The Giver: A Corpus-Based Analysis of Word Frequencies

TARA BRANDENBURG-WEEKS
Colorado State University, Fort Collins,
United State of America
brandenburg.tara@gmail.com

ALBATOOL MOHAMMED ABALKHEEL
Department of English Language and Translation,
College of Arabic Language and Social Studies,
Qassim University, Buraydah, Saudi Arabia

ABSTRACT

Choosing appropriate reading material for language learners is not an easy task. It is necessary to keep in mind the specific purpose of the exercise and that the material chosen must support that purpose. Often, these goals will include developing fluency, increasing vocabulary knowledge, and increasing the rate at which vocabulary grows. One of the texts and novels commonly used for extensive reading in high intermediate level English as a Second Language coursework in the United States is The Giver, a young adult novel written by American author Lois Lowry. Due to its importance, we chose to analyze the novel within the framework of word frequency counts through the lens of the percent of words a learner must know in order to understand the particular text. A corpus-based method is utilized in the study to determine the types and frequency of vocabulary that The Giver provides. Common groupings of related low frequency lexical items and collocations are explored. This application of a corpus-based analysis of the text draws a picture of the vocabulary that is presented in The Giver, and how it matches 1K, 2K and AWL. The analysis reveals that 89.07% of the vocabulary was found to fall within these high frequency word lists, which gives the novel a somewhat higher comprehensibility than average academic texts. Analysis of low-frequency words reveals key vocabulary that increase comprehensibility when explicitly taught, as well as use of compound words, prefixes, and suffixes, that—if paired with instruction in vocabulary analysis techniques—can also increase the comprehensibility of the text. A conclusion is drawn regarding the appropriateness of this novel for the target audience in the specified course types, with suggestions for modifications and supplements for the text's use within ESL/EFL classrooms.

Keywords: word frequency; reading comprehension; vocabulary knowledge and size; corpus-based method linguistics

INTRODUCTION

It is estimated that native speakers of English learn approximately 1,000 word families per year until leveling out at approximately 20,000 word families (Nation & Waring, 1997). However, only 2,000 words (2K word list) make up roughly 80% of English texts (Cobb, 1998; Nation, 2001). While many students may be driven to understand every single word in a text, the most efficient method of learning vocabulary is to focus on the most frequent word families. General vocabulary knowledge increases rapidly within the 2K word range; however, the rate of growth does not continue as rapidly after that. In fact, it has been found that an additional 3,000 words only increases text coverage by approximately 8% (Cobb, n.d.). On the other hand, Schmitt et al. (2011) found that for each 1% increase in vocabulary knowledge, there is an average of 2.3% growth in comprehension. "Vocabulary size is thus a strong predictor of reading comprehension" (Lee & Wong, 2020, p. 16).

Reading accuracy is also affected positively by an increase in vocabulary size. Gao (2015) found that vocabulary size could give an indication of the accuracy rate in reading comprehension. Vocabulary knowledge is crucial for the development of reading (Cameron,

2001) and it has been found that there are also substantial positive correlations between vocabulary knowledge and reading comprehension (Wei, 2020). The importance of vocabulary growth cannot be underestimated, although Nation and Waring (1997) noted that while vocabulary knowledge alone is not sufficient for proficiency in reading, it does play an integral role in language development and comprehension. It is also a prerequisite for fluency and volubility in reading and writing (Hsueh-Chao & Nation, 2000; Milton, 2013).

While knowing the 2K word list can help students understand a great deal of a text, research suggests that understanding only 80% of a text is not enough. With comprehension of only the 2,000 word list, students will still find it difficult to comprehend a passage. This is because approximately two words in every line of text will be unknown. Instead, it has been suggested that 95% coverage of vocabulary in a text is necessary for sufficient comprehension, which could be provided for through an understanding of 3,000 to 5,000 word families (Nation & Waring, 1997).

While it is not possible for English language instructors to explicitly teach all vocabulary that students may encounter in a given text, carefully selecting key vocabulary words that are used frequently in a particular text can help students to increase comprehension. The present study, conducted by an ESL instructor as classroom-based-research in an Intensive English Program (IEP) reading class, analyzes the vocabulary usage in the novel, *The Giver* (Lowry, 1993) with two objectives:

1. Analyze the comprehensibility of *The Giver* for language learners who have already studied the 1K, 2K, and AWL vocabulary lists in order to determine its appropriateness as an IEP course text.
2. Utilize corpus based analysis of the text to prepare vocabulary word lists that instructors can use to help students increase their overall comprehension of the text in the most efficient way possible.

