Jichi Medical University Journal 2016, 39, 51-57

Research Reports

Reading Time to Increase Mid-Frequency Vocabulary

Robert F. Dilenschneider

General Education Department, English, Jichi Medical University, Yakushiji 3311-1, Shimotsuke, Tochigi, 329-0498 Japan

Abstract

This paper examines and expands upon the findings of previous research concerning the reading time required for language learners to increase their mid-frequency vocabulary. First, it summarizes previous studies that have examined word frequency levels and reading comprehension. Second, it reviews analyses of both the Corpus of Contemporary American English（COCA）and the British National Corpus（BNC）. Third it analyzes through means of word repetition and reading time. Finally, it discusses and pedagogical implications with regard to enhancing the exposure to unknown mid-frequency words.
（Key words : Reading time, Vocabulary）

Previous Studies

Previous studies related to vocabulary acquisition have examined both the categories of words within a text as well as the number of words that learners should know in order to read a text without the use of a resource such as a gloss, glossary or dictionary. With regard to the percentage of running words that are covered from a typical text, West（1953）, Schmitt and Schmitt（2012）and Nation（2014）have determined that words can be principally categorized into three word frequency levels. Their first category of high frequency words consist of the 2,000 most frequent word families and cover about 68.5% of running words in spoken and written texts. Their second category of mid-frequency words consists of the next 7,000 words beyond the high-frequency word list from the third 1,000 to the ninth 1,000 word families and provides an additional 9% coverage of an average text. Their third category is low-frequency words, consisting of words beyond the 9,000 word families. Although these low frequency words make up the largest group of words, they typically account for only 1-2% of words in a text. Not included in these three word categories are proper nouns, transparent compounds, and marginal words, making up approximately 3-4 % of the words in an average text（Nation, 2014）. With regard to the number of words necessary for unaided coverage, Prichard and Matsumoto（2011）and Schmitt, Jiang, and Grabe（2011）have determined that to read and understand texts without an external resource, language learners should know approximately 98% of the words in a text. Using these findings as a basis, it is useful to take a close look at how the coverage of different word frequency levels from a language corpus might be used to help learners achieve the 98% coverage necessary to read texts without the need for an external resource such as a dictionary.

Analyses of Corpora

The two types of language corpus that have been analyzed are the Corpus of Contemporary American English（COCA）and the British National Corpus（BNC）. First, the Corpus of Contemporary American English（COCA）is a collection of more than 450 million words derived from over 160,000 texts. It includes samples of words from written texts such as books, magazines, newspapers, and academic journals as well as roughly roughly 85 million words of conversational transcripts from television and radio programs. Davies（2008）examined the coverage provided by the COCA. Table 1 shows the coverage of the COCA.

Correspondence to：Robert Dilenschneider, General Education Department, English, Jichi Medical University School of Medicine 3311-1 Yakushiji, Shimotsuke-shi, Tochigi, 329-0498 Japan. E-mail：bdilenschneider@yahoo.com
Received：30 September 2016, Accepted：2 February 2017

Table 1. *Corpus of Contemporary American English (COCA) (Davies 2008)*

Frequency level	%cover added by level	Cumulative%
1st 1,000	78.5	78.5
2nd 1,000	7.9	86.4
3rd 1,000	3.0	89.4
4th 1,000	2.4	91.8
5th 1,000	1.3	93.1
6th 1,000	0.9	94.0
7th 1,000	0.6	94.6
8th 1,000	0.5	95.1
9th 1,000	0.4	95.1

Note. Numerals, words with apostrophes, and proper nouns are excluded.

Table 1 shows the coverage provided from both the high and mid-frequency word levels. For example, the first 3,000 high frequency words from the COCA provide 89.4% coverage (78.5% + 7.9% + 3.0% = 89.4%), and the cumulative coverage for the first nine 1,000 word families is 95.5% (78.5 + 7.9 + 3.0 + 2.4 + 1.3 + 0.9 + 0.6 + 0.5 + 0.4 = 95.5%).

The BNC is the other major corpus that has been analyzed. It is a 100 million word collection of samples of written language from a wide range of sources, such as novels, newspapers, periodicals, journals and academic books that represent the type of vocabulary used from the latter part of the 20th century. Unlike the COCA, the BNC does not include conversational transcripts from television and radio programs. The British National Corpus was divided into twenty 1,000 word-frequency levels plus other groups of words, such as proper nouns, marginal words and compounds. Nation (2006) used the British National Corpus (BNC) to examine the coverage of words in different frequency levels. Table 2 shows the coverage provided by the BNC.

