The Markov Estimation of Semantic Association (MESA): Good for Vocabulary Measurement, Ineffective for Vocabulary Teaching

Article *in* THE READING MATRIX · April 2018

2 authors:

Jeff McQuillan
Center for Educational Development, Inc.

Warren Ediger
PQ

The Reading Matrix: An International Online Journal
Volume 18, Number 1, April 2018

The Markov Estimation of Semantic Association (MESA): Good for Vocabulary Measurement, Ineffective for Vocabulary Teaching

Jeff McQuillan
Center for Educational Development

Warren Ediger
Adult ESL Specialist

ABSTRACT

There is considerable evidence that incidental vocabulary acquisition through reading accounts for a large portion of the growth in word knowledge for both first (L1) and second (L2) language acquirers. In this paper, we evaluate the Markov Estimate of Semantic Association (MESA) technique for detecting small, incremental gains in vocabulary knowledge, as well as its application as a vocabulary teaching tool.

Keywords: vocabulary acquisition, vocabulary instruction, word knowledge, measurement

INTRODUCTION

Several scholars have proposed that we acquire a large proportion of our vocabulary, in both a first and second language, through reading (Krashen, 1989; Nagy, Herman, & Anderson, 1985; Smith, 2012). This knowledge is thought to come both *incrementally* (little by little) and *incidentally* (as a by-product of our main activity in reading, comprehension).

The measurement of incremental growth in partial word knowledge – which may fall short of a full definition, translation (in L2 acquisition), or acceptable synonym on a standard vocabulary assessment – is made difficult by the complexity of vocabulary knowledge. Nation (2001), for example, describes three general aspects of word knowledge – form, meaning, and use – for both receptive (reading, listening) and productive (writing, speaking) word knowledge. Each of the three general aspects is further subdivided, yielding a total of nine aspects of word knowledge. Table 1 summarizes the receptive dimension of "what is involved in knowing a word" (p. 27), according to Nation.

Table 1. What is Involved in Knowing a Word (Receptive Dimension)

Form	*Spoken*: What does the word sound like?
	Written: What does the word look like?
	Word parts: What word parts can I recognize (e.g., affixes that modify word meaning or form)?
Meaning	*Form and meaning*: What meaning does the word form suggest to me?
	Concept: What is included in the concept (oak and maple, e.g., are subsumed within the concept "tree")?
	Associations: What other words does it make me think of?
Use	*Grammatical functions*: In what patterns would I find this word?
	Collocations: What other words or word types does this word occur with?
	Constraints on use: When, where, and how often should I expect to meet this word?

From Nation (2001)

Miller (1999) also attempts to classify the kinds of things we know when we "know" a word. He notes the difficulty in trying to capture all the aspects of word knowledge, Nation's as well as several others. In particular, he emphasizes the importance of the role of "contextualization" in word knowledge. Miller defines contextualization as "a cognitive representation of the set of contexts in which a given word form can be used to express a given word meaning" (p. 13). Contextualization is what we know about how one word is related to another, the situations in which the word can be used, and the purposes and goals of the utterance in which the word appears. These relationships are primarily among word *meanings*, not word forms. They are part of our larger "cognitive structure," according to Miller, but have been little studied so far in any systematic way.

We introduce Nation's and Miller's proposed categories here to underscore the difficulty in assessing growth across a number of different aspects of word knowledge. Some researchers have attempted to tap knowledge of each aspect separately, asking questions about word class, spelling recognition, word meaning, etc. (e.g., Pellicer-Sanchez & Schmitt, 2010). Others have developed scales of knowledge "depth," such as Wesche and Paribakht's (1993) Vocabulary Knowledge Scale.

THE MESA TECHNIQUE

Frishkoff, Collins-Thompson, Perfetti, and Callan (2008) developed a unique approach to the problem of tapping incremental knowledge gains with its various components: the *Markov Estimation of Semantic Association* (MESA). MESA aims to detect incremental vocabulary growth after multiple exposures to a novel word in text. Their approach does not capture all of the aspects discussed by Miller and Nation, but does attempt to combine several of these into a single metric of word knowledge. Our purpose here is not to report on our own experience with the approach, but to discuss what MESA's strengths and weaknesses are as a research tool for the study of vocabulary, and whether it is appropriate as a teaching tool in a classroom.