The literature review section outlines existing corpus studies that establish the high frequency word lists used as a baseline against which to evaluate the appropriateness of *The Giver* as an instructional text. It also establishes the need for corpus-based vocabulary instruction to increase comprehension. The methods section outlines how a stratified sample of *The Giver* was analyzed with *Compleat Lexical Tutor* to produce word frequency lists. The results section of the study presents recommendations for publishers to incorporate vocabulary instructional support into *The Giver* as well as provide pedagogical recommendations for instructors using this novel in their classes.

LITERATURE REVIEW

Schmitt et al. (2011) asserted that, in fact, a learner must know 98% to 99% of the vocabulary in a text to be able to independently comprehend authentic texts in English. Laufer and Ravenhorst-Kalovski (2010) also found that language learners have to know at least 4,000 to 5,000 words for reading comprehension of academic texts, and that the optimal achievement for reading academic texts is 6,000 to 8,000 words.

Aside from the 2,000 most frequent words, it has been found that approximately 9% of running words in texts are academic, 5% are technical, and 5% are low frequency (Nation, 2001, p. 12). However, the composition of text in an academic corpus is slightly different than that of a general corpus. In fact, only 76.1% of running words in an academic text can be found on the 1,000 and 2,000 word lists, while 10% are found on the Academic Word List (AWL), and 13.9% are low frequency words (Coxhead, 1998 as cited in Nation, 2001, p. 13). The AWL

itself "covered 10% of the total words in a general academic text, but only covered 1.4% of the total words in a fiction with the same vocabulary size" (Yan & Yang, 2016, p. 45). Therefore, a learner who knows the 1K, 2K, and AWL lists, will be able to comprehend approximately 86.1% of the words in any given academic text. Hence, to reach the 95% coverage goal, understanding of additional low frequency words will be necessary for learners to attain sufficient command of a text.

While some researchers recommend explicit instruction of high frequency words (1K, 2K, and AWL), they also indicate that the development of vocabulary strategies, such as using context and affixes to guess the meaning of words, as well as using dictionaries and vocabulary cards is crucial to reading comprehension (Nation, 2001, p. 20; Nation & Waring, 1997). This is because low frequency word selection of a text greatly varies depending upon the genre. However, teaching strategies must still recognize the need for sufficient vocabulary size, as was supported by Nation and Waring (1997), who argued that a learner must know approximately 3,000 words before they will be able to learn the meaning of a word from context.

Schmitt et al. (2011) agreed with Nation and Waring's (1997) assertion, and argued that skills such as inference, fluency, reading strategies, motivation, word recognition, and the like, affect the extent to which a student will comprehend a text. They added that background knowledge on the topic of the assigned work also facilitates comprehension; however, the more comprehension of a text expected, the more vocabulary must be known. For example, teachers who aim for students to comprehend 60% or more of a text should expect students to comprehend about 98% of the words (Schmitt et al., 2011, p. 39). Hence, when determining the level of vocabulary coverage necessary for a particular text, one must also take into account the comprehension goals intended by the teacher as well as the reading strategies employed by the reader.

Frequency of vocabulary, the number of instances where the specific words are encountered in the text, affects learner's word knowledge (Tyler, 2012), but it is difficult to say "how often a word should recur in a textbook to enable learning," and "to establish how many words a learner of a foreign language should know to achieve adequate comprehension of a text of average difficulty" (Norberg & Nordlund, 2018, p. 464). Therefore, it is clear that when choosing reading texts and novels for language learners, it is crucial to keep reading goals in mind. Corpus linguistic methods of analysis enable teachers, educators, and publishers to examine texts and their appropriateness for the established goals, especially vocabulary that are most suitable for this kind of analysis (Jones & Durrant, 2010). This is added to the advantages of using "corpora as supplementary learning material to reinforce vocabulary knowledge" (Laosrirattanachai & Ruangjaroon, 2021, p. 2). In corpus-based analysis, teachers, educators, and publishers select and order a body of language "according to specific linguistic criteria in order to be used as a representative sample of the target language" and, with the help of computer analysis techniques, complement introspection can be done by "providing reliable quantitative data, and a solid and much broader empirical basis for hypotheses" (Millar & Budgell, 2008, p. 370).

Although researchers and practitioners of textbook composition and evaluation vary in their perspectives, including on how to make choices for reading assignments, they all "point to the textbook as something that should be constructed with due thought and care and critically examined" (Stockton, 2020, p. 5). It is important to keep this in mind when analyzing and examining these materials, due to the fact that "they play a vital role, since they are considered as the means of conveying the required materials and knowledge" (Sarem et al., 2013, p. 372). In the EFL context, the need for this analysis is a must because the "exposure is limited to the input provided by teachers and teaching materials" (Alcaraz-Marmol, 2015, p. 6). Implementing corpus linguistics for this task is of great importance. Several researchers in

corpus linguistics (e.g., Laufer & Ravenhorst-Kalovski, 2010; van Zeeland & Schmitt, 2013; Webb & Nation, 2008) have examined vocabulary threshold levels and the extent of lexical knowledge EFL/ESL learners need for comprehension (Laufer & Ravenhorst-Kalovski, 2010; Nation & Waring, 1997; van Zeeland & Schmitt, 2013; Webb & Nation, 2008). Such research has been found to be a reliable source of valuable information for teachers, learners, and publishers (e.g., Granger, 2002; Hoey, 2000; Millar, 2016; Molavi et al., 2014; Norberg & Nordlund, 2018; Schmitt & Schmitt, 2000; Zarifi & Mukundan, 2013).