Table 2. *Coverage of the British National Corpus by Word-Frequency Lists (Nation 2006)*

Frequency level	% coverage of tokens	%cumulative coverage
1st 1,000	77.96	81.14
2nd 1,000	8.10	89.24
3rd 1,000	4.36	93.60
4th 1,000	1.77	95.37
5th 1,000	1.04	96.41
6th 1,000	0.67	97.08
7th 1,000	0.45	97.53
8th 1,000	0.33	97.86
9th 1,000	0.22	98.08
10th 1,000	0.28	98.23
11th 1,000	0.15	98.38
12th 1,000	0.11	98.49
13th 1,000	0.09	98.58
14th 1,000	0.07	98.65
15th 1,000	0.06	98.71
16th 1,000	0.04	98.75
17th 1,000	0.04	98.79
18th 1,000	0.03	98.83
19th 1,000	0.02	98.86
20th 1,000	0.01	98.86
Proper nouns	2.57	–
Marginal words	0.31	–
Compounds	0.30	–
Not in the lists	1.02	

The BNC shows comparable cumulative coverage for texts like that of Davies's analysis of the COCA. For example, the first 3,000 high-frequency words provide about 90% text coverage (77.96% + 8.10% + 4.36% = 90.42%). The coverage for the following six 1,000 mid-frequency word levels is between 4% and 5% (1.77 + 1.04 + .67 + .45 + .33 + .22 = 4.48%). Low-frequency words from the tenth through the fourteenth 1,000 word families provide less than 1% coverage (.28 + .15 + .11 + .09 + .07 + .06 + .04 + .04 + .03 + .02 + .01 = .9%). In addition, proper nouns, marginal words, and transparent compounds account for 3% to 4% coverage (2.57 + .31 + .30 = 3.18%).

In written texts, the meanings and spellings of words differ somewhat between American English and British English. In terms of vocabulary, when referring to the same respective parts of a car, speakers of American English would say *hood* or *trunk* while speakers of British English would say *bonnet* or *boot*. In terms of spelling, American English writers use the letters or to spell the word as color while British English writers use the letters *our* to spell colour. However, despite these differences between the two types of English, social linguists such as McKay and Hornberger (1996) have stated that both American and British speakers are able to, "read each others' newspapers and novels without any serious impediments" (p. 75). Therefore, the use of either the COCA or BNC corpus can be used to show how it is possible for learners of either American English or British English to obtain the 98% criterion for adequate text coverage if additional words are included. For example, the analysis for the British National Corpus reveals that proper nouns, marginal words, and transparent compounds provide 3.18% coverage. If this figure is added to the 94.9% cumulative coverage for the first 9,000 BNC word families mentioned above, learners can obtain sufficient coverage to read texts without an external resource (3.18% + 94.9% = 98.08%). In the same respect, because the maximum difference between the corpora differs only by 1%, if the 3.18%, or an approximate 2.18% figure for extra words is added to the 95.5% cumulative coverage for the first 9,000 word families of the COCA, learners can obtain or come very close to obtaining

sufficient coverage to read texts without an external resource (3.18 + 95.5 = 98.68% or 2.18 + 95.5 = 97.68%).

Word Exposure Repetition

As seen from both the analyses from the COCA and the BNC, 98% text coverage can be obtained if learners are acquainted with the first 8-9,000 word families in addition to other words, such as proper nouns, compound words, and transparent words (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Nation, 2006; Schmitt, Jiang, & Grabe, 2011). However, for learners to acquire this percentage of text coverage, they must have sufficient exposure to new words. Although previous research has demonstrated that repetition strongly influences learning (Waring & Takaki, 2003; Webb 2007b), some words are more difficult to learn than others, and thus there is not a definitive number of repetitions that ensures the learning of new words.

The problem concerning the sufficient number of repetitions necessary for learners to gain an understanding of new words might be address by Nation (2014), who conducted an analysis of 25 classic novels that provided a rough estimate of the amount of reading needed for learners to experience at least twelve repetitions of a word at each of the nine 1,000 word levels. Using this analysis as a basis, it is possible to understand not only the number of novels learners need to read in order to experience at least twelve repetitions of a new word, but also to determine the amount of time learners need to devote to reading. Table 3 shows the results from the analysis of 25 classic novels.