In Frishkoff et al. (2008), a group of native English speakers (N = 21) were given 60 unknown words in English to read in sentences. The words were considered rare and unlikely to have been known by the subjects. Each word was seen six times in different sentences, in either useful or "good" contexts or misleading or "bad" contexts.

A "good" context would help the reader determine the meaning of the word, such the word *abrogate* in this example:

> This system has been weakened since 1983, and the current Liberal party government seeks to further weaken or *abrogate* it.

A "bad" context (so named by the researchers) was a sentence in which the target word was used in a context that was appropriate for another, similar sounding and similarly spelled distractor word. For example, the target word *abrogate* would be used in place of a distractor, *arrogate*, as a malapropism:

> Traditional distributors…*abrogate* to themselves the role of determining what's proper for their customers to read.

In a bad context, what Krashen (1999) terms a "deceptive" context, the target word is misused as a replacement for the distractor word, thus potentially misleading the reader into inferring an incorrect meaning.

Some of the target or tested words were presented to Frishkoff et al.'s subjects in all "good" contexts, some with three good and three bad contexts, and some with one bad and five good contexts. Accuracy scores were predicted to vary according to the "goodness" of the six contexts.

Subjects were given a synonym judgment test before and after the experiment, in which they were provided the 60 target words and asked to select the best synonym for each. Among the possible answers was a synonym for the distractor word, so the researchers could test just how far the bad contexts would lead readers astray. The synonym judgment test provided a pre- and post-test measure of word knowledge gain.

In addition to the synonym test, two additional measures of word knowledge were taken after each exposure to the target word in context. First, subjects also had to judge the semantic appropriateness of sample sentences that contained the target words. While not asked to give a definition, subjects had to understand enough of the word to determine if the test sentence using the target word made sense or not.

Second, subjects wrote down the meaning of the word (their best guess) after reading each sentence. These responses were then analyzed using the MESA statistical model in order to detect small, incremental growth in word knowledge after each exposure. Each guess the subjects gave as to the meaning of the target word was given a score indicating its "distance" from the correct definition.

The MESA model uses a variety of factors to determine distance, including:

- Stemming - words based on a common morphology;
- Synonymy - words of similar meaning;
- Co-occurrence - words that tend to appear together in the same context, such as "election" and "politics";
- Hypernymy and hyponymy - categorical relationships among words, e.g. *crow*, *eagle*, and *robin* are all hyponyms of the hypernym *bird*; and
- Associative strength - words one might give if doing a "free association" with the term.

If the subject's response had no possible links to a correct answer along any of the dimensions that were included in the model, knowledge was scored -1; a correct definition or synonym was scored 0. This scoring procedure allowed the generated answers to be placed on a continuous scale marking their approaching accuracy to the word's correct meaning, and thus revealing incremental growth that fell short of full knowledge of the word.

Table 2 shows how responses might be scored for the word *abditive* (meaning "hidden") presented in all "good" contexts (taken from Frishkoff et al., 2008, Example 1, p. 917).

Table 2. Example of MESA Scoring for Incremental Word Acquisition for "Abditive"

Trial	Response	MESA Score
1	Avoidant	-1.00
2	Attitude	-.83
3	Sneaky	-.50
4	Secretive	-.35
5	Secretive	-.35
6	Hidden	-.35

Notice that after each exposure up through the fourth one, the MESA scores increases – that is, it gets closer to 0, which is the score that indicates the subject fully understood the meaning of the word. The additional two "passes" or exposures to this particular word in a sentence did not improve word knowledge, since the MESA score remained the same (-.35).

Words found in more useful contexts were, not surprisingly, acquired more quickly than those placed in "bad" or misleading contexts, consistent with other evidence on the effects of context and word inference (Brusnighan & Folk, 2012; Li, 1988).

Gains Through Multiple Exposures

The average gain from pretest to post-test on the synonym test was about 12%, which means on average subjects picked up an accurate meaning of the equivalent of seven of the 60 target words after six exposures (totaling all partial knowledge the words together).

However, this 12% figure appears to have been for *all* context conditions combined, including the misleading ones. Words we encounter normally while reading would likely not appear in misleading contexts at all. We find words in helpful contexts (“good”) or neutral ones, where the other words of the sentence neither help nor hinder attempts to infer the meaning of the unknown word.

When words are presented in all good contexts, the “pickup” rate was closer to 17% (Figure 2a, p. 915). Subjects also were less likely to choose the distractor’s synonym when the word was presented in good contexts, showing another aspect of word knowledge growth.

