



All Theses and Dissertations

2010-04-15

Essential 500 Wordlist for the Foundations Program at the Brigham Young University's English Language Center

Inho Jung

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Linguistics Commons](#)

BYU ScholarsArchive Citation

Jung, Inho, "Essential 500 Wordlist for the Foundations Program at the Brigham Young University's English Language Center" (2010). *All Theses and Dissertations*. 2116.

<https://scholarsarchive.byu.edu/etd/2116>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu.

Essential 500 Wordlist for the Foundations Program
at the Brigham Young University's
English Language Center

Inho Jung

A project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of
Master of Arts

Norman Evans, Chair
Dee Gardner
K. James Hartshorn

Department of Linguistics and English Language
Brigham Young University

August 2010

Copyright © 2010 Inho Jung

All Rights Reserved

ABSTRACT

Essential 500 Wordlist for the Foundations Program
at Brigham Young University's
English Language Center

Inho Jung

Department of Linguistics and English Language
Master of Arts

In order to help ESL students in the Foundations Program at BYU's ELC, I have developed a 500-word list of essential English vocabulary. This list attempts to reach one of the goals of the institute's curriculum, which is to develop systematic, integrated vocabulary instruction.

Published literature reveals that studying with a list in conjunction with explicit instruction can enhance L2 vocabulary learning when the instruction is systematically integrated with other skills and activities. The literature also shows the importance of learning the most frequent and essential words first, which can be found in the General Service List and the Academic Word List according to learners' needs.

In order to develop an essential vocabulary list of 500 words for the Foundations Program at the ELC, I used several strategies. First, to investigate which words in their textbooks the students use, I analyzed the graded readers and listening books required in the curriculum of the Foundations Program. Second, after investigating the texts of the textbooks, the second step was to conduct a survey in order to investigate learners' lexical knowledge. To develop a reliable survey, reliable research strategies were conducted. The first survey was conducted at the end of Fall Semester 2009. Following the first survey, the second survey was conducted at the beginning of Winter Semester 2010. The administration of the two surveys revealed the students' self-reported knowledge about specific vocabulary items in the lists.

Following the administration and analysis of the two surveys, I generated the final 500-word list for the students in the Foundations Program at BYU's ELC. The words were based on the students' needs and knowledge, and were generated based on the essential words from the GSL and the AWL in order to meet the goals of the curriculum of the Foundations Program.

Keywords: essential English vocabulary, lists, survey, the GSL, the AWL

ACKNOWLEDGMENTS

First of all, I would like to thank my Heavenly Father for His guidance and comfort each step of the way, in my entering, continuing, and completing this Master's program. Without His guidance, I would not have been able to have the great experiences I have gone through at this great LDS school, Brigham Young University. Without His inspiration and comfort, I would not have been able to complete this Master's program.

I am grateful to my wife and children who have supported me in many ways. They are a source of perpetuating my education. I am also grateful to my committee members, who have encouraged and advised me to complete this project with great insights. I am especially grateful to Dr. Norman Evans, my Project Chair, for his patience, and his strict but encouraging guidance in the process of this Master's project.

Table of Contents

ACKNOWLEDGMENTS	ii
Introduction.....	1
Background	1
Purpose	2
Literature Review.....	3
Positive Values of Learning from Context in Vocabulary Acquisition	3
Limitations of Guessing from Context for L2 Learners.....	4
Strategy of Integration.....	6
Studying Words from Lists	7
The GSL (General Service List).....	8
The AWL (Academic Word List)	9
Specialized Corpus.....	11
Project Development.....	12
Selection of Texts: Graded Readers and Listening Packets	13
Data Collection Process	15
Data Analysis of the Specialized Corpus	15
Creating a List of Words	17
Trimming from 800 words to 675 words	18
First Survey.....	19
Developing a Reliable Survey.....	19
Results of the First Survey	21
Lessons from the First Survey.....	22
Second Survey	23
Generating Words for the Second Survey.....	23
Developing the Second Survey	24
Results of the Second Survey.....	24
Final Word Compilation.....	25
Applications and Suggestions.....	26
Applications	26
Suggestions for Further Research	27
Conclusion	28
References.....	30
Appendix A: Sample of words analyzed by Range and Frequency Analysis.....	33

Appendix B: Samples of High Frequency Words from the Off-list	35
Appendix C: Function Words and Numbers.....	36
Appendix D: 675 Words compiled for the First Survey	37
Appendix E: First Vocabulary Survey	38
Appendix F: Result Samples and Analyses of the First Survey	40
Appendix G: 675 Words compiled for the Second Survey.....	42
Appendix H: Second Vocabulary Survey	44
Appendix I: Final list of 500 Essential Words for the Foundations Program.....	46

Introduction

Background

Until the fall semester of 2009, the English Language Center (ELC) at Brigham Young University (BYU) had no systematic approach to vocabulary instruction in its curriculum. The main vocabulary instruction in the classes before had been to develop the skill, called guessing from context. Admittedly, depending on the textbooks and teachers, some explicit vocabulary instruction had been employed erratically but there was no systematic, program-wide approach to vocabulary instruction. However, since the ELC's curriculum had a drastic change in 2005, many of the programs within the curriculum have been developed. Among them, Content Based Instruction was launched in 2008, and in Fall 2009 the programs of the ELC shifted from five levels to two main programs of three levels each: the English Foundations Program, consisting of levels A, B, and C and the Academic English Program consisting of levels A, B, and C. A systematic type of vocabulary instruction also became a critical part of the new ELC curriculum. The Academic Word List (AWL) developed by Coxhead (2000) was chosen to be the main focus of vocabulary instruction for the Academic English Program. The list was purposely divided into sub lists and implemented into the Academic English Program; the vocabulary instruction was integrated into the four skills (reading, listening, speaking, and writing) for which classes are offered.

Nevertheless, systematic vocabulary instruction in the Foundations Program, unlike that of the Academic English program, has yet to be developed for many reasons. However, as the new curriculum is being established, the urgent need for systematic vocabulary instruction for the students in the Foundations Program has surfaced more than ever.

Purpose

The Foundations Program at the ELC has not had a systematic approach to vocabulary instruction whereas the Academic English Program has already made the first move to utilize the AWL. Thus, the General Service List (GSL) was first suggested to use for a vocabulary teaching focus for the Foundations Program. In general, the GSL developed by West (1953) has served many English educators and administrators to provide stable foundations in educating English learners. It seems that the GSL is a good resource to start with to acquire introductory vocabulary knowledge for general English. However, its size (2,000 word families) may be too large for the students in the ELC's Foundations Program to handle because the 2,000 word families could mean more than 10,000 individual words. To explain, the list consists of headwords, which means that this list assumes that a person knows other members of the word family. For example, the *advice* family consists of *advises, advising, advised, advisor, advisors, adviser, advisers, advise, and advisory*. Accordingly, attempting to teach all the words in the list may be unrealistic.

Teaching the AWL to the students in the Foundations Program could also pose some serious challenges. The first challenge would be the fact that they would not be ready to study the AWL because of their low proficiency and its specialized words for academic texts. The second challenge would be the fact that the AWL has already been implemented into the Academic English Program at the ELC and is currently being used in its classes. On account of the current challenges that the ELC has confronted, there has not been a systematic approach to vocabulary instruction for the Foundations Program even though many of the current research studies (Bengelei & Paribahkt, 2004; Nassaji, 2003; Webb, 2007) assert that the lower the proficiency, the more explicit instruction is needed. Thus, the students in the Foundations Program at the ELC *do* need their own manageable wordlist that can help them to meet the goals of the ELC curriculum effectively. Accordingly, the objectives of this project are (1) to build a

corpus based on the texts that the students have to study, (2) to identify the students' lexical knowledge, and (3) to generate a list of the most essential 500 words they need to know in order to meet the goals and objectives of the classes in the Foundations Program.

Literature Review

The main purposes of this literature review are to recognize the most common strategies of effective vocabulary instruction for L2 learners and to reveal the roles and efficacy of studying wordlists. Prior to discussing the roles of studying vocabulary with wordlists and their efficacy, it is imperative to raise key issues pertinent to one of the most popular vocabulary learning strategies — namely, guessing from context. Therefore, this review will discuss the positive values and limitations of guessing from context, high versus low proficiency, the strategy of integration, and finally the GSL, the AWL, and the use of specialized wordlists.

Positive Values of Learning from Context in Vocabulary Acquisition

Krashen (1986) asserts that “vocabulary is most efficiently attained by comprehensible input in the form of reading” (p. 440). He strongly believes that reading for pleasure outweighs all other strategies of vocabulary learning. According to his Comprehensible Input Hypothesis, true language acquisition will occur only through comprehensible reading. In particular, vocabulary will develop naturally through reading, which should be the primary means of vocabulary acquisition. He strongly supports the notion of gaining vocabulary through context. Some of the values of learning from context are discussed below.

