

A Brief Look at the Word Length of On-line Learner Dictionary Definitions

Kathryn Zidonis Dilenschneider
Robert Dilenschneider

Abstract

This study examines the word-length of definitions from three well-known on-line learner dictionaries. The defining vocabulary of twelve target words was analyzed to investigate the average number of words the three learner on-line dictionaries use to define words. Results show that target words defined by the Cambridge on-line learner dictionary yield shorter definitions compared to the Merriam-Webster and Oxford on-line learner dictionaries. The findings suggest the Cambridge on-line learner dictionary might be the best choice to create learning activities and materials that require shorter definitions.

Background

In recent decades, the popularity, availability and accessibility of on-line dictionaries have increased. For instance, with regard to popularity, Lan (2005) stated that 70% of the students interviewed at the Polytechnic University in Hong Kong used an on-line dictionary more frequently than a book dictionary. With regard to availability, the number of available on-line dictionaries from 1997 to 2011 has increased. *Onelook dictionary.com*, for example, expanded its index from just 188 to 1,063 dictionaries, and in 2014, *World Language.com* offered hundreds of

dictionary products in 175 languages. Lastly, with regard to accessibility, Nation (2011) stated, “Increasingly, electronic dictionaries of various kinds are now available, and ownership is becoming less important than access because many dictionaries are available on-line, not only through computers but also through mobile phones and other electronic devices” (p. 211).

The growing number of on-line dictionaries provides a variety of features and functions that are not available in book-based dictionaries to assist with the learning of new words. For instance, like traditional book-based dictionaries, on-line dictionaries offer features such as word definitions, word spellings, parts of speech, pronunciation symbols, synonyms and example sentences. However, unlike traditional book-based dictionaries, on-line dictionaries also provide functions such as audio pronunciations, bilingual translations, podcasts, and word banks for learners to save and review words. As a result, learners may be attracted to these features and functions of on-line dictionaries whenever they encounter unknown words in reading passages.

Statement of the Problem

Despite the popularity and features of using on-line dictionaries, some dictionaries may incorporate more words to define a target word than others. There are legitimate reasons why longer definitions from some types of dictionaries may be a disadvantage for learning compared to shorter definitions from other dictionaries. First, long definitions may make it difficult for learners to understand a target word. In her review, *Dictionary Definitions: Problem and Solutions*, author Adamska-Salaciak states, “Even in the absence of any particularly complex syntactic structures, the sheer length of the entry can prove an obstacle to comprehension, especially for users whose native language is typologically distant from the language of the dictionary” (Salaciak 2012, 279). As a result, if learners need to read lengthy definitions positioned in a gloss in order to comprehend a reading passage, they may become frustrated with having to read several unfamiliar words within a definition just to gain an understanding of a single word. This may be particularly unnerving if

learners are under a time constraint.

Second, another obstacle of long definitions learner encounters is that it may be difficult for them to create vocabulary flashcards for self-study. For instance, learners might be limited by the amount of space they are afforded to create flashcards on a digital platform such as Anki, Brainscape, or Quizlet. Shorter definitions, however, would assure that learners could include the necessary information provided by a reliable source to clearly understand the meanings of unknown words.

Third, long definitions provided by an on-line dictionary might also present a challenge for language instructors to create learning materials. The amount of space afforded to instructors on a page to include both the content and the meaning of unknown words might be compromised if target words essential to the understanding of a reading passage have lengthy definitions. For instance, if six unknown words need to be defined in a gloss in the margin of a page either beside or under a reading passage, a dictionary that provides short and succinct definitions would be ideal because it would allow both the reading passage and the defined words in a gloss on the same page. However, if a dictionary provides, on average, just two extra words to define each word, the amount of space for a gloss could be diminished due to the extra twelve words needed to define the six unknown target words. As a result, all six unknown words necessary to comprehend a reading passage might not be included or adequately defined in a gloss. In addition, the content of a reading passage might have to be reduced or shifted over to another page if dictionary definitions include several words to define all six unknown words in a gloss. In the case of the former, the spatial layout intended for the essential content of a reading passage may be crowded out or may need to be omitted. In the case of the latter, shifting the content of a reading passage may force a reading exercise to flow onto additional pages. This in turn, as a potentially negative domino effect, may compromise both vocabulary learning and passage comprehension, limiting the inclusion of essential exercises intended to reinforce the understanding of a reading passage.