The MESA measure confirmed that words presented in good contexts were more easily inferred than those in bad contexts. More importantly, the measure showed an incremental, mostly linear pattern of more accurate word knowledge after each “good” context exposure, indicating that each inference was getting nearer and nearer to the correct definition.

“Effort” and Spacing Effects

Frishkoff and her colleagues extended their work in a later study, Frishkoff, Perfetti, and Collins-Thompson (2011), using the MESA measure to examine gains in word knowledge in good versus neutral contexts, to test the effects of “spacing” of the exposures, and to see how much knowledge was retained of the words with a one-week delayed post-test.

Table 3 shows the pretest, immediate post-test, and delayed post-test scores for words found in all “good” and in “mixed” contexts. When the subjects (all English-speaking college students) were presented with words in all “good” contexts, the amount of partial word acquisition was impressive, with “depth of knowledge” scores increasing substantially from the first exposure to the sixth one (an increase on average of 75%). More importantly, while 16 to 19% of this knowledge was “lost” by the time of the one-week delayed post-test, most of the gains were retained (64 to 67%).

Table 3. Word Knowledge Retention on Immediate and One-Week Delayed Post-Tests in Frishkoff et al. (2011)

Context	Pretest	Post-Test	Delayed Post-Test	Loss to Delayed Post-Test	Gain Pretest to Delayed Post-Test
Good	1.15	2.38	1.93	-18.9%	67%
Mixed	1.12	2.20	1.84	-16.4%	64%

Calculated from Frischkoff et al. (2011) Figure 2. Maximum score = 3

The researchers also commented that mixing helpful and neutral contexts did not lead to more word acquisition than providing the words in all helpful contexts, as might have been predicted by theories of "effort" in language acquisition. Such theories (e.g., Schmidt & Bjork, 1992) posit that the more effort that is required to infer the meaning of the word, the more likely you are to learn and retain that word. Frishkoff and colleagues found little evidence for this. It appears that the brain does not need a "challenge" in vocabulary acquisition.

Another important finding was the failure of "spacing" in word exposure to have strong effects on word knowledge acquisition. Frishkoff and her colleagues noted that:

> [S]pacing of practice effects have been attributed to passive decay of the memory trace, which appears to explain learning (and forgetting) curves in simple, associative learning tasks (e.g., word learning from definitions; Pavlik & Anderson, 2005). By contrast, word learning from context may engage more active processes that are not captured by this simple model.

The use of the associative learning model to understand vocabulary acquisition, and to develop techniques in vocabulary teaching, is common in both first and second language reading fields. Frishkoff et al.'s (2011) data suggest, however, that exposure to words in a context (even a single sentence, as in their study) may produce qualitatively different kinds of knowledge that are not subject to the "laws" of learning and forgetting seen in studies of paired associate word learning (Smith, 1998).

Still, the exposure to the words in Frishkoff et al.'s study was relatively close together compared to other "massed" versus "distributed" (spaced) research designs, so these findings on spacing effects should be treated as preliminary. [1]

Types of Word Knowledge

It should be noted that the MESA method does not measure incidental acquisition *per se*, but rather word knowledge gain through deliberate inference. Subjects know that they will be tested on their knowledge of the words, and are asked to figure out the meaning of the word from context.

It isn't clear if this process mimics what readers do in normal reading when they encounter an unknown word. A reader may or may not stop to infer consciously the meaning of a word, a decision that no doubt depends on many factors. Some researchers (e.g., Brusnighan & Folk, 2012) appear to assume, however, that *all* increases in word knowledge must come from some combination of conscious study and deliberate inference.

"Read-and-test" incidental word acquisition studies, in which subjects are told to read a passage and are then given an unannounced vocabulary test (see Krashen, 2004, for a review), indicate that word knowledge can be gained unconsciously when readers are focused on global comprehension, forming part of what Krashen (1981) has called our "acquired" knowledge.[2]

Vocabulary knowledge can be acquired unconsciously even when subjects are not focused on language comprehension, as an experiment by Hamrick and Rebuschat (2014) found. Hamrick and Rebuschat's subjects saw a series of slides with images of objects accompanied by a set of spoken pseudo-words. Subjects assigned to the incidental condition were told to categorize the objects they saw as either animate or inanimate, with no mention made by the experimenters of the spoken pseudo-words themselves. Those in the intentional condition were told to try to memorize the meanings of the words.