One of the greatest values of context is the connection of a word with an instance. Anderson, Stevens, Shifrin, & Osborn (1978) point out an important value of context. They believe that native speakers recall the original context of where they first encountered a new word when they read or hear the word a second time. In other words, rich contexts provide them with a concrete and strong connection between words and their previously accumulated

knowledge. This connection can help the learners retain the new word longer and actually use it, whereas rote memorization of a word's definition can only help learners know its meaning and form. Nation (2001) elaborates on this issue, "One very important value of context in learning vocabulary is that a variety of contexts will evoke a variety of enriching instantiations" (p. 241).

These contexts are naturally given to native speakers. They gradually build up contexts and guessing skills during their lifetime. For many years they grow up in different environments, such as schools or homes with rich and natural exposure to various words. Consequently, when native speakers encounter new words through reading books or magazines and listening to movies, music, or talks, they can naturally connect the new words with previous experiences. In this way, they can easily develop an intuition for guessing new words from context throughout their entire lives. Of course, they may learn vocabulary through explicit and intentional learning in educational settings, but the number of such items would be too small to compare to the number that they learn incidentally.

Schatz & Baldwin (1986) claim another important value of guessing from context. They say that this strategy facilitates comprehension without consulting outside sources like dictionaries. For native speakers, it seems very natural to skip new words, but they understand the general content and then acquire new words from context later. There is no doubt about the values of the reading and guessing strategy. Considering this research, can this useful strategy also be applicable to L2 learners?

Limitations of Guessing from Context for L2 Learners

Another important context factor is that many naturally occurring contexts give minimal clues, no clues, or false clues to word meaning. Even though guessing from context may be the most common skill used by L1 learners for vocabulary acquisition, it does not mean that it is the most effective way for L2 learners. There is much research that casts doubt on the practice of

guessing words from context among L2 learners. Sternberg (1987) shows his strong skepticism in the effectiveness of use of context clues for L2 learners. He cautions that it may not be an optimal approach for L2 learners. For example, one study revealed how hard this strategy can be for even L1 learners, and how much harder it could be for L2 learners. Schatz & Baldwin (1986) examined the effectiveness of context clues on 101 native speakers in the 11th and 12th grades. The test consisted of two types: “a-word-in-context test and a word-in-isolation test” (p. 442). Their findings were quite surprising and different from what most reading teachers would expect. The results revealed that context clues did not help the students with low-frequency words and that students had the same level of confusion without the context clues. However, one important finding was that when the word was repeated in context many times, it was easier for them to guess its meaning, but when the word was used to explain new information, it was difficult to guess its meaning.

Their findings suggest some important pedagogical aspects. First, even native speakers have a hard time using context clues in their reading when it comes to guessing the meanings of low-frequency words. This result implies how difficult it would be for non-native speakers. Second, students need to encounter words frequently in context in order to succeed in gaining new words through context. Yet, we cannot extrapolate that L2 learners will be able to guess high frequency words as accurately as L1 learners. It is important to learn from this discovery that teachers have to carefully select materials that contain fewer, low-frequency words and more high-frequency words to intentionally and systematically expose students to the target words.

Even though many previous studies examined the effectiveness of learning from context with different methods and approaches, they only proved that learning from context did not work effectively for L2 learners, due to their limited exposure to the English language and lack of lexical knowledge. So why doesn't this strategy work with L2 learners? The answer may be

found in the statement by Folse (2004) that, “the true pedagogical value of guessing may be for reading comprehension and not for vocabulary learning” (p. 82). He also suggests that ESL teachers may have to teach this as a good reading strategy, and not as a vocabulary development strategy. In order to make this strategy work, ESL teachers should also combine this strategy with explicit instruction or activities focused on words. Nation (2001) also admits the focus of this strategy could be on improving reading skills instead of vocabulary size. Another factor that may impede guessing from context is the learners’ proficiency levels.

Bengelei & Paribahkt (2004) studied the relationship between reading proficiency and level of success in inference. Their study indicated that the higher the proficiency, the better the success in lexical inferencing. They also concluded that low proficiency learners would perform worse in lexical inferencing, and their comprehension would suffer more.

Their study can be interpreted to mean that for beginning learners, learning words through context may not be effective because they are not good at comprehending information due to their lack of semantic knowledge compared to advanced learners. The study suggests that the beginning learners need to develop their vocabulary as soon as possible by explicit learning so that they can develop their skills of guessing from context. Overall, this empirical study reports that context clues are not always beneficial for every level. Thus, in order for the low proficiency learners to make context clues work effectively in their reading for vocabulary acquisition, explicit vocabulary instruction integrated into other strategies like inferencing may be much more effectual.

Strategy of Integration

It seems salient from research (Bengelei & Paribahkt,2004; Schatz & Baldwin,1986; Sternberg,1987) that vocabulary instruction for low proficiency learners should be explicit and integrated with other skills and strategies. To test the positive effects of the strategy of

integration, Wesche & Paribakht (1994) conducted an experiment on two different groups of adult ESL students in the U. S: the reading-only group (reading followed by more reading) and the reading-plus group (reading followed by vocabulary exercises). The result demonstrated that the reading-plus group, representing explicit vocabulary review with exercises, had more gains than the reading-only group, representing incidental vocabulary learning. In addition, the reading-plus group showed better knowledge of the target words than the reading-only group.

This result implies that inferencing words from context may not be as effective for ESL students, and explicit instruction of vocabulary may be more effective for them to acquire productive knowledge of unknown words. What we can also learn from this study is that combining comprehension with explicit vocabulary instruction can enhance both guessing skills and vocabulary knowledge of L2 learners.

Another study examined the effect of the strategy of integration in vocabulary acquisition. To see the effect of vocabulary retention through extensive reading combined with other skills, Laufer (2003) conducted an experiment in which she compared test scores of those who completed a sentence writing activity after reading, and those who merely did extensive reading. The study concluded that the scores of those who completed the sentence writing activity were significantly higher in the short term than those who only read extensively, and they also maintained their higher scores when given a delayed test. The study also found that explicit instruction involves diagnosing the words that learners need to know, presenting words for the first time, elaborating word knowledge, and developing fluency with known words.

Studying Words from Lists

It seems obvious that reading and guessing from context may be helpful in understanding overall context and connecting the words into rich instances, but with explicit vocabulary instruction integrated into other skills or activities such as writing activity or vocabulary exercise,

learners' vocabulary acquisition can be accelerated more. There may be many ways available to teachers of introducing words explicitly to students.

One of the salient ways of explicit vocabulary instruction is teaching from lists. According to Folse (2004), this strategy along with the grammar-translation method has not been welcomed by many language teachers due to its dullness. However, with regard to the return of teaching from lists, Hulstijn states, "vocabulary lists are not 'in,' but we may be seeing a comeback now" (as cited in Folse, 2004, p. 37). Folse (2004) stresses that there is little evidence that using lists hinders L2 learning, which implies that when the lists are used wisely with other activities, they can be very useful. Carter (1987) supports the power of beginners studying words from a list. He adds that novice learners may learn the words better when they are presented in lists. Nation (1993) also claims that for the beginning learners, it is critical to experience a "vocabulary flood," which may successfully be presented as a list.

Finally, it seems ideal that teachers use lists to teach clearly target vocabulary to beginning learners as they integrate them with other strategies and skills to enhance their retention and overcome boredom. There are many lists that have been developed according to the needs of the learners and among these, the most commonly known and professionally developed wordlists are the GSL and the AWL. They will be discussed next.

The GSL (General Service List)

Conservative estimates suggests that there are around 114,000 word families in English, not including proper nouns (Goulden, Nation & Read, 1990), with more liberal estimates suggesting this number may be as high as two million lexemes (words with distinct meanings) (Crystal, 1995). This number can be overwhelming to both native speakers and non-native speakers. Especially for the non-native speakers, trying to learn those words may sound like swimming in an endless ocean without knowing where to go. Furthermore, they do not have the

luxury of time and money to waste in the ocean of words. They need a more viable, focused, and manageable sized list of words to study. In response to this need, a general word list was developed. The GSL by West (1953) contains 2,000 word families that cover almost 80% of the running words that are most frequent in general texts. Nation (2001) claims that this list contains the most crucial words that all beginning learners of English should master because “these words cover a very large proportion of the running words in spoken and written texts and occur in all kinds of uses of the language” (p. 13). For example, Nation (2001) asserts that the first 1,000 words cover around 77% of words in most texts and the second 1,000 words cover approximately 5% of words in most academic texts. He also reveals that the first 1,000 words covers 84.3% and the second 1,000 words cover 6% of words in conversation. This list has 2,000 head words that are of general service for learners and has served many teachers, administrators, and researchers as well in many ways. It is highly recommended that for beginning learners who would like to pursue a college education, the words in the GSL should be the first list of words to master and then the AWL. The chief reason why they are highly recommended is that the words in the two lists do not overlap and cover approximately 90% of words in most academic texts (Nation, 2001).

For these reasons, attempting to develop an essential word list from the GSL for the Foundations Program at the ELC seems prudent, logical, and timely. It is also reasonable and practical to teach the students in the Academic English Program with the AWL.