METHODOLOGY

CORPUS DATA

The present study analyzed word frequencies in *The Giver,* which provides a useful and meaningful sample for this project because it is one of the required texts for a high-intermediate "Reading and Vocabulary" course in an intensive English program (IEP) at a large Midwestern university in the United States. The rationale of curriculum developers in selecting this particular novel for the course is related to the sequencing. In the preceding IEP course levels, students use abridged novels from the Penguin Reader Series, which are graded by level of difficulty. Throughout these levels, students advance through graded-reader levels as well as learn reading skills such as skimming, scanning, pronoun reference, and guessing the meaning of words based on the context. The "Reading and Vocabulary" course is intended to provide a bridge for students between the graded readers of the low-intermediate class and the academic articles presented in the low-advanced reading class, scaffolding the student toward independent reading for credit-bearing university classes, in which they will enroll after only an additional 3 months of IEP instruction.

Although *The Giver* has faced skepticism for use in ESL courses because it uses some words eccentrically and in unusual senses—such as *release*, which in the novel means euthanasia—proponents of the use of the text argue that it provides a cultural basis upon which students can engage with Americans, as a majority of college-age Americans will have read this book in high school. Secondly, although it is considered a young adult novel, the deep philosophical and political themes provide opportunities for discussion, the introduction of related non-fiction readings, and critical reflection on societal practices. To analyze the word usage of the novel to determine vocabulary supplements and text modifications necessary for using this novel as an ESL/EFL text, a corpus was created using Chapters 1, 3, 15, and 21. Hence, samples selected from the beginning, middle, and end approximate a representative sampling of the novel. The use of the small data sample (4,157 tokens) focused on just one novel was based on Zahar et al.'s (2001) study, which investigated the relationship between frequency of words, types of context, and learning. Implementing a corpus-based methodology to analyze a single-story in a vocabulary acquisition study with French-Canadian school-aged ESL learners, they found that "frequency needs were found to be related to learner level, and contextual richness was unrelated to learning" (Zahar et al., 2001, p. 541). Horst (2000), who also replicated the experiment in two separate studies with longer texts written for native speakers, found similar results.

ANALYSIS OF FREQUENCIES

The selected portions of *The Giver* were uploaded to the *VocabProfilers* tool that is freely available on fourth version of the *Compleat Lexical Tutor* website developed by Cobb (2004), using frequency lists based on Laufer and Nation's Lexical Profiler (LFP). Of the total 4,157 words (tokens) inserted in the corpus, there were only 1,247 different words (types). Of those

types, only 86.77% fell into either the 1K or 2K lists used in the *VocabProfilers* tool, as demonstrated in Table 1. This means that for individuals who know the 2,000 most frequent words of the English language, which is the intended level of the course, 13.23% of the words will be unknown. In other words, one or two words could be unknown in every 10 words of text.

An important factor to take into account when analyzing this data is that all character names (i.e., *Jonas*, *Lily*, *Gabriel*, *Gay*, *Asher*, *Gabe*, *Ash*), which accounted for 199 tokens and eight types, as well as a number of place names, such as *Jonas's dwelling* and *the Annex*, were included. Given that a large portion of the off-list words includes proper names, this may make it difficult to use the unknown word list as a means of judging text difficulty.

TABLE 1. Word frequency counts of *The Giver*

	Families	Types	Tokens	Percentage
K1 Words	470	686	3622	80.19%
K2 Words	173	197	297	6.58%
AWL Words	55	67	104	2.30%
Off-list Words		301	494	10.94%

After excluding proper nouns and hyphenated high frequency words, it was found that true low frequency tokens actually only comprise 282 types and 55 word families. Some of the word families include, for example, *birthing, birthmother,* and *birthmothers.* As *birth* and *mother* are both on the K2 list and *–ing* is a very frequent syntactic suffix, it may be seen that these "low frequency" words are actually composed of high frequency pieces and therefore, students trained to analyze the pieces of words could decode the meaning of such words.

Another consideration when interpreting the frequency data, which the LexTutor Concordancer does not calculate, is the use of prefixes to form more specific meanings, resulting in low frequency words. For example, *undone, unstrapped, unstrapping,* and *unwrapped* without the prefix *un-* would each appear on the K2 word list. As learning prefixes, including *un-*, is a part of the course curriculum, it may also be misleading to assume that students would not understand these words solely because they are categorized in the corpus as low frequency.