After the exposure to the images and spoken words, both groups were given a vocabulary test. Results are summarized in Table 4. Subjects in both conditions were able to select the meaning of a significant number of the words above chance (25%). The incidental group knew 44.4% (SD: 7.5%) of the words; those in the intentional group scored much higher (M = 73.3%, SD 10.7%).

There was also a clear frequency effect for these gains by both groups, greater for the incidental group (r = .51) than the intentional group (r = .26), as can also be seen in Table 4 by comparing percentages correct in the High Frequency and Low Frequency conditions.

Table 4. Accuracy Percentages in Incidental and Intentional Conditions in Hamrick & Rebuschat (2014)

	Guess	**Intuition**	**Memory**	**High Frequency**	**Low Frequency**
Incidental	35.8%	48.5%	61.4%	65%	32%
Intentional	54.2%	61.9%	88.9%	85%	65%

From Hamrick & Rubuschat (2014), Table 9.2 and Figure 9.2 (p. 125). Frequency percentages estimated from Figure 9.2.

Hamrick and Rebuschat also asked their subjects to state whether each of their answers to the vocabulary test were based on "guess," "intuition," or "memory," taking "memory" as an indication of "explicit" knowledge and "intuition" and "guess" as indications of "implicit" knowledge. Accuracy scores were higher for answers that were from "memory" than from "intuition" and "guess," yet all three categories showed scores above the level of chance. Hamrick and Rebuschat concluded that exposure to novel words can lead to both "implicit" and "explicit" knowledge.

Note that there was no delayed post-test in the experiment, which may have found the gap between the incidental and intentional conditions narrow. McQuillan (2016) re-analyzed the results of several experiments involving second-language adults in which one group of subjects was exposed to unknown words through reading for meaning, and a second group received both a reading and explicit instruction treatment. Accuracy scores from the "reading plus instruction" groups typically declined much more rapidly than the "reading only" groups from immediate to delayed post-tests. Differential declines in retention are consistent with the notion that two different types of knowledge are being measured: implicit/acquired and explicit/learned (Krashen, 1981).

Further evidence of implicit or unconscious vocabulary acquisition can be found in Dabrowska (2014), who tested adult subjects on the meanings of rare words in English. Even when they said they were guessing at the word meanings, her subjects scored above chance on a multiple-choice measure, indicating that they knew more about the meanings of the words than what was available in their conscious knowledge.[3]

USING MESA AS A TEACHING TOOL

The MESA methodology represents an important advance in detecting small changes in word knowledge after multiple exposures. Frishkoff and her colleagues have shown that even when subjects are asked to produce the meaning of novel words they encounter in text (versus merely to recognize the correct meaning in a multiple-choice format, as is often done in "read-and-test" studies), there is evidence of considerable cumulative growth in knowledge, knowledge that often falls short of the "correct" answer.

More recently, however, Frishkoff and her colleagues (Frishkoff, Collins-Thompson, Hodges, & Crossley, 2016) decided to transform their MESA method from a *measurement* tool into an *instructional* method to teach vocabulary. The researchers asked a small group of adults (N = 21) attempt to infer the meanings of 60 target words, all rare words in English (e.g., *impavid*, *brio*). The words were embedded into six sentences each, for a total of 360 sentences. The sentence with the target word was written to provide either a "good" context ("high-constraint") or neutral context ("low-constraint") for inferring the meaning of the word.

Half of the words were presented in all high-constraint sentences (that is, six sentences with good context) and half were contained in "mixed" constraint sentences (three high-constraint and three low-constraint sentences). The sentences did not form a coherent text, so subjects read in effect 360 unrelated sentences.

As when used as a measurement tool, the MESA software determines how "close" the subjects are to the correct answer. Half of the subjects were given this sort of "accuracy feedback" on their answer after each guess. Feedback consisted of either "correct," "partially correct," or "wrong," depending on how near it was to the correct answer. The other subjects received no feedback at all. Note that none of the subjects received *corrective* feedback – that is, none was told the correct meaning of the word after his guess.

Frishkoff et al.'s (2008) hypothesis was that giving students feedback after each guess would help them infer the meaning of more words. Their results, however, were decidedly mixed. Due to the small sample size of the experimental and control groups, the researchers ran both a traditional ANOVA analysis as well as a nonparametric "bootstrap" analysis, in which 50,000 rounds of sampling were used. Table 5 lists the results of how well their subjects did on the both statistical approaches.