The AWL (Academic Word List)

After the creation of the GSL by West (1953), many researchers have also created different sets of wordlists for academic purposes. One of the earliest sets was developed by Champion and Elly (1971), which covered 19 academic disciplines. Praninskas (1972) also generated the American University Word List (AUWL) for non-native English speakers in

university. Later Lynn (1973) and Ghadessy (1979) created a word list using student annotations in textbooks. Using all the previous lists, Xue and Nation (1984) developed the University Word List. The list was widely used by teachers and students because the words in the list were most frequent in most academic texts. However, this list was challenged by Coxhead (2000), who pointed out that the list contained many pitfalls and lacked consistent selection principles, and was based on a small data set so that it did not have a variety of topics in balance. Furthermore, she claimed the need for a new academic word list by stating:

There is a need for a new academic word list based on data gathered from a large, well-designed corpus of academic English. The ideal word list would be divided into smaller, frequency-based sublists to aid in the sequencing of teaching and in materials development. (p. 214)

To substantiate her claim, she generated a new academic word list. She selected texts from four disciplines: Arts, Commerce, Law, and Science. Each of the disciplines contained around 875,000 running words equally distributed over eight subject areas. Finally, the Academic Corpus was developed and along with it, a new Academic Word List was also developed, covering approximately 10% of the Academic Corpus and around 86% of the GSL. Moreover, the AWL consists of 570 word families, and “more than 94% of the words in the list occur in 20 or more of 28 subject areas of the Academic Corpus” (Coxhead, 2000, p. 226).

On account of the professional development of the AWL, it has become the most essential list of words for academic texts. Thanks to the new AWL and the GSL, the words that are essential for learners preparing for academic goals are focused, efficient, and manageable.

These two professionally developed lists seem very appropriate for the ELC at BYU where most of the students are eager to enter a university. When two lists are purposefully

adapted and operated, the new curriculum that the ELC's administration has been implementing would be efficient, responsive, and stable.

Specialized Corpus

With the aim of generating essential words for the Foundations Program at the ELC, it is necessary to discuss the matters regarding a specialized corpus. Nation (2001) states that it is possible to create a specialized list of words for specific texts depending on the need of the curriculum such as a list of words for basic speaking, for reading the newspaper, or for business letters. Hunston (2002) also elaborates on a specialized list: "There is no limit to the degree of specialization involved, but the parameters are set to limit the kind of texts included" (p. 14). For example, Carrie Thompson (2005) generated 500 essential words for ESL missionaries by analyzing the texts and selecting the words that were frequently used in them. The words such as *apostasy*, *priesthood*, *atonement* are not found in either the GSL or the AWL, but they are very important and frequent in the texts that the missionaries use.

Another example of a specialized corpus was created by (Schonell et al, 1956). From the corpus, they attempted to generate a special set of wordlists for speaking. In particular, they focused on adult conversations in the workplace. The 2,000 word families they generated covered 99% in their corpus. Later, this hand-made specialized list of 2,000 words was examined and compared with Lexical coverage of Spoken Discourse (CANCODE) by (Adolphs & Schmitt, 2003). CANCODE stands for Cambridge and Nottingham Corpus of Discourse in English, and is a compilation of around 5 million words spoken throughout the U. K and Iceland in various settings from 1994 to 1999. Adolphs & Schmitt (2003) learned that the 2,000 word families still cover approximately 95% in the CANCODE corpus. The above studies imply the necessity of lists that meet the needs of learners with particular purposes in a given context.

In summary, it has been explained that guessing from context has great promise but has certain limitations as well, especially to the L2 beginning learners. Studying words from a list has great potential when it is integrated into other strategies, skills, and activities. Furthermore, among many lists developed, the two lists, the GSL and the AWL have many resourceful values. The words in the lists are absolutely indispensable for general and academic texts. Finally, a specialized corpus is for learners with specific needs. A list of words can be generated from the special corpus to address the special needs of learners.

Project Development

As the studies from the literature review have revealed, learning outcomes can be optimal when explicit vocabulary instruction, accompanied by a professionally developed list—like the GSL or the AWL—is combined with other strategies and skills. In this respect, the administrators and staff at the ELC, who have tried to build up the vocabulary imbedded in the curriculum, are heading in the right direction to making it fully responsive to the needs of the students. Since Fall 2009, the Academic English Program has already implemented the AWL in classes and the words in the list have been systemically taught and presented to the students through different classes and activities. However, for the Foundations Program, adopting a list like the GSL has been a challenge because of its size as well as low levels of student proficiency. Nevertheless, low proficiency students tend to have a fervent desire to learn English as fast as they can because they do not want to waste their time and money by staying at the ELC for a long time. Therefore, this urgent need must be met. Finding the essential words that students must know is the first step to developing a sound curriculum. Much of the curriculum has been developed using the Analysis of needs, Design, Development, Implementation, and Evaluation (ADDIE) model. Obviously, the first step in the current project is the first step in the model: Analysis of needs. In order to find the needs of the students in the Foundations Program, it is

imperative to evaluate the texts the teachers use to teach with and what words their students already know and do not know in the texts. As a result, it is hoped that a special corpus for the program can be developed and a special list of words can be generated from it.

Selection of Texts: Graded Readers and Listening Packets

The very beginning step in knowing what words are in learners' texts is to develop a special corpus. As a consequence, with the purpose of collecting texts, I consulted with the supervisors of the Foundations Program. They informed me about the program, textbooks, and other extensive materials currently in use. For our analysis, we decided to focus only on reading and listening texts, and not to incorporate writing and grammar textbooks. This was because the administration of the ELC decided that reading and listening were the main modes through which explicit vocabulary instruction would be given. The ELC primarily uses writing and speaking activities to transition students' receptive lexical knowledge into productive knowledge.

The Foundations Program uses three textbooks and 65 graded readers to teach reading. These are divided into three different levels. For listening, there were also three levels of listening textbooks. In the initial stage of text selection for a corpus development, it was suggested to select all the textbooks. However, after having a meeting with my committee members, it was suggested that the reading textbooks currently in use should be excluded from the project because of the nature of vocabulary in their texts. They were mainly expository, focusing on formal, academic language, whereas the graded readers and Listening Packets were primarily narrative, emphasizing informal conversational vocabulary. Thus, based on the grounds that the words in the expository texts are quite different from those in narrative texts, the reading textbooks were eliminated from the process of the text selection. In the end, out of the 65 required reading books for the Foundations Program, 38 books were randomly selected and divided into two main categories: fiction and non-fiction. They were again divided by

proficiency level (A, B, C). In addition to the graded readers, the Listening Packets, which were going to be used in all listening classes in Winter 2010, were also selected as texts for the project's corpus development. The Listening Packets consisted of three books for each level. All together, 38 graded readers and three listening books were selected. (See Table 1 for the graded reader titles).

Table 1

Selected Graded Readers used by the English Foundations Program

	Fiction	Non-Fiction
Level A	Frog and Toad Are Friends Frog and Toad All Year Days with Frog and Toad Flying home Mary Queen of Scots Anne of Green Gables Call of the Wild Railway Children	The Coldest Place on Earth Pele A Look at Dogs Amazing Trains Hot Air Balloons
Level B	Robinson Crusoe The Elephant Man Phantom of the Opera Prince and the Pauper Henry VIII and His Six Wives	Lincoln Green Planet Martin Luther King Civil War
Level C	Charlotte's Web Les Miserable Sarah, Plain and Tall The Adventures of Huckleberry Finn Freedom Side Einstein Young Thinker Number the Stars	Mississippi River Mark Twain Energy Spiders World War II Colonial America Pioneers Brain Underground Railroad

Data Collection Process

In order to elicit necessary data, I used Frequency and Range software program. It is a software program developed by Nation (Heatley et. al. , 2002) that provides data on the frequency and range of texts. Because the Frequency and Range software program only allows thirty-two files to be run through at any one time, it was necessary to divide the 38 selected graded readers and three listening books into smaller groups. Subsequently, the 44 books within each level were divided into three groups: fiction, non-fiction, and listening. In order to analyze the selected texts, the first step was to transform the texts into electronic formats. With the help of OmniPage, an optical character recognition application that converts images such as scanned paper documents and PDF files into document file formats, the printed texts in those books were scanned and transformed into electronic text. The texts in the Listening Packets were already in an electronic format so they were ready to be changed into the proper electronic text. The 38 books from the ELC were scanned and turned into electronic text. From this process, a specialized corpus was generated for the Foundations Program.

Data Analysis of the Specialized Corpus

The specialized corpus for the Foundations Program consisted of the texts of 38 graded books and three listening books. In order to analyze the texts, the scanned books were divided into six files which were named as follows: Level A fiction, Level A non-fiction, Level B fiction, Level B non-fiction, Level C fiction, and Level C non-fiction. In addition to the reading materials the listening books, known as Listening Packets were added. They consisted of three files named as follows: Level A, Level B, and Level C. All together there were nine files and they were run through the Frequency and Range software program. The frequency and range analysis of the nine files revealed that there were 3,111,948 tokens, namely, the entire word

count from the texts, including repeated words. The analysis also found 11,529 types, or distinct words. (See Table 2 for the profile of specialized corpus for the Foundations Program).