Methodology

The Purpose of the Study

The purpose of the study is to examine three on-line learner dictionaries in terms of the number and percentage of words used to define words. The results from this study can provide information as to what type of on-line learner dictionary might be best for EFL/ESL learners to access in order to learn new words. This study is also valuable for second language instructors because it might shed light on which on-line learner dictionary might be best to create language materials for teaching new vocabulary.

Word Frequency Levels and Word Families

When learners read non ESL targeted books, blogs, newspapers and magazines or, if they access a dictionary while reading, they will typically encounter a wide variety of words. Learners may have a solid understanding of some words but only a vague understanding of others. As a result, learners are likely to have a spectrum of understanding for the words they encounter. Therefore, in order to understand why some words are easier for them to learn than others, it is useful to take a close look at the coverage of different word frequency levels that learners encounter in books and dictionaries.

The first category of words concerns high-frequency words (Henceforth HFW). These words consist of the 2,000 most frequent word families (e.g., the *General Service List* [GSL] [West, 1953]) and cover about 68.5% of running words in spoken and written texts (Nation, 2014). If an additional 9.2% of technical words concerning the topic and subject areas belonging to the first 2,000 word families were also included, the cumulative coverage for this category of words would amount to 77.7% ($68.5 + 9.2 = 77.7\%$) (Nation, 2014).

Other words not necessarily included in this category are words such as proper nouns, transparent compounds, and marginal words. Transparent compounds, such as *ashtray* or *aftershave*, are words, “where the meaning of the compound is

transparently related to the meaning of the parts” (Nation, 2014). Marginal words, such as *ahh*, *er*, *ooh*, or *ssh*, are utterances that are rarely found in written texts and are not necessarily found in dictionaries. Nevertheless, together these three types of words make up about 3-4% of the words in an average text (Nation, 2014).

The analysis of the British National Corpus has indicated that the first 2,000 words, along with proper nouns, transparent compounds, as well as marginal words, account for around 90% of words in an average text (Nation, 2014). However, Schmitt and Schmitt (2012) suggested that high-frequency words should be seen as the first 3,000 word families. This expanded list, along with proper nouns, transparent compounds, and marginal words, provide approximately 95% coverage of an average written text.

The next category of words concerns mid-frequency words (Henceforth MFW). These words consist of the next 7,000 words beyond the high-frequency word list and stretch from the third 1,000 to the ninth 1,000 word families. MFWs provide around an additional 9% coverage of an average text. However, even though these words occur less frequently than HFW, they are useful to know because they largely consist of general-purpose vocabulary that can help learners read texts without reference to an external resource. For instance, together with high-frequency words and proper nouns, MFWs can help learners reach 98% coverage of a text (Nation, 2014).

The last category of words is low-frequency words (Henceforth LFW), which are words beyond the 9,000 word families. Although these words make up the largest group of words, they typically account for only 1-2% of the words in a text. Although learning the meanings of these words is beneficial, Nation (2014) suggested that, due to their specialized tendency and minimal coverage, low-frequency words should be learned incidentally through reading and listening.

Research Questions

Depending on their level, ESL learners will typically be more familiar with HFWs because they occur more often in texts than MFWs or low-frequency words LFWs. Therefore, to explore words that occur less often, this study involves mid-frequency

target words belonging to the 4,000 to 9,000 word-family levels and low-frequency target words belonging to the 10,000 to 24,000 word-family levels. The research questions are as follows :

1. What is the total number of words and the average number of words per definition the Merriam-Webster, Oxford and Cambridge on-line learner dictionaries use to define target words among the three word categories?
2. What is the total number of words and the average number of words per definition the Merriam-Webster, Oxford and Cambridge on-line learner dictionaries use to define mid-frequency and low-frequency target words?