Table 5. Accuracy of Feedback Versus No-Feedback Group on Vocabulary Measure in Frishkoff et al. (2017)

	Multiple Choice Pretest/Post-test	**MESA Score Gains**	**Trial 3 to 4 MESA**
ANOVA	Not Significant	Not Significant	Not Significant
Bootstrap	*Significant*	Not Significant	Not Significant

The multiple-choice measure was a comparison of pretest and post-test accuracy scores on a synonym judgment test. Interestingly, the subjects did know some words before beginning the experiment, scoring 32% overall, which was 12% above the chance level of 20%.

The ANOVA test for the multiple-choice measure was not significant for the feedback condition, but was in the bootstrap procedure. The MESA Score Gain measured the increase in MESA scores from the first to the sixth exposure. Here there was no significant difference either in the ANOVA or bootstrap statistical tests. Finally, the researchers tested the difference between the MESA scores on Trials 3 and 4, since this was the point at which the "mixed" sentence condition changed from sentences with helpful contexts to neutral ones. Again, there was no significant difference between the feedback and no-feedback groups on either statistical test.

The researchers concluded that "these findings show that real-time Feedback can improve performance . . . and can thereby lead to improved word learning from context" (p. 18, online version). But with only one significant finding in the six statistical tests they conducted, the results of their experiment do not give us a great deal of confidence in that conclusion. As a practical matter, the difference in post-test accuracy score gains between the feedback and no-feedback groups was not dramatic: The feedback group made a gain of 40%, while the no-feedback group gained 32%. Readers gained a good deal without the use of the software.

In the classroom, the use of the MESA software would obviously require considerable investment of time, both for teachers and students. A much more reliable, and cheaper, alternative to improve vocabulary is simply to read. Nation (2014) estimated that reading about an hour a day could move you from the 5,000-word-family level in English to the 9,000-word-family level in a little over two years. This is enough to read most texts written for adults in English, and no special software is required. Voluminous reading has numerous benefits beyond merely improving vocabulary (Krashen, 2004), and is much more pleasant than reading a series of unrelated sentences on a computer screen.

SUMMARY

No single measure is likely to capture what we know when we acquire a new word, and different theories of word acquisition have led to the development of different assessments. Nation (2001) and Miller (1999) highlighted the difficulty of the effort, but also indicated the variables that should be taken into account. Added to the problem of assessing vocabulary knowledge is the issue of *partial* word acquisition, since while "fast mapping" can account for knowledge of a word's meaning even with a single exposure (Carey & Bartlett, 1978), the full meaning of words usually requires several exposures.

The MESA technique is one attempt to integrate various sources of knowledge into a single metric of "knowing" word meanings. Its use of several of the factors involved in word acquisition, combined with multiple exposures to a word under differing conditions, gives us a more complete picture of how well a word is known and the incremental increases of knowledge after each exposure. As a measurement tool of what is consciously learned through inference, then, MESA holds promise as a way of tapping into aspects of word knowledge that up to this point required a good deal more effort to measure. Frishkoff and colleagues have provided good evidence that their tool measures small changes of word knowledge, and that it does so in a more comprehensive way than alternative approaches.

The evidence of using the MESA model to teach vocabulary, however, is much less impressive. The attempt by Frishkoff, Collins-Thompson, Hodges, and Crossley (2016) to use MESA to teach rare word meanings through the use of a "feedback" condition resulted in only marginal improvements, with only weak evidence for its effectiveness. We have much stronger evidence showing that engaging in extensive reading is by far the most effective means of improving vocabulary knowledge (see Krashen & Mason, 2017, for a review).

ENDNOTES

1. Krashen (2004) notes that, if "distributed" spacing turns out to be an important element in word acquisition, normal reading provides that already: "Encountering words in natural text provides, of course, distributed exposure to vocabulary" (p. 48).

2. Stahl (1990) also points out that there is some evidence that incidental acquisition is unaffected by reading ability – that is, both good and poor readers acquire words from text at about the same rate, when faced with texts of equal relative difficulty. Studies in which readers are asked to infer the meaning of words from context, however, typically *do* see an effect of reading ability (e.g., Cain, Oakhill, & Elbro, 2003). If Stahl is correct, this is another indication that the two tasks are tapping different types of word knowledge. Swanborn and de Glopper (1999) also found that reading ability did not explain any additional variance in incidental word acquisition rates once grade level was controlled for. This is consistent with other evidence that older acquirers are generally faster than younger ones (Krashen, Long, & Scarcella, 1979).