Table 2

Profile of the Specialized Corpus for the ELC Foundations Program

	Families	Types (%)	Tokens (%)
K1 Words (1-1,000)	970	2,887 (25.49)	255,976 (81.37)
K2 Words (1,001-2,000)	927	2,109 (18.81)	20,287 (6.45)
AWL Words (academic):	394	721 (5.07)	3,171 (1.01)
Off-List Words:	N/A	5,812 (50.63)	35,143 (11.17)
Total	2,291	11,529	311,948

In the specialized corpus for the Foundations Program at the ELC, out of 11,529 types, there were 970 word families from K1 Words, 927 word families from K2 Words, and 394 AWL word families. The total number of word families in the categories of K1, K2, and AWL came to 2,291. (See Appendix A for sample of words analyzed by Range and Frequency Analysis).

Surprisingly, there were a great number of types in the off-list, accounting for approximately 50% of the different words (types) in the corpus. This high percentage of the off-list words in the corpus means that the texts may be very difficult for the students in the program. However, after reexamining the 5,812 types from the off-list, I found that 1,013 of the types consisted of proper nouns and numbers accounting for about 17% of the total off-list types. Most of them were off-list words with high frequency. (See Appendix B for samples of the off-list types). Furthermore, I found 285 non-words, such as *yoursef* and *yonr*, which covered around 4% of the total off-list types. These words were distorted from the scanning process. In addition to these proper nouns, numbers, and non-words, there were still 4,514 content words in the corpus that were from the

off-list, covering around 30% of the entire corpus. These could seriously influence the comprehension of the texts even though most of them were low frequency. These are important words to know; however, compared to the high frequency words from the GSL and the AWL, they are not as commonly used in general contexts. Therefore, considering the students' time and urgent need of mastering the more generally used words from the professionally developed lists, I decided that the off-list words should not be included in the final list development. Instead, I generated a corpus of 2,291 word families based on the K1, K2, and AWL from the texts of nine files.

Creating a List of Words

To create a list of the most frequent words from the corpus, the 2,291 word families from K1, K2, and AWL were put in an EXCEL spreadsheet for range and frequency sorting. The reason these 2,291 word families were selected for the frequency analysis was their pertinence to the goal of the current project, which states, "to generate a list of the most essential 500 words they [students] need to know in order to meet the goals and objectives of the classes at the Foundations Program." The goal of this project was also closely tied to the goals of the ELC's curriculum for the Foundations Program. For Foundations A and B, the goal is to prepare students for "basic interpersonal communication," and the goal for Foundations C is to establish a "language base in which potential future acquisition of academic skills may occur"(<http://elc.byu.edu/curriculum/#>). To fulfill these goals, the administrators and teachers of the ELC first and foremost wanted the learners to master the words in the GSL and the AWL lists, because according to Nation (2001) the words in these two lists cover between 84% and 92% of the words used in conversation, fiction, newspapers, and academic text. He highly recommends teaching the GSL words first, and then the AWL words to novice learners. Based on this reasoning, I generated this corpus of 2,291 words that the students at the program should learn first.

Although a corpus of 2,291 word families had been generated, it was vital to select carefully which words would be put in the final list of 500 words. The basis for selecting 500 words is that 500 may be a feasible number of words for the students in the Foundations Program to study for a semester. As the first step, I used the sorting function of EXCEL in order to select those words that occurred in at least five files and at least 26 times overall. I accomplished this by setting a range cutoff point of occurrence within at least 5 of the files, and then I set a frequency cutoff of at least 26 occurrences overall. Finally, I was able to generate a preliminary list of 800 word families out of the 2,291 word families. After that, I ran them through VocabProfile, an online program (adapted by T. Cobb) that analyzes a text, counts the words in the text, and categorizes them. It shows all the K1, K2, AWL, and off-list words, as well as calculating the ratios of all tokens, types, and word families. Using this program, I outlined the profiles of the 800 word families as follows.

Approximately 78% of the high frequency words in the 800 word families were from the first 1,000 words of the GSL; 20% were from the second 2,000 words, and only 1.6% were from the AWL. As the above data indicates, the highest frequency words in the reading texts for the Foundations Program were from the GSL, taking up 98% of the preliminary list. This result also implies that the publishers of the graded readers and the author of the listening textbooks intentionally chose the words for the appropriate levels. In addition, the high frequency words in the texts were very basic content words like *go, eat, or sleep*, and included function words like *a, an, the, or on*.

Trimming from 800 words to 675 words

These preliminary words needed to be trimmed to be a refined list of words. One way to do this was to eliminate function words and numbers. First, function words were eliminated from the list because due to their nature, they were taught primarily in grammar classes, not in reading

classes at the ELC. As a result, the function words like *between*, *also*, and *almost* were eliminated. Second, numbers were eliminated because they were seemingly too easy for the ELC students. (See Appendix C for the eliminated function words and numbers). Finally, a list of 675 preliminary words was generated, and prepared into a survey to find out how many words in the list the students knew. (See Appendix D for the list of 675 words for the First Survey).

First Survey

Developing a Reliable Survey

Once we refined the list to contain only content words, it was ready for a survey administration. The main purpose of this survey was to find out which words in the list students knew or did not know. It was hoped that after this survey, the most commonly known words to the students at the Foundations Program, could be taken out from the list. In other words, the ideal goal of the survey was to collect 500 word families that were essential and unknown to the students after deleting the known words. For this project, the definition of “unknown” words denotes words that were not known to the students in their receptive knowledge. In order to make a as reliable survey as possible, I created a survey form adapted from the checklist test used by Meara (1989). The checklist test was challenged by the criticism that “there is no way of knowing how validly the test-takers are reporting their knowledge of the words” (Read, 2000, p. 88). This concern was taken into consideration. Nevertheless, others have found when checklist tests were carefully designed with strategies such as adding pseudo words to the test, checklist tests have been found to be reliable enough to generally test students’ vocabulary knowledge (Anderson & Freebody, 1983; Meara & Buxton, 1987). To bring the highest reliability possible to the survey, I utilized several strategies. First, I created four descriptions for the students to choose from. *I know this word, I think I know this word, I am not sure if I know this word, I do not know.* (See Figure 1 for a part of the actual survey).

Figure 1. Sample from the actual survey

	WORD	I know this word	I think I know this word	I am not sure if I know this word	I do not know this word
1	aware				
2	label				
3	blind				
4	penny				
5	crime				
6	glomater				

Instead of using Yes/No type questions, I hoped that this four-answer type of survey would better represent the participants' levels of knowledge. Second, in order to minimize fatigue for the students, I constructed a survey that contained only 50 items. More than 50 items could make students exhausted and they could take it carelessly (Nation, 2001). As a result, it was hoped that they could take the 50 item survey as truthfully as possible. Third, to increase reliability of a survey of fifty items, five pseudo words were intentionally included. In short, there were 45 genuine items and five pseudo words in the survey. Fourth, I wrote succinct instructions in the survey for the students to beware of the pseudo words in the survey to prevent the students from marking the items inaccurately. The final strategy I carried out was a very intricate challenge. It was to figure out how to assemble different survey sets of 45 words out of the 675 words and to make all the students in three levels take them so as to bring out the maximum reliable outcome of the survey. To do this, I designed the survey in a way that all 675 words were used as survey questions for students in all ELC foundation levels at least three or four times. In order to do so, I divided the 675 words into three pools of 225 words. The first pool contained words from the most frequent word to the 225th word, the second from the 226th to the 450th, and the third from the 451st to the 675th. I put 15 words in a survey from each survey pool so that one student in any level could do a survey on words from different frequency

ranks. Even the students at the Foundation level A would be surveyed from the most frequent words to the least frequent words in the list. Finally, I created 15 different sets of surveys to be administered to the 76 students who were enrolled in Fall 2009 at the ELC. (See Appendix E for the first survey).

Results of the First Survey

The survey was administered to 70 out of the 76 students enrolled in the ELC during Fall 2009. In order to analyze the data in the surveys, I assigned a number to each item of a survey: *I know this word* corresponded with the number 4; *I think I know this word* with 3; *I am not sure if I know this word* with 2; and *I do not know this word* with 1. The purpose of this numbering was to rank the words in groups numerically, from the words that students claimed they knew down to the words that they claimed they did not know. In addition, it served to sort out unreliable surveys. (See Table 4 for the cutoff point system).