A final note regarding the interpretation of word frequencies is the number of instances of particular low frequency words. Among those with numerous tokens are *newchild* (15), *dwelling* (11), *announcement* (7), and *bike* (5). These words are distributed throughout the book, as can be noted when examining the page numbers of the concordance lines, because they refer to key concepts. For instance, rather than using the term *baby*, the term *newchild,* demonstrates that the concept of childbirth and childrearing is incredibly different in the novel than in society. The term *newchild* substitutes *baby*, which has very few occurrences throughout the novel. Beyond the fact that readers learn the word early on in the book and see it repeatedly, the meaning should be fairly transparent as both *new* and *child* fall within the 1K word list. The other common low frequency terms refer to everyday items in the life of characters in *The Giver* that may not necessarily be a part of ordinary society, meaning they are very frequent words in the novel, but not frequent words in everyday English. Although these words are categorized as low frequency in corpus analysis of the novel, these considerations must be taken into account when using the data to come to conclusions about readability.

RESULTS AND DISCUSSION

PATTERNS OF LOW FREQUENCY LEXICAL ITEMS

Although this novel uses quite a few low frequency words, most of them are conceptually related and occur in groups. The main lexical relationships among low frequency words are explored in this section. Because vocabulary learning is enriched by presenting vocabulary within a rich, comprehensible context, it is important to draw out vocabulary themes within the low frequency words in order for the course instructor to activate schemata that will help learners understand and remember vocabulary. In this section, the researcher first identified the most common low-frequency words, then manually reviewed and tagged the low-frequency word list to identify related clusters of vocabulary.

One of the most common themes of low frequency words relates to infants. The word *birthmother* occurs four times, and the words *newchild/newchildren* occur 14 times within the selected passages. Other words related to infancy are *babbling* (1), *crib* (1), *nurturer* (1), *nurturing* (2), and *toddler* (2). Although there are references to babies throughout the book, they tend to cluster around scenes in which the baby, Gabriel, appears as a main character, such as in the following selection: "Gabriel played happily on the floor, BABBLING his baby talk, looking with glee now" (Lowry, 1993, p. 164).

On top of ordinary infancy-related words occurring close together in sections regarding Gabriel, many unusual low frequency words arise in these sections as well. As the community in which the novel is set has very unusual rules and conventions regarding childhood, some interesting terms are used. Rather than being taken care of by the biological parents, babies are taken care of by *nurturers*. The following passage partially describes the social function of *birthmothers* and *nurturers* in childrearing within the community:

> Father turned the NEWCHILD onto his tummy in the basket. He sat beside it and rubbed its small back with a rhythmic motion. "Anyway, Lily-billy," he said affectionately, "the BIRTHMOTHERS never even get to see NEWCHILDREN. If you enjoy the little ones so much, you should hope for an Assignment as NURTURER. (Lowry, 1993, p. 21)

In this passage, *Father* (who is in fact an adoptive father society deemed more worthy of raising children than the true birthmother) is himself a *nurturer* by profession. As this concept is fairly different from the normal sense of childrearing, the terminology used by the author to refer to family members is unusual, but comes at a very high frequency within the novel as childhood and the treatment of babies is a recurrent theme throughout the work. However, the author makes a point to explain exactly what these specialized meanings of the words mean within the context of the novel, as with the description of *birthmother* in the preceding passage.

Another grouping of low frequency words includes institutional punishment words, which in the following selected passages include: *chastise* (1)*, *chastisement* (1)*, *condemned* (1)*, *conviction* (1)*, *fugitives* (1)*, and *transgression* (1). In the authoritarian society of the novel, rules permeate every aspect of life of the citizens and punishment is taken very seriously. Punishment occurs frequently and in everyday situations in the novel, as demonstrated in the following passage:

- Lily, he decided, would have to learn that soon, or she would be called in for CHASTISEMENT because of her insensitive chatter. (Lowry, p. 19)

Hence, the words necessary to describe various forms of punishment also arise frequently. The author chooses vocabulary words that make the situations seem severe, such as *condemned.* For example, as the preparations for running away are described, Lowry

incorporates the consequences of what—to many readers—might seem to be minor problems, but in *The Giver* are worthy of punishment by death, as in the following passage:

- He thought of the rules he had broken so far: enough that if he were caught, now, he would be CONDEMNED. (Lowry, p. 165)

In fact, a majority of the punishment-related low frequency words can be found in this chapter as descriptions of the potential consequences of the main character running away.