3. Dabrowska suggests that words pass from an implicit to an explicit stage, depending on the strength of their "entrenchment" (p. 13). An alternative hypothesis is that all word knowledge is acquired implicitly, quite apart from our ability to make that knowledge explicit when called upon to do so.

Jeff McQuillan, Ph.D., is a Senior Research Associate at the Center for Educational Development in Los Angeles, California.

Email to jeff@eslpod.com.

Warren Ediger is a teacher and writer living in La Habra, California.

REFERENCES

Brusnighan, S. M., & Folk, J. R. (2012). Combining contextual and morphemic cues is beneficial during incidental vocabulary acquisition: Semantic transparency in novel compound word processing. *Reading Research Quarterly*, *47*(2), 172-190.

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language*, *30*(3), 681-694.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference*, *15*, 17–29.

Dabrowski, E. (2014). Implicit lexical knowledge. *Linguistics*, *52*(1), 205-223.

Frishkoff, G. A., Perfetti, C., & Collins-Thompson, K. (2011). Predicting robust vocabulary growth from measures of incremental learning. *Scientific Studies of Reading*, *15*(1), 71-91.

Frishkoff, G., Collins-Thompson, K., Perfetti, C., and Callan., J. (2008). Measuring incremental changes in word knowledge: Experimental validation and implications for learning and assessment. *Behavior Research Methods*, *40*(4), 907-925.

Frishkoff, G., Collins-Thompson, K., Hodges, L., & Crossley, S. (2016). Accuracy feedback improves word learning from context: Evidence from a meaning-generation task. *Reading and Writing, 29*(4), 609-632.

Hamrick, P., & Rebuschat, P. (2014). Frequency effects, learning conditions, and the development of implicit and explicit lexical knowledge. In J. Connor-Linton & L.W. Amoroso (Eds.), *Measured Language: Quantitative Studies of Acquisition, Assessment, and Variation* (125-139). Washington, D.C.: Georgetown University Press.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford University Press.

Krashen, S. (1989). We acquire vocabulary and spelling by reading: Additional evidence for the Input Hypothesis. *Modern Language Journal, 73*(4), 440-464.

Krashen, S. (1999). *Three Arguments Against Whole Language & Why They Are Wrong*. Portsmouth, NH: Heinemann.

Krashen, S. (2004). *The Power of Reading: Insights from the Research*. 2nd edition. Portsmouth, NH: Heinemann.

Krashen, S., Long, M., & Scarcella, R. (1979). Age, rate, and eventual attainment in second language acquisition. *TESOL Quarterly, 11*, 573-582.

Krashen, S. & Mason, B. (2017). Sustained silent reading in foreign language education: An update. *Turkish Online Journal of English Language Teaching (TOJELT), 2*(2), 70-73.

Li, X. (1988). Effects of contextual cues on inferring and remembering meanings of new words. *Applied linguistics, 9*(4), 402-413.

McQuillan, J. (2016). Time, texts, and teaching in vocabulary acquisition: A rebuttal to Cobb (2016). *Reading in a Foreign Language, 28*(2), 307-318.

Miller, G. (1999). On knowing a word. *Annual Review of Psychology*, 50, 1-19.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly*, 233-253.

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

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*(2), 1-16.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do *Things Fall Apart*? *Reading in a Foreign Language, 22*(1), 31-55.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Smith, F. (1998). *The Book of Learning and Forgetting*. New York: Teachers College Press.

Smith, F. (2012). *Understanding Reading: A Psycholinguistic Analysis of Reading and Learning to Reading*. 6th edition. New York: Routledge.

Swanborn, M., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261-285.

Stahl, S. A. (1990). *Beyond the instrumentalist hypothesis: Some relationships between word meanings and comprehension.* Center for the Study of Reading Technical Reports, No. 505. Champaign, IL: University of Illinois, Urbana-Champaign .

Wesche, M., & Paribakht, T. (1993). Assessing second language vocabulary knowledge: Depth versus breadth. *Canadian Modern Language Review, 53(*1), 11-40.