Table 4

Cutoff Point System

Categories	Descriptions	Given number	Average Cutoff Point Range
Known	I know this word	4	4.0
	I think I know this word	3	3.0-3.9
Unknown	I am not sure if I know this word	2	2.0-2.9
	I do not know this word	1	1.0-1.9

The results of the survey were very surprising. First of all, there were 10 unreliable surveys, which were about 14% of the total surveys issued. All tests having average scores on the five pseudo words above 3 points were considered unreliable surveys. Second, regarding the rest

of the surveys, out of the 675 words surveyed, there were only 4 words that were unknown to the students, which were *shake* (2.5), *centre*(1.8), *lean* (2.5), and *bill* (2.5). In order to judge whether a word was known by the students, any words with scores below a 3.0 average were considered unknown words. An average score of 3.0 or above meant that the words were known to students. I decided to choose this cutoff point because marking a 3.0 or above on the survey indicated that the student felt confident that he or she had encountered that word before, even if he or she were unsure of its meaning out of context. Anything below a 3 indicated that the student either did not know the word, or felt fairly confident that he or she had not previously encountered it in his or her English studies. In short, the students at the Foundations Program claimed through their self-report survey that they knew 99.4% of the 675 words that were found to be the most frequent words from the texts. (See Appendix F for result samples and analyses of the first survey).

Lessons from the First Survey

The results from the first survey taught me vital lessons. One important lesson I learned from the first survey is that almost all the students in the Foundations Program claimed that they knew 624 words out the first 1,000 GSL, which is more than 60%, and 162 words in the second 1,000 GSL, or around 16%. These results imply that they have already acquired most of the first 1,000 words in the GSL and some portion of the second 2,000 words in the GSL. This insight taught me that I should lower the cutoff score of frequency and range and create a new list of words for the second survey without adding words from the first 1,000 words in the GSL. In other words, the students needed to learn lower frequency words from the second 1,000 GSL and apparently already knew some words from the AWL. Overall, the first survey served as a valuable, informative pilot survey.

Second Survey

Generating Words for the Second Survey

After gaining some valuable insights about the students' lexical knowledge in the Foundations Program from the first survey, I generated the second list of words from the same corpus of the 2,230 word families from which the first survey was created. In the process of generating the second list, the frequency within the entire corpus was between 8 and 1 occurrences. The cutoff point for range was between 4 and 1 occurrences across the nine files of the corpus. Within these parameters, there were 771 word families; there were 96 words from K1, 383 words from K2, and 292 words from AWL. First of all, the 96 words from K1 were eliminated based on the fact that students reported that they knew almost all the words in the first survey (99%). Words in the K1 list are of comparable difficulty and occur with comparable frequency. With this in mind, considering that the students were comfortable with 62% of the first 1,000 words in our survey, it would be logical to presume that the students would also be comfortable with the remaining 48% of the K1 words. After eliminating the 96 K1 word families, I created another set of 675 words. (See Appendix G for the entire list words). The 383 words comprised 56.8% of the second wordlist words. These same 383 words comprised 38.3% of K2. The remaining 292 words (43.2%) of the list comprised 51.2% of the words in the AWL. (See Table 3 for the profile of the second wordlist).

Table 3

Profile of the Second Wordlist

	Families	Percent of 675 wordlist	Percentage
K2 Words (1,001-2,000)	383	56.8%	38.3% (K2)
AWL Words (570)	292	43.2%	51.2% (AWL)
Total	675	100%	

Developing the Second Survey

I created the second survey exactly like the first survey, using its design and scoring system. The only difference was the pseudo words. I created five new pseudo words so that the students who took the first survey would not recognize the pseudo words in the second survey. (See Appendix H for the second survey). The student body that took the survey in Winter 2010 differed from that of Fall 2009. In Winter 2010, there were 54 new students who had not taken the first survey. Out of the 83 students enrolled at the ELC, 81 students participated in the second survey.

Results of the Second Survey

The results of the second survey revealed many intriguing insights. (See Table 4 for the results of the second survey).

Table 4

Results of the Second Survey

Descriptions	Given Number	Cutoff Score	Number of Words in the survey	% (675)	K2	AWL
I do not know this word	1	1.0 -1.9	92 words	13.6%	69	23
I am not sure if I know this word	2	2.0 -2.9	235 words	34.8%	145	90
I think I know this word	3	3.0 -3.9	248 words	36.7%	118	130
I know this word	4	4.0	100 words	14.8%	51	49
Total			675 Words	100%	383	285

Of the 81 participants, only 3 students returned unreliable surveys, compared to the 10 unreliable surveys returned from the first administration. As far as the details of their results were concerned, the students were familiar with 348 words; the cutoff score was 3.0 or above. These covered approximately 52% of the second wordlist. They were not confident with the remaining 48% of the words; the cutoff score was below 3.0. Another interesting finding was that the students reported that they knew 169 words out of the 383 K2 words used, which made up about 44% of the total K2 words in the survey. One can deduce from these data that the students probably know approximately 44% of the entire K2 list. Furthermore, regarding the AWL words in the second survey, I found that the students claimed to know 179 words out of 285, or 63% of the AWL words in the survey. From the data, one can also assume that the students would know roughly 63% of the 570 AWL words. One can see that, surprisingly, the students seemed to have a greater familiarity with the AWL vocabulary, than the K2 vocabulary. This result implies that the AWL may not necessarily be more difficult than the K2 in the GSL to learn for the students.

Final Word Compilation

The main goal of this project was to compile a list of 500 essential words that were not a part the students' receptive knowledge in the Foundations Program at the ELC. From the second survey, I learned that students were not familiar or did not know 327 words from the second list of 675 words. These words were automatically placed on the final list of 500. The remaining 173 words needed for the list were taken from the "I think I know" category. I used the words with the lowest average scores from the survey. Finally, I compiled the 500 words, ranking them according to difficulty based upon the results from the survey. (See Appendix I for the final list of 500 words). These words came from the K2 and the AWL. (See Table 5 for more details).

Table 5

Profile of the final 500-word list

	Families	Percent
K2 Words (1,001-2,000)	294	58.80%
AWL Words (academic)	206	41.20%
TOTAL	500	100%

The final 500 words are composed of 294 K2 words and 206 AWL words. They are based on the students' knowledge and generated from the essential words from the GSL and the AWL to meet the curriculum goals of the Foundations Program.

Applications and Suggestions

In this section, I will discuss ideas for application such as incorporating the list into the ELC's curriculum and syllabi. I will also discuss suggestions for further research, such as examining the reliability of the material and doing test item correlations.

Applications

This essential list of 500 words for the Foundations Program at the ELC can be used in various ways. First of all, it can be incorporated into the current curriculum of the ELC. Just as the AWL has been used for the Academic English classes, this list can similarly be utilized to teach the students in the Foundations Program. Since there are 206 words from the AWL words in the list, it may be redundant to teach these words in the Foundations program. Yet, it can be a smooth transition for the students to move to the Academic English Program, where words from the AWL are taught. For example, these 206 words could be taught receptively in the Foundations Program, and then taught productively in the Academic English Program. With this list, they can efficiently concentrate on learning the essential words which they do not know or of which they are not sure.

Furthermore, the teachers of the reading, writing, speaking and listening classes in the Foundations Program can design their syllabi in order to enhance vocabulary instruction and systematically teach the words in their classrooms. In addition, the teachers can give a more concrete vision of vocabulary learning to their students, which can be very motivational in that the students are given an essential, meaningful, and manageable number of words to study. Additionally, the students could easily study the words independently if the list were made into a small pocket-sized book.

Another useful application of this material is that the teachers at the ELC can use the specialized corpus generated from compiling the texts of the graded readers and listening books to teach these 500 essential words effectively. To teach them, the teachers can use an Online Concordancer, a web program that gives a list of several words, phrases, or distributed structures along with immediate contexts from a corpus or text collection. Using the program, from the texts that the students read or listen to, the teachers can generate sentences that contain words that they want to teach or they can generate vocabulary exercises. As a result, the students can efficiently encounter the words associated with contexts that they read from the extensive reading materials and reinforce their retention of the words.

Finally, this 500-word list can be used to help the students strengthen essential vocabulary knowledge in the reading and listening areas of the ELC because this list is generated from the texts of the textbooks that the ELC is currently using. In addition, this list can also meet other general English needs involving basic vocabulary because the list is generated based on the GSL and the AWL, which are the essential words for the global English needs.

Suggestions for Further Research

Based on the methods of this project development, some essential further research is possible. Since the material is based on only two surveys over two semesters, Fall 2009 and

Winter 2010, it is still not certain whether this wordlist is reliable for future use at the ELC. Since the students enrolling at the ELC are different every semester, this material should be tested and analyzed to check its reliability.

Another suggestion is, in order to examine the reliability of this material, the same students who took the survey could take a vocabulary test on the words of the survey. It is possible that the survey may have been misleading, depending on the students' honesty, physical condition, or psychological state. Therefore, having students take a test based on the same vocabulary items of the survey could be a good future research project to see the correlation between the survey and the test.

Furthermore, as a verification process, a researcher could conduct item analysis by analyzing the items on the vocabulary test that the students in the Foundations Program take for placement at the beginning of each semester, and by comparing the items in the survey with those in the test so as to find any correlations.

