while the AVL had 121, with only 16 shared between the two lists. Another difference was the number of word families that fulfilled the criteria of a frequency of 500 or more. The AVL had 69 unique word families (not including word families beyond the top 570) that met the criteria, while the AWL had only 12. Furthermore, beyond the top 570 word families, the AVL had 38 word families that met both the range and frequency criteria previously mentioned.

The findings of this study suggest that although the AWL demonstrated noteworthy coverage of the academic vocabulary found in the IEP corpus, the AVL has much higher coverage in all areas analyzed, often by substantial margins. Thus, the AVL may be a more comprehensive list for NNESs to use in terms of focusing their studies on the most high-priority academic vocabulary in an IEP setting.

Limitations

It is necessary to acknowledge limitations that occurred in this study.

Word families. First, the use of word families greatly limits the efficacy of the AVL. This was necessary in order to make the AVL compatible and comparable with the AWL for comparison purposes. However, this process eliminated one of the key differences between the lists and a primary purpose for the AVL's creation. This limitation was also necessary in order to utilize a program such as Range that organizes information based on word families and is unable to recognize the difference between grammatical parts of speech.

Range software. A second limitation in this study is the Range program's inability to differentiate between academic and non-academic senses of words contained in the corpus. Essentially, the program is only capable of counting word forms and predetermined groupings of those forms into word families. Words such as *use* (in the academic noun sense—e.g., *The most*

common use of a text) are grouped together with their non-academic counterparts (e.g., *use* in the verb sense—e.g., *She will use another recipe*). The AWL and the AVL are meant to be academic lists, but the Range program counts all word forms without regard for their parts of speech or meanings. Therefore, the different usages of a word family were counted in both the frequency and range counts, resulting in overinflated data.

Implications

This study investigated the coverage of the AWL and AVL in an IEP setting. According to the findings of the study, the AVL has greater coverage of the English used in this setting (and likely other similar settings) in terms of overall coverage, range, and frequency. Therefore, it would be beneficial for NNESs to focus their studies on the words contained in the AVL. The words in the AVL appear more frequently in academic texts than words in the AWL, but both lists are certainly better than no lists or with no starting points for addressing crucial academic word knowledge.

Suggestions for Future Research

There is a need for additional studies comparing the AWL and the AVL so that students and practitioners may know how to study and instruct most effectively in the realm of academic vocabulary. The few studies currently present in the literature do not provide sufficient evidence that the AVL demonstrates a greater coverage of academic vocabulary. It is recommended that future researchers conduct similar studies in different academic settings, such as at other IEPs, as well as at universities and in K-12 settings. Furthermore, the development of computer programs to compare academic core lemmas with academic core word families could provide insightful data, and might lead to a more nuanced understanding of the possibilities and limitations of academic core word lists.

Summary

This research was conducted in an intensive environment where NNESs study textbooks that teach English as a second language to help them function in both daily life and academic settings. The AWL and the AVL were created to assist NNESs with their academic language. As this study suggests, the AVL appears to have greater coverage of the academic vocabulary that is present in a typical IEP corpus. For this reason, the AVL should be the focus of study for NNESs desiring to enter academic settings equipped with adequate knowledge of the academic vocabulary they will encounter.

APPENDIX A

Book included in the English Language Corpus

Foundations A Books

Jeffries, L., & Mikulecky, B. S. (2010). *Basic reading power*. White Plains, NY: Pearson Longman. (49,969 words)

Johnson, T. (1998). *Farm life long ago*. Austin, TX: Steck-Vaughn Co. (448 words)

Myers, A. (1998). *Hot air balloons*. Austin, TX: Steck-Vaughn Co. (469 words)

Folse, K. S. (2011). *Oxford American dictionary: vocabulary builder*. New York, NY: Oxford University Press. (24,341 words)

Tanka, J., & Baker, L. R. (2014). *Interactions listening/speaking* (6th ed.). New York, NY: McGraw Hill. (461 words)

Foundations B Books

Anderson, N. J. (2013). *Active skills for reading 1* (3rd ed.). Boston, MA: National Geographic Learning. (42,105 words)