A third grouping of low frequency words pertains to the use of bicycles. As bicycles are the primary form of transportation in the community, they are mentioned frequently in the novel, although they are not a central focus. Of the words contained in the sample, the following low frequency words are used: *bike* (5)*, bikeport* (1)*, bikes* (1)*, pedal* (1)*, pedaled* (1)*, pedaling* (1)*,* and *port* (2). Although the common meaning of *port* is not exclusive to bicycles in everyday English, the eccentric use in the novel specifically refers to a *bikeport*—or in other words, what most English speakers would refer to as a bike rack, as in the following examples:

- Father put his bike into its **port** (Lowry, p. 20)
- as he wheeled his bike into its narrow **port** beside the door (Lowry, p. 4)
- He had hesitated for a moment, standing beside the **bikeport** in the darkness (Lowry, p. 165)

As the only use of *port* in the corpus refers to a *bikeport* in particular, we can see that this low frequency word is not only used regularly and distributed throughout all sections of the text, it is also used with a very low frequency meaning. The sense used in the novel does not match any of the senses described in *Merriam-Webster's Online Dictionary*, with the first two definitions being "a place where ships may ride secure from storms" and "a harbor town or city where ships may take on or discharge cargo" (port, n.d.). As one can see, ESL students who come across the term *port* in the novel and choose to look it up in a dictionary, will find an incongruous definition that may muddle the intended meaning of the author.

PATTERNS OF COLLOCATIONS

In order to fully consider the effect of collocations in this novel, five common low frequency words that occur throughout the selected portions of the novel were examined by analyzing patterns in concordance lines. The common collocates found in *The Giver* corpus were then compared to the most commonly found collocates in the *Corpus of Contemporary American English* (*COCA*; Davies, 2008) utilizing *Compleat Lexical Tutor* to determine whether the usage in the novel is representative of ordinary English patterns. The Collocates tool of AntConc (vers. 3.5.9; Anthony, 2020) was used to determine MI scores.

1. Dwelling. Within the eight hits for *dwelling* and two for *dwellings*, the following concordance lines revealed a few collocational patterns:

   ➢ he had left the DWELLING at night (Lowry 1993, p. 165)
   ➢ Before he had left the DWELLING, he had laid (Lowry, p. 166)
   ➢ kind they had played back in the DWELLING, with tickles (Lowry, p. 167)
   ➢ unexpected night away from the DWELLING. (Lowry, p. 164)
   ➢ side on the path behind his family's DWELLING. (Lowry, p. 2)

We can see that the verb *left* (MI 7.97823) collocates with *dwelling* as well as prepositions, including *back* (MI 4.80831)*, away* (MI 6.39327)*,* and *behind.* However, none of these were found to be common collocates according to *COCA* (MI 6.39327).

2. Newchild. There are 12 hits of *newchild* within the selected portions of the novel. Although the words *new* and *child* both occur as low frequency words apart from each other, the compound word *newchild* is very uncommon in most English. One collocation that appeared frequently (5 occurrences), but not in every instance, is *prep + the basket* (MI 7.14816) as demonstrated in the following concordance lines:

   ➢ Now, seeing the NEWCHILD and its expression (Lowry, p. 21)
   ➢ placed beside the NEWCHILD in his basket. (Lowry, p. 21)
   ➢ hovered over the NEWCHILD in its basket, (Lowry, p. 25)
   ➢ Father turned the NEWCHILD onto his tummy in the basket. (Lowry, p. 22)
   ➢ watching the NEWCHILD play. He was exhausted. (Lowry, p. 167)

   The phrase *in its/his/the basket* occurred three times in this sampling. Although it may not be a common collocate in English at large, it occurs in the novel. As the word *newchild* was not found in the *COCA*, it was not possible to draw comparisons to common collocates. For that reason, this vocabulary item and its collocates are not useful items for language learners to study beyond the purpose of this novel.

3. Announcement. *Announcement*, which occurs six times in the corpus, appears to commonly occur with nothing other than *the* (MI 1.86276) in the sample, as one can see in the following examples:

   ➢ remembered with humiliation, that the ANNOUNCEMENT "Attention. This is a (Lowry, 1993, p. 23)
   ➢ because. the public ANNOUNCEMENT had been sufficient to produce (Lowry, p. 23)
   ➢ There would be an ANNOUNCEMENT like that quite soon, he felt (Lowry, p. 23)
   ➢ bewildered by it. Not by the ANNOUNCEMENT or the necessary apology; those were (Lowry, p. 23)
   ➢ over the speaker, the ANNOUNCEMENT that had singled him out without using his (Lowry, p. 25)
   ➢ with the microphone all day, making ANNOUNCEMENTS. He laughed silently (Lowry, p. 22)

   However, if a sampling of the entire book were analyzed in the corpus, it is possible that collocational patterns might arise.

4. Recreation. There are four instances of the word *recreation* included in the corpus. The only collocate appears to be *recreation + noun,* including *area* (MI 9.56320), *period* (MI 10.14816), and *director* (MI 10.14816).