Conclusion

I have learned many valuable lessons from the process of developing this material. First of all, I have learned the value of the Frequency and Range software, as well as the VocabProfile program in analyzing text. Because of these tools, teachers can easily analyze words in the textbooks that their students are using. Thanks to these tools, I was able to analyze the texts of the graded readers and Listening Packets and learn in detail about the vocabulary profiles of the textbooks. I was able to find out that many graded readers being used at the ELC contained a serious number of off-list words that might affect students' comprehension negatively, which taught me that I should be more careful in selecting extensive reading materials for my students.

In addition, I learned the great value of the GSL and the AWL in learning essential vocabulary. It may not be necessary for an institution like the ELC to scan their textbooks and run them through a Frequency and Range analysis. If the primary purpose of an institution is to teach words from the two lists, focusing on graded readers use, administering a survey for students' vocabulary knowledge can be sufficient. Analyzing the texts of the graded readers may not be necessary because the entire point of those texts is to drill the words from the two lists. However, if an institution wishes to teach from non-graded readers in a specific discipline such as biology, then analyzing their textbooks and creating a specialized list of words would be beneficial. Students' money and time are valuable, and teaching general English words without a systematic approach can be a disservice to them.

Another important lesson I learned is to know students' needs and lexical knowledge. From the experience of the first survey, I learned that the students knew most of the words on the list that I generated, which I did not expect. Had I understood them better, I could have targeted the right words in the corpus that they needed to know. Above all, I have an understanding of basic principles of teaching vocabulary. Once we know the needs of our students, we should also be well acquainted with the texts they use. After we analyze the texts, we can create or choose an existing list of words they need to learn. Finally, integrating the words with other skills and strategies can reinforce their retention.

References

- Adolphs, S. & Schmitt, N. (2003). Lexical coverage of spoken discourse (CANCODE), *Applied Linguistics*, 24 (4), 425-438.
- Anderson, R. C. & Freebody, P. (1983). Reading comprehension and the assessment and acquisition of word knowledge. *Advances in Reading/Language Research*, 2, 231-256.
- Anderson, C., Stevens, C., Shifrin, Z., & Osborn, J. (1978). Instantiation of word meanings in children. *Journal of Reading Behavior*, 10, 149-157.
- Bengelei, S. & Paribahkt, T. (2004). L2 reading proficiency and lexical inferencing by university EFL learners. *Canadian Modern Language Review*, 61, 225-249.
- Campion, M. E. & Elley, W. B. (1971). *An Academic Vocabulary List*. Wellington: New Zealand Council for Educational Research.
- Carter, R. (1987). Vocabulary and second/foreign language teaching. *Language Teaching*, 20, 3-16.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- English Language Center. (2009). *Curriculum portfolio*. Retrieved March 2, 2010 from <http://elc.byu.edu/curriculum/#>.
- Folse, K. (2004). *Vocabulary myths: Applying second language research to classroom teaching*. Ann Arbor: The University of Michigan Press.
- Ghadessy, P. (1979). Frequency counts, words lists, and materials preparation: A new approach. *English Teaching Forum*, 17, 24-27.
- Goulden, R., Nation, P. & Read, J. (1990). How large can a receptive vocabulary be? *Applied Linguistics*, 11, 341-363.
- Hanston, S. (2002). *Corpora in applied linguistics*. Cambridge Applied Linguistics: University Press.

- Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). *Range Frequency Programs* [Computer software] Wellington, New Zealand: Victoria University of Wellington.
- Available at: http://www.vuw.ac.nz/lals/staff/paul_nation/index.html. Cambridge.
- Krashen, S. (1989). We acquire spelling and vocabulary by reading: Additional evidence for Input Hypothesis. *Modern Language Journal*, 73, 440-464.
- Laufer, B. (2003). Vocabulary acquisition in a second language learning: Do learner really acquire most vocabulary by reading? Some empirical evidence. *Modern Language Journal*, 59, 567-587.
- Lynn, R. W. (1973). Preparing word lists: a suggested method. *RELC Journal*, 4 (1), 25-32.
- Meara, P. & Buxton, B. (1987). An alternative to multiple choice vocabulary test. *Language Testing*, 4, 142-151.
- Nation, I. S. P. (1993). Measuring readiness for simplified material: A test of the first 1,000 words of English. In M. Tickoo (Ed.), *Simplification: Theory and application* (pp. 193-202). Singapore: RELC.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Praninskas, J. (1972). *American university word list*. London: Longman.
- Prince, P (1996). Second language vocabulary learning: The role of context versus translations as a function of proficiency. *The Modern Language Journal*, 80, 478-493.
- Read, John. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Schatz, E. & Baldwin, R. (1986). Context clues are unreliable predictors of word meanings. *Reading Research Quarterly*, 21, 439-453.
- Schonell, F. J., Meddleton, I. G., & Shaw, B. A. (1956). A study of the oral vocabulary of adults.

Brisbane: University of Queensland Press.

Sternberg, R. (1987). Most vocabulary is learned from context. In M. McKeown & M.

Curtis (Eds.), *The nature of vocabulary acquisition* (p. 89-106). Hilldale, NJ:

Lawrence Erlbaum.

Thompson, C. (2005). *500 essential English words for ESL* (Master's project).

Retrieved January 30, 2010, from <http://pdf.byu.edu>.

Wesche, M. & Paribakht, T. (1994, March). *Enhancing vocabulary acquisition through*

reading: A hierarchy of text-related exercise types. Paper presented at the

meeting of the American Association for Applied Linguistics (AAAL), Baltimore, MD

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

Co.

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

Communication. 3, 215-229.

Appendix A: Sample of words analyzed by Range and Frequency Analysis

1-2K &AWL

FAMILIES	RANGE	Family	FREQ	SING	6	307
THE	9	17352	SING	6	307	
BE	9	14128	ANIMAL	6	265	
I	9	9296	FIGHT	6	247	
AND	9	9030	OWN	6	231	
TO	9	8111	AFRAID	6	221	
A	9	7856	EARTH	6	187	
HE	9	7829	ARM	6	180	
IN	9	4866	CARRY	6	150	
OF	9	4807	LEAD	6	146	
YOU	9	4724	ALMOST	6	145	
THIS	9	3952	WITHOUT	6	137	
SHE	9	3838	FLY	6	130	
IT	9	3724	LAND	6	120	
THEY	9	3611	SUN	6	117	
NOT	9	3490	HIDE	6	115	
HAVE	9	3198	DESCRIBE	6	115	
SAY	9	3013	KITCHEN	6	110	
DO	9	2907	APPEAR	6	109	
FOR	9	2199	EGG	6	108	
ON	9	2159	LESS	6	104	
WE	9	1871	LAY	6	102	
GO	9	1852	COMPUTE	6	98	
BUT	9	1847	BIRD	6	97	
AT	9	1759	ALONE	6	96	
WITH	9	1641	ROAD	6	95	
THERE	9	1291	POWER	6	92	
SOME	9	1185	MARK	6	91	
ABOUT	9	1173	DIRECT	6	90	
WHAT	9	1137	RETURN	6	88	
THEN	9	1081	DEEP	6	88	
COME	9	1079	MISS	6	87	
CAN	9	1078	SKY	6	86	
AS	9	1077	HORSE	6	85	
WILL	9	1073	TOMORROW	6	80	
FROM	9	1070	DAUGHTER	6	79	
LOOK	9	1039	CHURCH	6	78	
ONE	9	1025	GOD	6	75	
			PROTECT	6	73	

1-2K &AWL	RANGE	FREQ		
RESTAURANT	5	20	APPARENT	1
BASE	5	20	APPROXIMATE	1
EXCEPT	5	19	COMMENCE	1
BOWL	5	19	CONFINE	1
TEMPERATURE	5	19	CONTRADICT	1
PREFER	5	19	CONTRIBUTE	1
ASSUME	5	19	DEDUCE	1
REMIND	5	19	DISTORT	1
THIRTY	5	18	ENABLE	1
WESTERN	5	18	ENCOUNTER	1
POLITE	5	18	ENFORCE	1
CASE	5	18	EQUATE	1
SYSTEM	5	18	ERODE	1
STOVE	5	18	EVOLVE	1
MATCH	5	18	FORMULA	1
DATE	5	18	INCLINE	1
SHARE	5	18	INTEGRATE	1
EAGER	5	18	INTERVENE	1
BELONG	5	18	INVEST	1
FLOOD	5	18	ISOLATE	1
CONSIDER	5	18	LOGIC	1
PROGRESS	5	18	MATURE	1
DESK	5	17	NOTION	1
WASTE	5	17	OBVIOUS	1
POLITICAL	5	17	PERSIST	1
SIGNAL	5	17	PROHIBIT	1
RELIEVE	5	17	REFINE	1
ORDINARY	5	16	REINFORCE	1
GRACE	5	16	RESTRICT	1
DEAL	5	16	SCHEME	1
POUND	5	16	SIMULATE	1
BICYCLE	5	15	SURVEY	1
CREDIT	5	15	SUSTAIN	1
CHARGE	5	15	TERMINATE	1
ENTIRE	5	15	TREND	1
DELIVER	5	15	ULTIMATE	1
DIVIDE	5	15	VEHICLE	1
SPORT	5	15	VIOLATE	1