Instruments

On-Line Dictionaries

The on-line learner dictionaries examined in this study are from Merriam-Webster, Oxford and Cambridge. All three companies are well known and offer different types of dictionaries. However, this study examined the three learner dictionaries that use simpler vocabulary to define words. These on-line learner dictionaries can be either accessed through their websites from a desktop computer, and in addition, application versions of these on-line dictionaries can be downloaded onto a smartphone or tablet computer for students to use virtually any place, anytime.

Vocabulary Profiler Software

The Compleat Lexical Tutor (Cobb, 2018) software program was used to examine the defining vocabulary for the target words in this study. This on-line software program provides several different vocabulary profilers in which texts and passages can be inputted to analyze word frequencies. For this study, the BNC-COCA-25 vocabulary profiler was used to provide a thorough investigation of defining vocabulary. This vocabulary profiler utilizes a combination from the 100 million words of written material from the British National Corpus (Nation, 2013) and from the 450 million word Corpus of Contemporary English (Davis, 2012). It categorizes

HFWs from the 3,000 most frequent word families, mid-frequency words MFWs from the next 7,000 words that stretch from the third 1,000 to the ninth 1,000 word families, LFWs beyond the 9,000 word families that extend fifteen word families from the tenth 1,000 to the twenty-fifth 1,000 word families, and off-list words that are identified as being above the twenty-fifth 1,000 word-family level.

Spreadsheet Software

The Microsoft Excel 2013 spreadsheet software program was used to inspect the defining vocabulary for the target words in this study. Its cells arranged figures in rows and columns and were used to organize data and its formulas were used to calculate the sums and averages of data for analysis and interpretation.

Target Words

A total of 12 target words used in this study were medical terms related to the three word categories including a body part (liver, lung, intestine, pancreas) a disease (influenza, pneumonia, leukemia, measles) and a symptom (amnesia, heartburn, indigestion, vertigo). The target words were selected at random and the analysis by the Compleat Lexical Tutor indicated that six target words have word family levels that belong to the MFW level (6-MFTW of 4, 4, 5, 5, 6, 7 word-family levels) and six target words have word family levels that belong to the LFW level (6-LFTW of 10, 10, 10, 10, 11, 12 word-family levels). No HFWs were selected or used in this study. Table 1 shows the target words for each word category and their word family levels.

Table 1. Total Words and Their Family Levels

Body Parts		Diseases		Symptoms	
Word	Family Level	Word	Family Level	Word	Family Level
liver	4	influenza	5	amnesia	10
lung	4	pneumonia	6	vertigo	10
intestine	5	leukemia	7	indigestion	11
pancreas	10	measles	10	heartburn	12

Procedure

The procedure to retrieve and organize the data for the defining vocabulary of each word involved six steps using the on-line learner dictionaries, vocabulary profiler and spreadsheet software programs. First, using the on-line learner dictionaries, the spelling of each target word was typed into a learner dictionary search window. Second, after typing the spelling of the target word, a submission button near the search window of an on-line learner dictionary was clicked to access the definition of a target word. Third, when the definition of a target word was displayed from a learner dictionary, the definition was then copied and pasted into a submission window of the BNC-COCA-25 vocabulary word profiler. Fourth, after the definition of a target word was pasted, a “submit” button within the BNC-COCA-25 vocabulary word profiler was clicked to view the output of the word families for each word in the definition of the target word. Fifth, the output of word families represented with figures for each token or word in the definition was then transferred to a Microsoft Excel spreadsheet. The output figures of tokens from the BNC-COCA-25 vocabulary word profiler were copied and pasted into their corresponding high-, mid-, low-, and off-word word frequency cell categories within the Microsoft Excel spreadsheet. Sixth, the figures that were organized into the spreadsheet were examined and calculated.

Methods of Calculation

The methods of calculation involved three steps to identify, total and average the number of words the three learner dictionaries use to define target words. First, the number of high, mid- and low frequency words used to define a target word were identified by using the Compleat Lexical Tutor BNC-COCA-25 word profiler output. For example, to define the target word *intestine*, the Merriam-Webster on-line learner dictionary used 14 HFWs, and one LFW, the Oxford on-line learner dictionary used 25 HFWs, two MFWs and one LFW, and the Cambridge on-line learner dictionary used only nine HFWs. This process was performed for all 12 target words.