   ➢ Objects are not to be removed from the RECREATION area and that snacks are (Lowry, p. 23)
   ➢ taken the apple home, against the RECREATION area rules. That (Lowry, p. 25)

➤ disposed of the apple and made his apology to the RECREATION Director the next (Lowry, p. 23)
➤ It had happened during the RECREATION period, when he had been playing (Lowry, p. 23)

In all of the examples within *The Giver* corpus, it appears that the word *recreation* is used in noun phrases to modify head nouns. This falls into alignment with the results of *COCA*, which also demonstrates a high level of collocation with nouns. Where *recreation area* (MI 5.29) and *recreation director* (MI 3.15) are demonstrated to be frequently used collocations in *COCA*, *recreation period* was not found to be a common collocate on *COCA*.

5. Bike. The low frequency word, *bike,* which appears five times in the corpus, seems to collocate with prepositions, as can be seen in these examples:

➤ still, as he wheeled his BIKE into its narrow port beside the door (Lowry, p. 4)
➤ Father put his BIKE into its port. Then he picked up the basket (Lowry, 1993, p. 20)
➤ had dropped his BIKE on its side on the path (Lowry, p. 2)
➤ comfortably ride the larger BIKE when he was so accustomed (Lowry, p. 166)
➤ lifted him from the BIKE, and watched him investigate (Lowry, p. 167)

In three of the instances, *bike* is directly followed by either *into* (MI 7.36680) or *on* (MI 5.12579), whereas in one instance the preposition *from* (MI 5.91934) occurs two words to the left. However, once again, none of these were found within the top 50 collocates on *COCA*.

This analysis reveals that not only are many low frequency words used in eccentric sense, they also tend to be paired with low frequency or eccentric collocates, which presents a particular difficulty for language learners to comprehend and to acquire generalizable collocations while utilizing *The Giver* as a language learning text.

MODIFICATION

If this novel is a mandatory part of a curriculum, it may not be possible to modify the text itself; one must consider providing supplemental activities instead. On the other hand, it would be an interesting venture to propose an ESL edition of the novel to publishing companies. This section provides suggestions for both revisions for publishers as well as supplements teachers could use to scaffold the novel.

To begin with, textbook publishers could provide a glossary of terms with their definitions and collocates in the back of the book. This would give readers the authentic experience of reading a novel in English, with the ease of being able to quickly look up unknown words when necessary. For example, the word *newchild*, would be a particularly useful word to provide in a glossary. Perhaps the glossary could provide explanations for terms used in unusual senses in the novel as well as an explanation that the given word was created by the author to achieve a particular stylistic effect. Where the term is an eccentricity of the novel, the glossary could also include the commonly used synonym, for example, *baby* or *infant* rather than *newchild.* Alternatively, publishers include footnotes with definitions of low frequency words, such as *churn*, which in the following example may not be easily understood through the context.

- He had been frightened then. The sense of his own community silent, waiting, had made his stomach churn. He had trembled. (Lowry, 1993, p. 2)

Moreover, simply adding a brief definition at the bottom of the page, such as this one taken from the *Merriam-Webster Learner's Online Dictionary* (2020),

- 4 [no obj]: to experience a lot of confused activity
  ◊If your stomach churns or something churns your stomach, you feel sick from nervousness, disgust, etc.

could help students quickly look up the meaning of words without interrupting the fluency of their reading by having to flip to a glossary in the back of the book.

Another helpful addition could be the use of illustrations with captions that demonstrate the meaning of low frequency words. Since low frequency words may not necessarily be useful to memorize and draw sufficient attention to bring them into the active knowledge, the use of illustrations can help students understand meaning quickly without interrupting the flow of their reading or drawing unnecessary attention and effort to the word. For example, when three low frequency bicycle vocabulary words are introduced in the novel in the sentence, "as he WHEELED his BIKE into its narrow PORT beside the door" (Lowry, 1993, p. 4), an image of a boy pushing a bike toward a bike rack could illustrate the scene with the sentence below it as a caption, so that students can understand with minimal effort.

In addition, it may be beneficial to provide a pre-reading section for each chapter that pre-teaches the meaning of low frequency words. For example, a brief activity with an answer key, as demonstrated in Figure 1, could prepare students to recognize words as they come across them in a chapter.

Match each word on the left to its definition on the right.

1. childcare	a) the things that are done to take care of children especially when their parents are away or at work
2. aircraft	b) the state of feeling foolish in front of others
3. squint	c) an area of land that has a particular quality or appearance
4. riverbank	d) to look at something with your eyes partly closed
5. landscape	e) a machine (such as an airplane or a helicopter) that flies through the air
6. pilot	f) the ground at the edge of a river
7. embarrassment	g) a person who flies an airplane, helicopter, etc.