Appendix B: Samples of High Frequency Words from the Off-list

TYPE	RANGE	FREQ	F1	F2	F3	F4	F5	F6	F7	F8	F9
YORK	8	37	2	8	2	13	4	6	1	0	1
MARY	7	93	9	51	23	1	0	5	2	0	2
OK	7	69	1	33	8	0	1	0	12	2	12
JOHN	7	62	14	7	12	16	0	11	1	0	1
FRANCE	7	53	7	11	7	21	3	3	0	1	0
BRITISH	7	41	0	2	1	22	10	4	1	0	1
MOVIES	7	15	0	4	1	5	1	0	1	2	1
GEORGE	7	12	1	1	0	2	1	5	1	0	1
ANNE	6	267	55	186	0	3	1	0	11	0	11
SCOTT	6	91	0	0	0	3	69	4	2	11	2
ENGLAND	6	89	41	18	8	17	4	1	0	0	0
LONDON	6	64	42	16	1	3	1	1	0	0	0
AMERICAN	6	56	0	8	7	21	4	14	0	2	0
CANADA	6	49	0	5	22	15	0	3	2	0	2
BRAZIL	6	41	7	4	1	3	21	5	0	0	0
FRENCH	6	39	2	1	6	11	2	17	0	0	0
BOOTS	6	24	0	4	13	1	2	2	0	2	0
ER	6	16	3	0	2	3	0	4	2	0	2
MOVIE	6	16	0	5	0	1	0	1	3	3	3
MEXICO	6	11	0	0	1	5	2	1	1	0	1
ITALIAN	6	10	0	1	2	4	1	0	1	0	1
ENGLISHMAN	6	9	1	1	3	1	2	1	0	0	0
PICNIC	6	8	0	1	0	1	0	1	2	1	2
TOM	5	237	83	0	128	21	1	4	0	0	0
HENRY	5	156	118	16	13	8	0	1	0	0	0
EDWARD	5	126	122	1	0	1	1	1	0	0	0
AMERICA	5	80	2	0	7	31	2	38	0	0	0
EMERGENCY	5	47	0	0	1	1	0	0	10	25	10
EUROPE	5	46	2	2	5	29	0	8	0	0	0
CRAZY	5	42	0	19	19	2	0	0	1	0	1
JEAN	5	42	0	0	22	5	0	1	7	0	7
WILLIAM	5	38	12	2	15	7	0	2	0	0	0
ST	5	37	3	0	9	18	5	2	0	0	0
CABIN	5	36	0	0	21	7	0	6	1	0	1
JOE	5	31	0	5	0	1	0	0	3	19	3
ITALY	5	30	1	0	15	11	2	1	0	0	0
DE	5	28	11	0	8	6	2	1	0	0	0

Appendix C: Function Words and Numbers

a about across after again ago almost along always and another any around as at away back be because before behind below beside between both but by can could do down each eight either even every first for four from front have he how however I if in instead into it just may maybe might mister more most Mrs. much must near never next nine no not of off oh on one or other out seem seven she sir six so some still than the then there they this though three through to too two under until up very we what when where which who why will with without would yes yet you inside lot

Appendix D: 675 Words Compiled for the First Survey

1-1,000 [502]

accept account act active add age agree air all allow alone already animal answer appear apply
 April arm arrive ask attack bad bank bear beauty become bed begin believe best big bill bird
 black blood blue board boat body book box boy break bridge bright bring brother build burn
 business buy call captain car care carry catch centre certain chance change child choose church
 city clear close coal coast cold college color come command company complete continue control
 cost count country course cover cross crowd cry cut danger dark daughter day dead decide deep
 describe destroy die difference difficult direct doctor dog dollar door draw dream dress drink
 drive drop dry early earth east easy eat egg end enemy English enjoy enough enter escape
 evening expense experience explain express eye face fact factory fall family famous far farm fast
 father feel few field fight fill find fine finish fire fish five floor flower fly follow food forest
 forget form free Friday friend full future game garden gas gather general gentle get gift girl give
 glass go god gold good great green ground group grow half hand hang happen happy hard head
 hear heart heavy help here high hill history hold home hope horse hot hour house hundred
 husband idea ill important inch include interest join judge June keep kill kind king know lady
 lake land language large last late laugh law lay lead learn leave left less letter level lie life light
 like line listen little live long look lord lose love low machine make man march mark marry
 master matter mean meet memory middle mile million mind minute moment money month
 morning mother motor mountain mouth move music name nation native need new news
 newspaper night north note now number offer office often oil old once only open operate order
 organize over own paper part party pass past pay people perhaps person picture piece place plan
 plant play please point poor popular possible power prepare present press pretty price problem
 protect pull put question quite race raise rather reach read ready real really reason receive red
 remain remember rest return rich ride right rise river road rock roll room round rule run safe
 same Saturday save say school sea seat second secret see sell send separate serious serve set
 settle several shake shall ship shoot short show side sign silence simple since sing sister sit size
 sky sleep small smile snow soft soldier son soon sound south space speak special speed spend
 spread spring stand star start state station stay step stone stop store story strange street strong
 student study succeed such summer sun sure surprise table take talk teach tear tell ten test thing
 think thousand throw time today together top touch town train travel tree true try turn twenty
 type understand unite university use usual village visit wait walk wall want war watch water way
 wear week well west whether white whole wife wild win wind window winter wish woman
 wonder wood word work world write wrong year young


1001-2,000 [160]

accident afraid afternoon ahead alive anger asleep attention baby bake beat birth bottom brave
 breakfast breathe brown bus busy cake camp card chair check chicken Christmas clean clever
 clock cloth club coat comfort conversation cook corner cousin cream cup damage dance dinner
 dirt during engine exact examining excite excuse fat firm foot forgive forward fright fun grand
 grass guess gun hair hall harbor hat hate hide hit hole hullo hunger hurt ice inform inn inside
 invent invite island journey jump key kitchen knock lean leg lot loud luck lunch manage meat
 medicine message mistake neck nice noise nose nurse ocean pack pain parent park passenger pen
 perfect pick police prison probable proud quick quiet rail rain repeat roof rush sad search sheet
 shirt shoe shop shout skin slave slow smell sorry steal stick storm straight sudden suit swim tall
 tea telephone tent terrible thank thick thin tie tire tomorrow tonight track trip uncle wake warm
 wash weather weigh wet worry worse yellow AWL [13] Chapter compute feature final goal
 identify job reside respond similar task tense transport





Appendix E: First Vocabulary Survey

Level: Foundation A-1

Directions:

In the following table, mark how well you know each word. Check  the box that best describes how well you know the word. Check only one box for each word, and do not use a dictionary. Be careful, there are some words in this list that **are not real English words.**

Example:

WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
child				
name				
test				
sound				

	WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
1	toy				
2	liberal				
3	program				
4	resource				
5	cattle				
6	copper				
7	opposite				
8	community				
9	bitter				
10	pearl				
11	arthret				
12	pet				
13	scent				
14	brief				
15	comment				
16	rude				

	WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
17	rogther				
18	question				
19	fast				
20	dress				
21	important				
22	island				
23	carry				
24	station				
25	poor				
26	worry				
27	lead				
28	uprouse				
29	leg				
30	decide				
31	fire				
32	lose				
33	idea				
34	simerabley				
35	point				
36	serve				
37	silence				
38	trip				
39	village				
40	spring				
41	march				
42	experience				
43	captain				
44	heart				
45	teager				
46	invite				
47	tear				
48	student				
49	task				
50	skin				

THING	4	4	4	4	4	4	24	4
MOTHR	4	4	4	4	4	4	24	4
RIGHT	4	4	4	4	4	4	24	4
NIGHT	4	4	4	4	4	4	24	4
DOG	4	4	4	4	4	4	24	4
PUT	4	4	4	4	4	4	24	4
BEGIN	4	4	8		4	4	24	4
HEAR	4	4	8		4	4	24	4
KING	4	4	8		4	4	24	4
ONLY	4	4	8		4	4	24	4
BIG	4	4	8		4	4	24	4
CHILD	4	4	8		4	4	24	4
WALK	4	4	8		4	4	24	4
MOVE	4	4	8		4	4	24	4
STOP	4	4	8		4	4	24	4
HAND	4	4	8		4	4	24	4
BEST	4	4	8		4	4	24	4
NAME	4	4	8		4	4	24	4
PLACE	4	4	8		4	4	24	4
FATHER	4	4	8		4	4	24	4
HAPPY	4	4	8		4	4	24	4
BOY	4	4	4		4	4	20	4
RIVER	4	4	4		4	4	20	4
SIT	4	4	4		4	4	20	4
WRITE	4	4	4		4	4	20	4
WAY	4	4	4		4	4	20	4
CRY	4	4	4		4	4	20	4
LITTLE	4	4	4		4	4	20	4
USE	4	4	4		4	4	20	4
ROOM	4	4	4		4	4	20	4
HEAD	4	4	4		4	4	20	4
STAND	4	4	4		4	4	20	4
DOOR	4	4	4		4	4	20	4
FEEL	4	4	4		4	4	20	4
WATER	4	4	4		4	4	20	4
FACE	4	4	4		4	4	20	4