Azar, B. S., & Hagen, S. A. (2005). *Basic English grammar* (3rd ed.). White Plains, NY: Pearson Education. (61,160 words)

Hardy-Gould, J. (2008). *Henry VIII and his six wives*. Oxford, England: Oxford University Press. (7,176 words)

Vicary, T. (2008). *The elephant man*. Oxford, England: Oxford University Press. (8,200 words)

Foundations C Books

Azar, B. S., Koch, R. S., & Hagen, S. A. (2011). *Fundamentals of English grammar* (4th ed.). White Plains, NY: Pearson Education. (74,946 words)

Azar, B. S., Koch, R. S., & Hagen, S. A. (2011). *Fundamentals of English grammar: answer key* (4th ed.). White Plains, NY: Pearson Education. (107,350 words)

Coerr, E. (1977). *Sadako*. New York, NY: Puffin Books. (7,749 words)

Funke, C., & McKee, D. (2013). *The wizard of Oz*. London, England: Puffin. (43,637 words)

Hammontree, M., & Doremus, R. (2014). *Albert Einstein*. New York, NY: Aladdin. (23,666 words)

Jeffries, L., & Mikulecky, B. S. (2010). *Reading power 2*. White Plains, NY: Pearson Longman. (73,145)

Kummer, P. K. (2000). *The pioneer way*. Austin, TX: Steck-Vaughn. (3,789 words)

Stine, M. (1992). *Laura Ingalls Wilder, Pioneer Girl*. New York, NY: Parachute Press, inc. (17,163 words)

Stadelhofen, M. M. (1982). *The freedom side*. Syracuse, NY: New Readers Press. (10,328 words)

Academic Preparation

Akinyemi, R. (2008). *Nelson Mandela*. New York, NY: Oxford University Press. (16,012)

Akinyemi, R. (2008). *Rainforest*. New York, NY: Oxford University Press. (6,434)

Blass, L., & Hartmann, P. (2007). *Quest 1. listening and speaking*. New York, NY: McGraw-Hill. (61,640)

Davies, P. A. (2008). *Information Technology*. New York, NY: Oxford University Press. (9,413)

Discovering Careers for Your Future: Advertising & Marketing. (2005). New York, NY: Ferguson. (19,972 words)

Higgins, M. M. (2007). *Benjamin Franklin: revolutionary inventor*. New York, NY: Sterling. (21,435 words)

Hirschmann, K. (2004). *Dolphins*. Boston, MA: Thomson Heinle. (4,977 words)

McLean, A. C. (2008). *Martin Luther King*. New York, NY: Oxford University Press. (9,815 words)

Stewart, S. (2008). *Recycling*. New York, NY: Oxford University Press. (8,789 words)

Academic A Books

Andrews, L. W. (2004). *Emotional intelligence*. New York, NY: F. Watts. (12,807 words)

Azar, B. S., & Hagen, S. A. (2009). *Understanding and using English grammar* (6th ed.). White Plains, NY: Pearson Longman. (90,963 words)

Azar, B. S., & Hagen, S. A. (2009). *Understanding and using English grammar: Answer key* (6th ed.). White Plains, NY: Pearson Longman. (177,747 words)

Bella, L. L. (2010). *World financial meltdown*. New York, NY: Rosen Central Publ. (8,611 words)

Bingham, J. (2009). *Post-impressionism*. Lewes, East Sussex: Heinemann Library. (8,482 words)

Blass, L. (2006). *Quest 2 listening and speaking*. New York, NY: McGraw-Hill. (59,510 words)

Hartmann, P. (2007). *Quest 2: Reading and writing* (2nd ed.). New York, NY: McGraw-Hill. (52,487 words)

Freedman, J. (2010). *First bank account and first investments smarts*. New York: Rosen Pub. (8,564 words)

Hall, M. C. (2008). *Leonardo da vinci*. Edina, MN: ABDO. (18,330 words)