Answers: 1a, 2e, 3d, 4f, 5c, 6g, 7b

FIGURE 1. Sample of Chapter 1 vocabulary preview

These same suggestions could be used as supplements to the text for a classroom teacher using the text in its "as is," unrevised form. A teacher focusing on 2K and AWL vocabulary could create vocabulary glossary handouts or exercises for each chapter.

CONCLUSION

Language learners need to understand 95-99% of the words in a text to achieve comprehension. In *The Giver,* 89.07% of vocabulary was found to fall within the 1K, 2K, and AWL high frequency word lists. This gives the novel a somewhat higher comprehensibility than average academic texts, which were found to be comprised of 86.1% high frequency words (Nation, 2001). In order to achieve comprehensibility, language learners need scaffolding and academic support to understand low frequency vocabulary, including explicit vocabulary teaching strategies (Nation, 2001, p. 20; Nation & Waring, 1997).

A potential confounding variable in the corpus results presented in this study is that many of the words categorized in the corpus as low frequency words, upon further examination, were revealed to be high frequency words used as compound words, words with prefixes, and invented words composed of two real words. Excluding these words based on the assumption that learners could utilize reading strategies to deduce word meaning could have revealed a much higher percentage of high frequency words. Furthermore, the usage of such vocabulary should be considered a positive attribute of the text as an instructional material because it increases comprehensibility of the text and provides opportunities for students to practice analyzing word parts to deduce meaning and the analysis of root words.

One disadvantage to using *The Giver* in IEP curriculum revealed by analysis of low frequency words and their collocates is frequent eccentric vocabulary usages that could lead learners to develop incongruous understandings of words and their common usages. For example, the novelist refers to bike racks as *"bikeports."* If the learner is unaware of the idiosyncratic nature of this term and utilizes it in spoken or written English, it could lead interlocutors to be unable to understand. This is because, as Alcaraz-Marmol (2015) points out, learners' understanding is based upon what they are exposed to in class (p. 6); therefore, it is important for instructors to provide further explanation around eccentric vocabulary usage. Clearly, to prevent the development of erroneous vocabulary knowledge, educational intervention is necessary.

With its greater percentage of high frequency words, this novel presents greater opportunity for learners to practice reading comprehension, increase vocabulary size, and build reading fluency than most academic texts. On the other hand, it also runs the risk of leading learners to develop false understandings of vocabulary. Due to the eccentric nature of low frequency words in this novel, publishers and educational institutions that choose to use it as part of an English language learning curriculum should take steps to prevent the development of incongruous vocabulary knowledge and scaffold understanding of the text.

Firstly, direct instruction of the 1K, 2K, and AWL vocabulary lists should take place prior to reading the novel to facilitate comprehension. To address low frequency and eccentric vocabulary, providing glossaries, footnotes, illustrations, and vocabulary preview activities could make the text more comprehensible. Reading of the novel should also be supported by guided reading strategy instruction that teaches students how to analyze root words to deduce meaning, utilize context clues to make inferences, and analyze word families. As some words were found to be frequent in the novel, but not in common usage in other corpora of the English language, students should also be supported in learning to differentiate between context-bound vocabulary and vocabulary that they should study and memorize.

REFERENCES

Alcaraz-Mármol, G. (2015). Dispersion and frequency: Is there any difference as regards their relation to L2 vocabulary gains? *International Journal of English Studies*, *15*, (2), 1-16. https://doi.org/10.6018/ijes/2015/2/201471

Anthony, L. (2020). AntConc (Version 3.5.9) [Computer Software]. Waseda University. https://www.laurenceanthony.net/software

Cameron, L. (2001). *Teaching languages to young learners*. Cambridge University Press. https://doi.org/10.1017/CBO9780511733109

Cobb, T. (n.d.). Web concordancers. Lextutor. Retrieved October 10, 2020, from https://www.lextutor.ca/conc/

Cobb, T. (n.d.) VocabProfilers. Lextutor. Retrieved October 8, 2020, from http://www.lextutor.ca/vp/

Cobb, T. (1998, 2019) Research base: Why & how to use frequency lists to learn words. Lextutor. Retrieved October 10, 2020, from http://www.lextutor.ca/research/

Davies, M. (2008-) The corpus of contemporary American English (COCA): One billion words, 1990-2019. Retrieved October 20, 2020, from https://www.english-corpora.org/coca/

Gao, Y. (2015). An empirical study on the correlation between vocabulary size and reading comprehension. *Course Education Research*, (16), 199-200.

Granger, S. (2002). A bird's-eye view of learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching*(pp. 3–33). John Benjamins Publishing. https://doi.org/10.1075/lllt.6.04gra

Hoey, M. (2000). The hidden lexical clues of textual organisation: A preliminary investigation into an unusual text from a corpus perspective. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 31–42). Peter Lang.