Second, the words from the definition were added to find the sum or total number of high, mid- and low frequency words that are used to define the four target words in a category. For example, the four target words of *lung*, *liver*, *intestine* and *pancreas* belonging to the Body Part category were defined by the Oxford learner dictionary with 66 HFWs ($13 + 12 + 25 + 16 = 66$), five MFWs ($0 + 1 + 2 + 2 = 5$) and no LFWs ($0 + 0 + 0 + 0 + 0 = 0$). This type of calculation was also performed for the four target word belonging to the Disease category and the Symptom category.

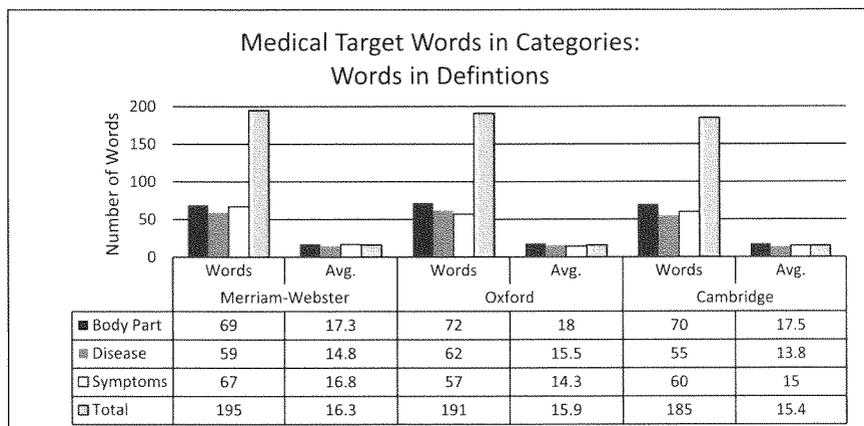
Results

Research question 1 : What is the total number of words and the average number of words per definition the Merriam-Webster, Oxford and Cambridge on-line learner dictionaries use to define target words among the three word categories?

The total number of words to define target words and the average number of words in a definition differs slightly among the three on-line learner dictionaries for the three word categories. The Merriam-Webster on-line learner dictionary uses 69 words to define the four Body Part target words with an average of 17 words per

definition ($69 \div 4 = 17.3$). The number of words and average number of words per definition for the same four target words from the Oxford and Cambridge on-line learner dictionaries is 72 (18.0) and 70 (17.5) words, respectively. The total number of words and the average number of words per definition for the same respective dictionaries is 59 (14.8), 62 (15.5) and 55 (13.8) for the four target words belonging to the Disease category and 67 (16.8), 57 (14.3) and 60 (15.0) for the four target words belonging to the Symptoms category. The total number of words and the average number of words per definition to define all 12 target words from Merriam-Webster, Oxford and Cambridge on-line learner dictionaries is 195 (16.3), 191 (15.9) and 185 (15.4) words, respectively. Table 2 reveals the total number of words and the average number of words per definition each on-line learner dictionary incorporated to define the four target words in each word category and for the 12 target words as a whole.

Table 2. Total Number of Words and Average Number of Words Per Definition



Note. Avg. = Approximate averages for the number of words per definition based on four target words.

There are three findings concerning the total number of words and average number of words per definition to define target words from the three on-line learner