Lüsted, M. A. (2011). *Social networking: MySpace, Facebook, & Twitter*. Edina, MN: ABDO Pub. Co. (16,651 words)

Ramen, F. (2007). *Drug abuse and society prescription drugs*. New York, NY: Rosen Publishing Group. (8,454 words)

Silverstein, A., Silverstein, V. B., & Nunn, L. S. (2009). *The depression and bipolar disorder update*. Berkeley Heights, NJ: Enslow. (16261 words)

Solomon, S. E. (2010). *Mental disorder*. Edina, MN: ABDO. (16,173 words)

Trueit, T. S. (2004). *Dreams and sleep*. New York, NY: Franklin Watts. (11,840 words)

Academic B Books

Calhoun, Y. (2007). *The environment in the news*. New York, NY: Chelsea House. (23,795 words)

Ching, J. (2009). *Outsourcing U.S. jobs*. New York, NY: Rosen Pub. (9,181 words)

Hartmann, P. (2006). *Quest 3 reading and writing*. New York, NY: McGraw-Hill. (93,325 words)

Hartmann, P. (2006) *Quest 3 listening and speaking*. New York, NY: McGraw-Hill. (93,229 words)

Lew, K. (2011). *Evolution: the adaptation and survival of species*. New York, NY: Rosen Pub. (8,997 words)

Lusted, M. A. (2010). *Poverty*. St. Paul, MN: ABDO. (17,104 words)

Watson, C. (2011). *Unloved and endangered animals: what you can do*. Berkeley Heights, NJ: Enslow. (17,086 words)

Lynch, D. (2003). *J.R.R. Tolkien: creator of languages and legends*. New York, NY: Watts. (21,190 words)

University Prep Books

Smalley, R. L., Ruetten, M. K., & Kozyrev, J. R. (2011). *Refining composition skills* (6th ed.). Independence, KY: Heinle ELT. (133,806 words)

REFERENCES

Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. Literary and Linguistic Computing, 7, 1–16.

AWL Exercises Homepage. (n.d.). Retrieved January 13, 2016, from http://www.englishvocabularyexercieses.com/AWL/

Bauman, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent and Adult Literacy,54*, 4-12.

Beck, I. L., Perfetti, C., and McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. Journal of Educational Psychology, 74, 506-521.

Beglar, D., & Murray, N. (2009). *Contemporary topics. academic listening and note-taking skills*. White Plains, NY: Pearson Longman.

Biber, D. (1993). Representativeness in corpus design. Literary and Linguistic Computing, 8, 243–257.

Biemiller, A. (1999). Language and reading success (Vol. 5). Cambridge, MA: Brookline Books.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. Reading Psychology 24, 323-35.

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. Journal of Educational Psychology, 93, 498–520

Carter, R. 2012. Vocabulary: Applied Linguistic Perspectives. Routledge

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

Chall, J.S., Jacobs, V.A., & Baldwin, L.E. (1990). The Reading Crisis: Why Poor Children Fall Behind. Cambridge, Mass.: Harvard University Press.

Chen, Q. & Ge, G.C (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). English for Specific Purposes, 26, 502-514.

Cole, D. (2010, November 14). Academic word list exercises. Retrieved January 13, 2016, from http://dcielts.com/ielts-vocabulary/awl-exercises/

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

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly,34*(2), 213-238. doi:10.2307/3587951

Coxhead, A. (2011). The Academic Word List 10 Years On: Research and Teaching Implications. *TESOL Quarterly,45*(2), 355-362. doi:10.5054/tq.2011.254528.

David, J. (2010). Closing the Vocabulary Gap. *Educational Leadership, 67*(6), 85-86.

Davies, M. 2012. 'Corpus of Contemporary American English (1990–2012),' available at http://corpus.byu.edu/coca/. Accessed June 2012

Dupuy, H.P. (1974). The rationale, development and standardization of a basic word vocabulary test. Washington, DC: U.S. Government Printing Office. (DHEW Publication No. HRA 74-1334).

Engels, L. K. (1968). The fallacy of word-counts. IRAL-International Review of Applied Linguistics in Language Teaching, 6(1-4), 213-232.