According to Table 3, to acquire mid-frequency words belonging to the 5,000-word frequency family, learners need to read nine novels of just over one million-word corpus to receive 13 repetitions of a word. Or, to acquire mid-frequency words belonging to the 8,000-word frequency family, learners need to read 20 novels of an approximate 2.4 million word corpus to receive 14 repetitions of a word. As the level of mid-frequency words increases, learners need to read a greater number of novels to experience at least an average of twelve repetitions of exposure to new unknown words. To understand this amount of reading in another way, the same number of tokens in Table 3 can be calculated into hours of reading per week. Table 4 presents the findings from Nation (2014), but also expands upon those

Table 3. *Corpus Sizes Needed to Gain an Average of Twelve Repetitions at Each of the 1,000 Word Levels Using a Corpus of Novels (Nation, 2014)*

1,000 word list level	Corpus size to get an average of at least 12 repetitions at this word level (repetitions)	Number of 1 timers / 2 timers out of 1,000	Number of families met	Numbers of novels
2nd 1,000	171,411 (13.4)	84/99	805 of 2nd 1,000	2
3rd 1,000	300,219 (12.6)	83/73	830 of 3rd 1,000	3
4th 1,000	534,697 (12.6)	93/73	812 of 4th 1,000	6
5th 1,000	1,061,382 (13.7)	101/79	807 of 5th 1,000	9
6th 1,000	1,450,068 (13.1)	89/82	795 of 6th 1,000	13
7th 1,000	2,035,809 (13.7)	92/63	766 of 7th 1,000	16
8th 1,000	2,427,807 (14.1)	96/70	755 of 8th 1,000	20
9th 1,000	2,956,908 (12.0)	96/70	805 of 9th 1,000	25
10th 1,000	2,956,908 (9.8)	88/78	754 of 10th 1,000	25

Table 4. *Amount of Reading in Tokens and Amount of Time per Week to Meet the 1,000 Word Families an average of Twelve Times (Nation, 2014)*

Frequency level	Amount of reading	Amount of time per week @ 200 wpm	Hours per week @ 100 wpm
2nd 1,000	171,411 (13.4)	21 minutes	42 minutes
3rd 1,000	300,219 (12.6)	38 minutes	1 hour 16 minutes
4th 1,000	534,697 (12.6)	1 hour 5 minutes	2 hours 10 minutes
5th 1,000	1,061,382 (13.7)	2 hours 12 minutes	4 hours 24 minutes
6th 1,000	1,450,068 (13.1)	3 hours	6 hours
7th 1,000	2,035,809 (13.7)	4 hours 5 minutes	8 hours 10 minutes
8th 1,000	2,427,807 (14.1)	5 hours 3 minutes	10 hours 6 minutes
9th 1,000	2,956,908 (12.0)	6 hours 10 minutes	12 hours 20 minutes
10th 1,000	2,956,908 (9.8)	6 hours 10 minutes	12 hours 20 minutes

findings if reader' s reads at half the speed. For example, the times listed in the third column in Table 4 assume that learners read an average of 200 words per minute and the times listed in column 4 assume that learners read at a rate of 100 words per minute.

Table 4 reveals that it is possible for learners to easily expand their vocabulary to include high frequency words; however, more reading time is required for learners to obtain sufficient repetitions of words as word frequency decreases. For instance, if a passage contains 2% unknown target words, assuming that learners are able to read five days a week for 40 weeks at a rate of 200 words per minute, to increase their vocabulary to the third 1,000 level, they need to read 38 minutes a week or between seven to eight minutes a day five days a week (38 minutes ÷ 5 days = 7.6 minutes). Learners who read at half that speed would take twice as long (15.2 minutes). However, for the same 40-week five-day period, learners wanting to increase their vocabulary to the eighth 1,000 level and who are able to read at a rate of 200 words per minute, need to read an hour a day for five days a week (5 hours and 3 minutes or 303 minutes ÷ 5 days = 60.6 minutes), while learners who are able to read at a rate of 100 words per minute would take twice as long (121 minutes). Table 5 converts the figures from Table 4 to show how much reading is required to increase vocabulary over a 40-week five-day period.