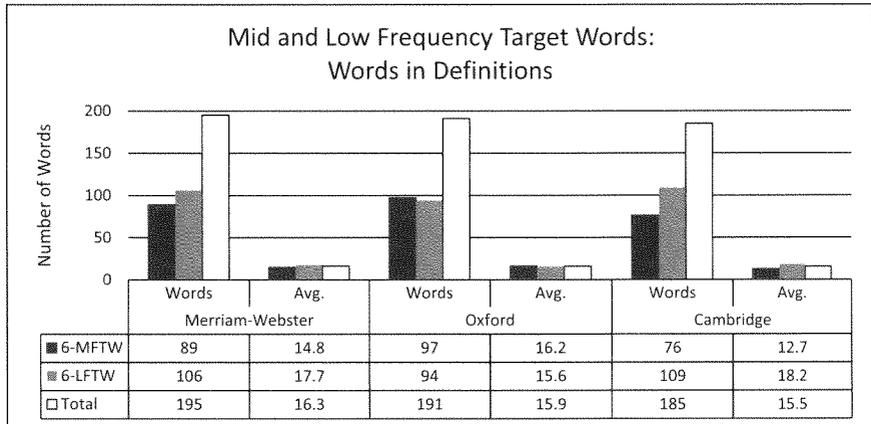
dictionaries. First, with regard to the four target words in each of the three word categories, the total number of defining words fluctuates from the three learner dictionaries among the three categories. For the Body Part category, the Merriam-Webster and Oxford dictionaries differ by three words, the Oxford and Cambridge differ by two words and the Merriam-Webster and Cambridge differ by one word. For the Disease category, the Merriam-Webster and Oxford dictionaries differ by three words, the Oxford and Cambridge differ by seven words and the Merriam-Webster and Cambridge differ by four words. For the Symptoms category, the Merriam-Webster and Oxford dictionaries differ by ten words, the Oxford and Cambridge differ by three words and the Merriam-Webster and Cambridge differ by seven words. Overall, it appears the largest differences occur for the target words belonging to the Symptoms category. Second, if figures that are .5 or higher are round up and figures that are .4 or lower are round down are taken into account, the average number of words per definition does not differ by more than two words among the three word categories. Third, with regard to the 12 target words as a whole, although the Cambridge on-line learner dictionary provides the fewest number of words per definition among the three on-line learner dictionaries (Merriam-Webster, 195 ; Oxford, 191 ; Cambridge, 185), the difference for average number of words per definition differs by one word (Merriam-Webster, 16.3 ; Oxford, 15.9 ; Cambridge, 15.4).

Research question 2 : What is the total number of words and the average number of words per definition the Merriam-Webster, Oxford and Cambridge on-line learner dictionaries use to define mid-frequency and low-frequency target words?

The number of words to define target words also differs between the two groups of six mid-frequency words and six low frequency words. The Merriam-Webster on-line learner dictionary uses 89 words to define the 6-MFTW with an average of 15 words per definition ($89 \div 6 = 14.8$). The total number of words and the average

number of words per definition for the same six target words from the Oxford and Cambridge on-line learner dictionaries is 97 (16.2) and 76 (12.7) words, respectively. The total number of words and the average number of words per definition for the 6-LFTW from the same corresponding on-line learner dictionaries is 106 (17.7), 94 (15.6) and 109 (18.2) words, respectively. Table 3 reveals the total number of words and the average number of words incorporated into definitions to define the groups of 6-MFTW and 6-LFTW from each on-line learner dictionary.

Table 3. Total Number of Words and Average Number of Words incorporated into Definitions for Mid-and Low-Frequency Target Words



Note. Avg. = Approximate averages for the number of words per definition based on six target words.

There are three findings with regard to the two groups of mid- and low frequency target words. First, with regard to the 6-MFTW, the Cambridge on-line dictionary provides the fewest words in its definitions and, if figures of .5 or higher are rounded up and figures of .4 or lower are rounded down are taken into account, it averages to two and three fewer words per definition compared to the Merriam-Webster and Oxford on-line learner dictionaries. Second, with regard to the 6-LFTW, the Oxford on-line learner dictionary provides the fewest words in its definitions with an average

of two fewer words than both the Merriam-Webster and Cambridge on-line learner dictionaries. Third, with regard to the 12 target words as a whole, although the Cambridge on-line dictionary provides the fewest number of words in its definitions, it averages the same number of words per definition as the Merriam-Webster and Oxford on-line learner dictionaries.

Discussion

This study reveals important findings to consider with regard to the number of words that are used to define words from on-line learner dictionaries. First, results varied among the three word categories. On one hand, the number of words used to define the four target words from the on-line learner dictionaries fluctuates among the three categories from one to ten words. In addition, for the 12 target words as a whole, the Cambridge on-line learner dictionary provides 10 fewer words in its definitions compared to the Merriam-Webster on-line learner dictionary ($195-185 = 10$) and six fewer words compared the Oxford on-line learner dictionary ($191-185 = 6$). On the other hand, the difference for average number of words per definition differs by one to three words among the three categories and only one word overall for all 12 target words. As a result, although the total number of defining words varies, the actual difference for the average number of words to define target words per definition is relatively small.