Francis, W. N., & Kucera, H. (1982). Frequency analysis of English usage. Boston: Houghton Mifflin.

Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, *28*(2), 241-265.

Gardner, D. (2013). Exploring vocabulary: Language in action. New York, NY: Routledge.

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

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

Hart, B., & Risley, T. R. (2003). The early catastrophe. The 30-million-word gap. American Educator, 27, 4-9.

Heatley, A. Nation, I.S.P, & Coxhead, A. (2002) RANGE program. Retrieved from http://www.victoria.ac.nz/lals/about/staff/paul-nation

Hiebert, E. H. & Lubliner, S. (2008). The nature, learning, and instruction of general academic vocabulary. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about vocabulary instruction*. (pp. 106-29). International Reading Association.

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

Hsueh-chao, M.H. & Nation, P. (2000). Unknown vocabulary density and reading comprehension. Reading in a Foreign Language, 13(1), 403-430.

Hyland, K. and P. Tse. 2007. 'Is there an ''Academic Vocabulary''?,' TESOL Quarterly 41: 235–53

Jeffries, L., & Mikulecky, B. S. (2014). *Advanced Reading Power 4: extensive reading, vocabulary building, comprehension skills, reading fluency*. White Plains, NY: Pearson Education.

Klink, P., & Loveland, T. (2015, November 1). Vocabulary Development in Technology and Engineering Education. *Technology and Engineering Teacher*, 8-13.

Kisslinger, E. (2009). *Contemporary topics* (3rd ed.). White Plains, NY: Pearson Longman.

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

Mikulecky, B. S., & Jeffries, L. (2007). *Advanced reading power: Extensive reading, vocabulary building, comprehension skills, reading faster* (2nd ed.). White Plains, NY: Longman.

Nagy, W., Anderson, R., Schommer, M., Scott, J. A., & Stallman, A. (1989). Morphological families in the internal lexicon. Reading Research Quarterly, 24, 262– 281. general service list of English words. London: Longman, Green.

Nagy, W., & Townsend, D. (2012). Words as Tools: Learning Academic Vocabulary as Language Acquisition. *Reading Research Quarterly,47*(1), 91-108. doi:10.1002/rrq.011

Nation, P. (2001). Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.

Nation, P. (2004). A study of the most frequent word families in the British National Corpus. Vocabulary in a second language: Selection, acquisition, and testing, (pp. 3-13).

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

Newman, J. A. (2016). A Corpus-Based Comparison of the Academic Word List and the Academic Vocabulary List. *BYU ScholarsArchive*.

Nippold, M. A., & Sun, L. (2008). Knowledge of Morphologically Complex Words: A Developmental Study of Older Children and Young Adolescents. *Language Speech and Hearing Services in Schools,39*(3), 365-373. doi:10.1044/0161-1461(2008/034).

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

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

Schweber, W. L. (2002). *Electronic communication systems: a complete course*. Upple Saddle River, NJ: Prentice Hall.

Sinclair, J. (1991). Corpus, concordance and collocation. Oxford: Oxford University Press.

Smith, M.K. (1941). Measurement of the size of general English vocabulary through the elementary grades and high school. Genetic Psychology Monographs, 24, 311-34

Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based case study. RELC Journal, 25(2), 34-50.

Thomas, A. (2013, March 20). Teach Students the All-Purpose Academic Word List. Retrieved May 24, 2017, from http://www.cdl.org/articles/teach-students-the-all-purpose-academic-word-list/

Townsend, D., Filippini, A., Collins, P., & Biancarosa, G. (2012). Evidence for the Importance of Academic Word Knowledge for the Academic Achievement of Diverse Middle School Students. *The Elementary School Journal,112*(3), 497-518. doi:10.1086/663301

Webb, S. A., & Chang, A. C. (2012). Second Language Vocabulary Growth. *RELC Journal,43*(1), 113-126. doi:10.1177/0033688212439367

West, M. 1953. A General Service List of English Words. Longman, Green.

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