Table 5 shows that language learners can increase their exposure to new words if they are assigned material to read from their language instructor over a typical five-day academic week. For example, in just 26 minutes learners who are able to read around 200 words per minute can experience approximately 13 repetitions of mid-frequency words from the fifth 1,000 word level (2 hours and 12 minutes or 132 minutes ÷ 5 days = 26.4 minutes). In the same respect, learners who are able to read 100 words per minute can experience approximately 12 repetitions of mid-frequency words belonging to the fourth 1,000 word level (2 hours and 10 minutes or 130 minutes÷5 days = 26 minutes). As a result, learners can gain significant exposure to mid-frequency words in under a half hour.

Language learners can devote even less time reading if they increase the time they spend reading from a five-day academic week to seven days a week. Thus setting aside time to read on a daily basis makes a strong argument for why language learners need to establish a daily routine of reading. Table 6, for example, converts the figures from Table 4 to show how much reading is required to increase vocabulary over a seven-day 40-week period.

Table 5. *Amount of Reading in Tokens and Amount of Time per Five-Day Week to Meet the 1,000 Word Families an Average of Twelve Times*

Frequency level	Amount of reading (word repetitions)	Amount of time @ 200 wpm	Amount of time @ 100 wpm
2nd 1,000	171,411 (13.4)	4.2 minutes	8.4 minutes
3rd 1,000	300,219 (12.6)	7.6 minutes	15.2 minutes
4th 1,000	534,697 (12.6)	13.0 minutes	26.0 minutes
5th 1,000	1,061,382 (13.7)	26.0 minutes	52.8 minutes
6th 1,000	1,450,068 (13.1)	36.0 minutes	72.0 minutes
7th 1,000	2,035,809 (13.7)	49.0 minutes	98.0 minutes
8th 1,000	2,427,807 (14.1)	60.6 minutes	121.2 minutes
9th 1,000	2,956,908 (12.0)	74.0 minutes	148.0 minutes
10th 1,000	2,956,908 (9.8)	74.0 minutes	148.0 minutes

Table 6. *Amount of Reading in Tokens and Amount of TIme per Seven-Day Week to Meet the 1,000 Word Families an Average of Twelve Times*

Frequency level	Amount of reading (word repetitions)	Amount of time @ 200 wpm	Amount of time @ 100 wpm
2nd 1,000	171,411 (13.4)	3.0 minutes	6.0 minutes
3rd 1,000	300,219 (12.6)	5.4 minutes	10.8 minutes
4th 1,000	534,697 (12.6)	9.2 minutes	18.7 minutes
5th 1,000	1,061,382 (13.7)	18.8 minutes	37.7 minutes
6th 1,000	1,450,068 (13.1)	25.7 minutes	51.4 minutes
7th 1,000	2,035,809 (13.7)	35.0 minutes	70.0 minutes
8th 1,000	2,427,807 (14.1)	43.2 minutes	86.5 minutes
9th 1,000	2,956,908 (12.0)	52.8 minutes	105.7 minutes
10th 1,000	2,956,908 (9.8)	52.8 minutes	105.7 minutes

Table 6 divides the amount of time spent reading in Table 4 by seven days. For example, in just 35 minutes, learners able to read around 200 words per minute can experience approximately 13 repetitions to mid-frequency words from the seventh 1,000 word level (4 hours and 5 minutes or 245 minutes ÷ 7 days = 35 minutes). Congruently, in 37 minutes learners who read at a slower pace of 100 words per minute can experience approximately 13 repetitions to mid-frequency words from the fifth 1,000 word level (4 hours and 24 minutes or 264 minutes ÷ 7 days = 37.7 minutes). In both examples, if learners make it a habit to read every day of the week they can experience significant exposure to mid-frequency words in just over a half hour.

Challenges and Pedagogical Implications

Although it is theoretically feasible to systematically increase vocabulary through reading, learners face some challenges in gaining sufficient exposure to new words. The first challenge is that the figures in Table 4 assume that learners are reading material where no more than 2 % of the tokens are unfamiliar. This percentage is about five unknown words for a typical 250-word page of a novel (250 x 2% = 500÷100% = 5). This assumption might be unrealistic, as a great number of words are not known by most foreign language learners. Another challenge is the actual number of word repetitions. The figures in Table 4 and the conversions thereafter in Table 5 and Table 6 are based on averages. However, several of these words might only appear one to three times in a novel. For example, for the fourth 1,000 word frequency level, 23 words appear once, 26 words appear twice, and 22 words appear three times, while at the sixth 1,000 word frequency level, 48 words appear once, 44 words appear twice, and 38 words appear three times (Nation, 2014). A third challenge is that the spacing between repetitions for some words in novels also varies. That is, the number of repetitions for words from Table 4 was based on a range of novels. However, some novels have few if any repetitions of target words compared to other novels. As a result, if there are large gaps between when learners encounter particular words, the recall of some words can fade and thus the opportunity to reinforce word meanings might be lost.