However, the number of words to define target words is particularly evident between the group of six mid-frequency target words and six low frequency target words. On one hand, for the 6-MFTW, the Cambridge on-line learner dictionary provides 13 fewer words in its definitions compared to the Merriam-Webster on-line learner dictionary ($89-76 = 13$) and 21 fewer words than the Oxford on-line learner dictionary ($97-76 = 21$). As a result, the average number of words per definition from the Cambridge on-line learner dictionary is two words fewer than the Merriam-Webster on-line learner dictionary ($14.8-12.7 = 2.1$) and four words fewer than the Oxford on-line learner dictionary ($16.2-12.7 = 3.5$). On the other hand, for the 6-LFTW, the Cambridge on-line learner dictionary provides three more words than

the Merriam-Webster on-line learner dictionary ($109-106 = 3$) and 15 more words than the Oxford on-line learner dictionary ($109-94 = 15$). The average number of words per definition from the Cambridge on-line learner dictionary is about the same as the Merriam-Webster on-line learner dictionary ($18.2-17.7 = 0.5$) but three words more than the Oxford on-line learner dictionary ($18.2-15.6 = 2.6$).

Nevertheless, despite these findings between the 6-MFTW and 6-LFTW, overall both the Cambridge and Oxford on-line learner dictionaries provide the fewest number of words to define both types of target words and on average provide the fewest number words per definition.

Conclusion

This study has provided a quick look at the number of words on-line learner dictionaries might use to define target words. Although this preliminary study did not conduct analyses to determine statistical significance, we believe its observational data suggest that the Cambridge on-line learner dictionary might use fewer mid-frequency words to define target words compared to the Merriam-Webster or Oxford on-line learner dictionaries. Therefore, due to its concise method of defining vocabulary, the Cambridge on-line learner dictionary appears to be the most accommodating on-line dictionary in terms of the word length of its definitions for both teachers to create and design learning materials and learners to more easily learn new mid-frequency words into their vocabulary.

This study provides an orientation to the word length of definitions provided from on-line learner dictionaries. Nonetheless, there are some issues that future studies can explore to further confirm its findings. First, the number of target words utilized in this study was limited. It would be useful for studies to incorporate a larger number of both mid-frequency and low-frequency words. Second, the words used in this study are all from the same part of speech, nouns. Future studies might incorporate and analyze how target words from different parts of speech such as adjectives, verbs and adverbs are defined. Such studies would confirm and build upon the findings of this study to help teachers and learners discover and use the

most conducive on-line learner dictionary for teaching and learning.

References

- Adamska-Salaciak, A. (2012). Dictionary definitions: problems and solutions. *Studia Linguistica Universitatis Jagellonicae Cracoviensis*, 129, 268-283. DOI 10.4467/20834624.SL.12.020.0804
- Cambridge Dictionary. (n.d.). Retrieved July 5, 2018, from <https://dictionary.cambridge.org/dictionary/learner-english/>
- Cobb, T. (2018). Compleat Lexical Tutor. (Version 8) [Computer Software]. Available from <http://www.lextutor.ca/>
- Davies, M. (2008) *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.
- Lan, Li, (2005). The growing property of on-line dictionaries. *English Today*, 83(3), 16-21.
- Merriam-Webster Dictionary. (n.d.). Retrieved July 5, 2018, from <http://www.learnersdictionary.com>
- Nation, I. S. P., & Webb, S. (2011). Content-based instruction and vocabulary learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 631-644). New York, NY: Routledge.
- Nation, I. S. P., & Anthony, L. (2013). Mid-frequency readers. *Journal of Extensive Reading*, 1, 5-16.
- Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language*, 26, 1-16.
- OneLookDictionary.com. (1996). Retrieved October 26 th, 2011, from <http://www.onelook.com>
- Oxford Learner dictionary. (n.d.). Retrieved July 5, 2018 from <https://www.oxfordlearnersdictionaries.com/>
- Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 1-20.
- West, M. (1953). *A General Service List of English Words*. London: Longman,

Green & Co.

WorldLanguage.com. (1996). Retrieved October 26, 2011, from <http://www.worldlanguage.com>