Appendix G: 675 Words compiled for the Second Survey

1001-2,000 [383] absent absolutely accuse ache afford agriculture aim airplane alike aloud altogether amuse angle annoy anxiety apology applause approve arrow artificial ash ashamed astonish audience autumn avenue awkward baggage balance band barber barely barrel beak bean beast behave behavior bell belt berry bind bitter blade bless blind boast borrow bound brass brick broadcast burial calculate camera canal cap cattle caution centimeter ceremony charm cheat chimney civilize clay clerk cliff coarse collar collect comb commerce companion competition complicate compose confess confidence congratulate conscience convenience copper copy cork cottage cough coward crack crime criminal critic crush cure cushion custom deaf debt decay deceive deer defend delay delicate descend deserve devil diamond dictionary dip disappoint discipline discuss dismiss disturb ditch donkey dot drawer duck dull elastic elder enclose engineer envelope envy essential evil excellent extraordinary extreme false fan fancy fashion fasten fate feast fierce flame flash flavor flesh flour fond forbid fork frame fur garage gay generous glory gradual grateful grave greed guard guest guilty hammer harvest haste heal heap hesitate hollow hook horizon host hut imitate immense informal ink inquire instrument insult interfere international jaw jealous jewel juice kilogram kneel knot ladder leaf leather lend limb litter loaf loan lodging lump lung mat meantime mechanic mend merchant mercy merry messenger miserable modest monkey multiply murder mystery neglect nephew net niece nonsense noon nuisance obey offend oppose opposite organ outline pad patriotic paw pearl pencil penny permanent persuade pet pin pinch pint pipe pity plaster plenty plural poem poison polish practical praise preach precious preserve pride priest procession program prompt pump punctual punish pupil pure purple qualify quart rabbit rare raw razor recommend refer refresh rejoice reputation resign retire revenge ribbon rice rid ripe rival roar roast rod rot rubbish rude rug rust sacrifice sake salary sample saucer saws scale scatter scent scissors scold scorn scrape screw seed seize seldom self severe shallow shave shield sincere slope sock solemn solid solve sore soup sour sow spare spill spit splendid spoon stain stamp steep sting stocking strap strict stripe suspicion swear sweat swell sympathy tame tap taxi temper tempt tend tender thorough thumb tide tidy tip title tobacco toe tough towel toy translate tray treasure tribe tube tune typical umbrella universe upright upwards urge vain verse voyage waist wax weed wheat wine witness worm worship wreck wrist zero


AWL [292] abandon academy access accompany accurate acknowledge acquire adapt adequate adjacent adjust alter alternative analyze annual apparent appreciate appropriate approximate assemble assist attitude automate aware behalf bond brief bulk capable capacity challenge chart chemical circumstance clarify classic clause code collapse commence comment commission commit community compensate conceive concentrate conclude concurrent conduct confer confine conflict consent consist constant constitute consult consume contact context contract contradict contrary contrast contribute convert core corporate crucial currency cycle data debate decade deduce define definite demonstrate depress design despite detect device devote dimension discriminate display distinct distort distribute diverse document domestic dominate draft drama edit element emerge enable encounter enforce equate equivalent erode error establish

estate estimate ethic evident evolve expand expert explicit exploit extract facilitate file flexible
focus format formula foundation founded framework function fund furthermore gender generate
generation guarantee identical ignorant illustrate immigrate impact imply incidence incline
income incorporate index individual infrastructure initial insert insight inspect integrate integrity
internal intervene invest investigate isolate journal label layer lecture legislate liberal license
likewise link logic manipulate manual mature medium mental migrate minimum ministry minor
monitor network nevertheless normal notion obtain obvious odd ongoing option orient overall
overlap overseas panel paragraph participate partner perceive percent persist phase phenomenon
philosophy plus policy portion pose positive potential predict predominant previous primary
prime principal principle priority proceed professional prohibit project psychology purchase
pursue random range react recover refine regime region register reinforce reject reluctance
research resolve resource restore restrain restrict reveal reverse rigid role scheme select series
simulate so-called sole specific sphere stable status strategy structure style submit sufficient sum
summary survey sustain tape target technical technique terminate text theme trace tradition
transform transmit trend trigger ultimate undergo uniform unique valid vary vehicle version via
violate virtual vision visual volume voluntary whereas widespread





Appendix H: Second Vocabulary Survey

Level: Foundation A-4

Directions:

In the following table, mark how well you know each word. Check  the box that best describes how well you know the word. Check only one box for each word, and do not use a dictionary. Be careful, there are some words in this list that **are not real English words.**

Example:

WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
child				
name				
test				
sound				

	WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
1	aware				
2	label				
3	blind				
4	penny				
5	crime				
6	glomater				
7	project				
8	qualify				
9	guest				
10	constant				
11	assemble				
12	diverse				
13	aloud				
14	cough				
15	false				
16	flour				

	WORD	I know this word	I think I know this word	I am I am not sure if I know this word	I do not know this word
17	wontlery				
18	rabbit				
19	rice				
20	sake				
21	seldom				
22	steep				
23	sweat				
24	acknowledge				
25	core				
26	despite				
27	index				
28	trovbelo				
29	ongoing				
30	series				
31	vision				
32	volume				
33	clay				
34	dolebry				
35	legislate				
36	orient				
37	participate				
38	technique				
39	trace				
40	transform				
41	applause				
42	artificial				
43	bound				
44	coarse				
45	schulton				
46	conscience				
47	ripe				
48	salary				
49	shallow				
50	sore				

Appendix I: Final list of 500 Essential Words for the Foundations Program

1001-2000 [294] accuse ache afford aim alike aloud altogether amuse angle annoy anxiety apology applause arrow artificial ash ashamed astonish audience autumn awkward baggage balance band barber barely barrel beak bean beast belt berry bind bitter blade blind boast borrow bound brass brick burial calculate cattle caution centimeter charm cheat chimney civilize clay clerk cliff coarse collar collect comb commerce compose confess congratulate conscience copper cork cottage cough coward crack criminal critic cushion deaf debt decay deceive deer defend delay delicate descend deserve diamond dip discipline discuss dismiss disturb ditch dot drawer dull elastic elder enclose envelope envy essential extraordinary false fan fancy fasten fate feast fierce flame flash flavor flesh forbid fork frame fur grave greed guard guilty hammer harvest haste heal heap hesitate hollow hook horizon host hut immense ink inquire instrument insult interfere jaw jealous jewel kilogram kneel knot ladder leaf leather limb liter loaf loan lodging lump lung mat meantime mend merchant mercy miserable modest murder mystery neglect net niece nonsense nuisance obey offend oppose organ outline pad patriotic paw pearl penny permanent persuade pin pinch pint pipe pity plaster plenty plural poem poison polish preach precious priest procession prompt pump punctual punish pupil pure qualify quart rare raw razor rejoice reputation resign retire revenge ribbon rid ripe rival roar roast rod rot rubbish rug rust sacrifice sake saucer saws scatter scent scissors scold scorn scrape screw seize seldom severe shallow shave shield sincere slope sock solemn solid solve sore sour sow spare spill spit splendid spoon stain steep sting stocking strap strict stripe suspicion swell sympathy tame tap temper tempt tend tender thorough thumb tide tidy tip tobacco toe tough towel toy tray treasure tribe tune upright upwards urge vain verse voyage waist wax weed wheat witness worm worship wreck wrist

AWL [206] abandon access accompany accurate acknowledge acquire adequate adjacent adjust alter alternative analyze annual apparent appreciate approximate assemble assist attitude automate aware behalf brief bulk capable capacity chart circumstance clarify clause code collapse commence commission commit compensate conclude concurrent confer confine conflict consent consist constitute consult consume contradict contrary contrast core corporate crucial currency cycle data decade deduce define definite demonstrate depress despite detect discriminate display distort distribute diverse document domestic dominate element emerge enable encounter enforce equate equivalent erode error estate estimate ethic evolve expand explicit exploit extract facilitate file flexible format founded framework function fund furthermore gender generation guarantee identical ignorant illustrate imply incidence incline income incorporate infrastructure initial insert insight integrate integrity intervene invest investigate isolate label lecture legislate liberal likewise link manipulate medium migrate ministry minor network nevertheless obvious odd ongoing overall overlap overseas panel participate partner perceive persist phase phenomenon philosophy portion pose positive potential predict predominant previous prime principal principle priority prohibit purchase pursue range react refine regime region reinforce reject reluctance research resolve resource restore restrain restrict reveal reverse rigid scheme simulate so-called sole sphere stable strategy submit sum summary survey sustain target technique terminate theme trace trend trigger ultimate undergo uniform valid vary vehicle via violate virtual whereas widespread