Despite the obstacles that language learners may encounter to gain sufficient exposure to new words, it is clear that mid-frequency vocabulary can be steadily enhanced as the amount of time spent reading is increased. Although the figures previously reported were based over a forty-week period and may seem like a lot of time, when broken down over the course of either a five-day or seven-day week, the actual amount of time that learners need to read is not demanding. As a result, significant exposure to mid-frequency words can be obtained if language learners set aside time to read before they go to bed. Additionally, language instructors might consider devoting reading time at the beginning of class as a warm up activity. In doing so, instructors might consider reading material written at an appropriate level for their students. However, even if language learners do not immediately have the approximate 98% vocabulary coverage necessary to read authentic texts without an external resource, they can quickly access word definitions simultaneously as they read simply by tapping on an unknown word they read in newspapers, articles and blogs displayed on a smart phone or tablet computer. Some research suggests that this type of word exposure is beneficial for learning vocabulary (Liu & Lin, 2011). Therefore, based on the research to increase the exposure to mid-frequency vocabulary in tandem with resources to learn those words, language learners should establish a habit to continuously read on a daily basis. Doing so will progressively enhance their exposure to mid-frequency words which will eventually enable them to acquire the 98% vocabulary coverage necessary to read authentic texts without an external resource such as a dictionary.

Bibliography

1) Davies, M. (2008) The Corpus of Contemporary American English: 450 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

2) Hu, M., & Nation, I. S. P. (2000). Vocabulary density and reading comprehension. *Reading in a Foreign Language, 13,* 403-430.

3) 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,* 15-30.

4) Liu, T., & Lin, P. (2011). What comes with technological convenience? Exploring the behaviors and performances of learning with computer-mediated dictionaries. *Computers in Human Behavior, 27,* 373-383.

5) McKay, S. L. & Hornberger, N. H. (Eds). (1996). Sociolinguistics and language teaching. New York, NY: Cambridge University Press.

6) Nation, I. S. P. (2006) How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63,* 59-82.

7) 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.

8) Prichard, C., & Matsumoto, Y. (2011). The effect of lexical coverage and dictionary use on L2 reading comprehension. *Reading Matrix: An International Online Journal, 11.*

9) Schmitt, N., & Schmitt, D. (2012). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching,* 1-20.

10) Schmitt, N., Jiang, X., & Grabe, W. (2011). The

percentage of words known in text and reading comprehension. *The Modern Language Journal. 95,* 26-43.
11) Waring, R., & Takaki, M. (2003). At what rate do learners learn and retain new vocabulary from reading a graded reader? Reading in a Foreign Language, 15, 130-163.
12) Webb, S. (2007b). The effects of repetition on vocabulary knowledge. Applied Linguistics, 28, 46-65.
13) West, M. (1953). *A General Service List of English Words.* London: Longman, Green & Co.

Jichi Medical University Journal 2016, 39, 51-57

中間周波数語彙を増やすための読書時間

ロバート　デイレンシュナイダー

自治医科大学　総合教育部門　英語　栃木県下野市薬師寺3311-1

要　　約

本研究では，研究者が中間語彙を増やすために必要な読み上げ時間を検る。まず，単語の頻度レベルと読解力を調べた先行研究を要約する。第二に，現代アメリカ英語（COCA）コーパスと英国コーパス（BNC）の両方の分析がレビューされている。第３に，単語の繰り返しおよび読み上げ時間による単語の露出を分析する。
最後に，未知の中間周波数語への暴露を強化することに関する課題と教育的含意について論じる。
（キーワード：Reading・Time・Vocabulary）

連絡先：ロバート　デイレンシュナイダー，自治医科大学　総合教育部門　英語，〒329-0498 栃木県下野市薬師寺3311-1，
E-mail：bdilenschneider@yahoo.com
受付：2016年９月30日，受理：2017年２月２日