Horst, M. (2000). *Text encounters of the frequent kind: Learning L2 vocabulary through reading*. Unpublished doctoral dissertation, University of Wales, Swansea.

Hsueh-Chao, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403-430. https://eric.ed.gov/?id=EJ626518

Jones, M., & Durrant, P. (2010). What can a corpus tell us about vocabulary teaching materials. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 387-400). Routledge.

Laosrirattanachai, P., & Ruangjaroon, S. (2021). Implementation of a data-driven hospitality Lexis learning programme. *3L: The Southeast Asian Journal of English Language Studies*, *27*(1). https://doi.org/10.17576/3L-2021-2701-01

Laufer, B. & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15-30. https://eric.ed.gov/?id=EJ887873

Lee, J. Y. V., & Wong, A. S. C. (2020). Vocabulary size and critical academic reading ability of secondary students in Sabah. *3L: The Southeast Asian Journal of English Language Studies*, *26*(4), 15-27. https://doi.org/10.17576/3L-2020-2604-02

Lowry, L. (1993). *The Giver*. Houghton Mifflin.

Merriam-Webster. (2020). *Merriam-Webster learner's dictionary*. Retrieved October 10, 2020, from http://www.learnersdictionary.com/

Millar, E. (2016). In search of a common core of key vocabulary among EFL coursebooks for secondary education in Cantabria using corpus linguistics applications. *III Jornadas Iberoamericanas de Innovación Educativa en el ámbito de las TIC*, 345-351. https://repositorio.unican.es/xmlui/bitstream/handle/10902/18970/SearchCommonCore.pdf?sequence=1

Millar, N., & Budgell, B. S. (2008). The language of public health—a corpus-based analysis. *Journal of Public Health*, *16*(5), 369-374. https://doi.org/10.1007/s10389-008-0178-9

Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *Eurosla monographs series, 2: L2 vocabulary acquisition, knowledge and use: New perspectives on assessment and corpus analysis*, (pp. 57-78). European Association of Second Language Acquisition.

Molavi, A, Koosha, M. & Hosseini, H. (2014). A comparative corpus–based analysis of lexical collocations used in EFL textbooks. *Latin American Journal of Content & Language Integrated Learning*, *7*(1), 66-81. https://doi.org/10.5294/laclil.2014.7.1.4

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

Nation, I. S. P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 6-19). Cambridge University Press.

https://www.scirp.org/(S(czeh2tfqyw2orz553k1w0r45))/reference/ReferencesPapers.aspx?ReferenceID =1590729

Norberg, C., & Nordlund, M. (2018). A corpus-based study of lexis in L2 English textbooks. *Journal of Language Teaching and Research*, *9*(3), 463-473. https://doi.org/10.17507/jltr.0903.03

Port. (n.d.). In *Merriam-Webster's Online Dictionary*. http://www.merriam-webster.com/dictionary/port

Sarem, S. N., Hamidi, H., & Mahmoudie, R. (2013). A critical look at textbook evaluation: A case study of evaluating an ESP course-book: English for international tourism. *International Research Journal of Applied and Basic Sciences*, *4*(2),372-380. https://www.academia.edu/14070844/A_Critical_Look_at_Textbook_Evaluation_A_Case_Study_of_E valuating_an_ESP_Course_Book_English_for_International_Tourism?auto=download

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26-43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., & Schmitt, D. (2000). *Vocabulary in language teaching*. Cambridge University Press.

Stockton, S. K. (2020). *A corpus-based analysis of junior and senior high school English textbooks in Japan* [Doctoral dissertation, Northern Arizona University].

Tyler, A. (2012). *Cognitive linguistics and second language learning: Theoretical basics and experimental evidence*. Routledge.

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Webb, S., & Nation, I. S. P. (2008). Evaluating the vocabulary load of written text. *TESOLANZ Journal*, *16*, 1-10. https://www.wgtn.ac.nz/lals/resources/paul-nations-resources/paul-nations-publications/publications/documents/2008-Webb-Evaluating-vocabulary-load.pdf

Wei, X. (2020). How does vocabulary knowledge relate to reading comprehension? In M. Dodigovic & M. P. Agustin-Llach (Eds.), *Vocabulary in curriculum planning: Needs, strategies and tools* (pp. 57-74). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-48663-1_4

Yan, L., & Yang, Y. (2016). Examining business English majors' business vocabulary knowledge development. *The Asian ESP Journal*, *12*(3, 40-69

Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *Canadian Modern Language Review*, *57*(4), 541-572. https://doi.org/10.3138/cmlr.57.4.541

Zarifi, A., & Mukundan, J. (2013). Phrasal verb combinations in corpus-based studies: A critical review. *International Journal of Applied Linguistics & English Literature*, *2*(4), 212-217. https://doi.org/10.7575/aiac.ijalel.v.2n.4p.212