ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/332514442

Estimating Input Quantity for L2 Vocabulary Acquisition: A Preliminary Study of Statistical Language Analysis

Article · February 2017

DOI: 10.32234/jacetjournal.61.0_10

CITATIONS
0

READS
22

1 author:

Akira Hamada
Meikai University

18 PUBLICATIONS **14** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project Lexical Processing and Memory Representations in Second Language Vocabulary Acquisition View project

Project Cognition and Education in Second Language Reading View project

All content following this page was uploaded by Akira Hamada on 18 April 2019.

The user has requested enhancement of the downloaded file.

JACET Journal 61 (2017) 109–129

Estimating Input Quantity for L2 Vocabulary Acquisition: A Preliminary Study of Statistical Language Analysis

HAMADA, Akira
Nihon University

Abstract

Acquisition of second language vocabulary requires such a large amount of input that determining how frequently new words should be introduced is important for teachers. Results from an experimental approach could not identify the frequency of input required for learners to efficiently acquire target words because of uncontrollable linguistic and individual factors. A corpus-based approach only showed whether representative words to be learned were included in certain textbooks or not. A new approach applies a computation model based on statistical language analysis, which can be a powerful tool to simulate learners' vocabulary growth from textual input. This preliminary study used models of reading-oriented word knowledge that simulated learners' vocabulary growth based on Latent Semantic Analysis. The knowledge models were assessed using Eiken vocabulary test items of varying difficulty. Latent growth curve modeling demonstrated that response accuracy increased as a result of incremental textual input and varied depending on test item difficulty. In addition, the product-moment correlation coefficients between the knowledge models provided evidence of substantial vocabulary growth from an unsophisticated vocabulary to one similar to adult knowledge. These findings indicate the applicability of statistical language analysis in estimating the input quantity needed for vocabulary acquisition.

***Keywords*:** vocabulary, corpus, latent semantic analysis, knowledge model, latent growth curve modeling

Introduction

Input Quantity Needed for Vocabulary Acquisition From Text

How much input is necessary for learners to acquire new words of a second language (L2)? When designing a textbook for vocabulary learning, educators run into difficulties determining which words should be introduced, how, and at what frequency. New words that students will encounter in real-life situations are often presented in a contextualized way. For example, textbooks typically provide students with lists of new words (e.g., *charge*, *reservation*, and *vacancy*) that are related to a certain topic or occur in a particular situation (e.g., "hotel booking dialogue"). Language teachers then decide how to input those words in a class, using exercise-based or task-based activities.

Determining how many times new words should appear in textbooks is another problematic issue. An experimental approach has attempted to gain insight into L2 word learning from textual input, especially concerning the factors that contribute to vocabulary growth. One of the operationalized variables often addressed in literature is the *frequency* of input exposure. For example, research on extensive reading has shown the relationship between input frequency and acquisition rates for target words (e.g., Horst, Cobb, & Meara, 1998; Pellicer-Sánchez & Schmitt, 2010; Waring & Takaki, 2003). More controlled experiments have tried to determine the optimal frequency of target word occurrences in a text by manipulating the type of context (Webb, 2008) and the type of words (Chen & Truscott, 2010). Both kinds of studies used the same type of meaning recall tests; unfortunately, the number of exposures to target words needed to exceed 50% recall probability varied wildly, ranging from as lows as three to 20 and more (Reynolds & Wible, 2014). Thus, although previous research has shown that multiple exposures to target words contribute to the development of L2 vocabulary knowledge, the threshold of the optimal frequency remains unknown.

Pellicer-Sánchez and Schmitt (2010) suggested that research designed to explore incremental word learning from multiple exposures in long L2 texts under naturalistic conditions was not sophisticated enough. In contrast, Elgort and Warren (2014) used a mixed-effect model, which took into account learner variables (e.g., age, proficiency, learning strategies) and target item variables (e.g., concreteness, frequency, saliency of use), to examine the complicated relationships between the variables that affect L2 vocabulary learning from textual input. Their results demonstrated that the required frequency of target word occurrences was modulated by learners' reading and lexical proficiency. However, their finding that "[the] number of encounters with a word needed for learning was higher for less proficient learners" (p. 396) was similar to the findings of prior studies and not useful for estimating the required quantity of input.

Another frequently used way of designing textbooks is a corpus-based approach. Chujo (2015) emphasized the advantage of corpus compilation from textbooks because it allows us to quantitatively analyze their contents. For example, corpus-based research can determine whether a particular textbook exposes L2 learners to high-frequency words or not and how many instances of those words are included. Comparisons of various aspects of the texts can give us useful information such as the number, variability, and difficulty of words (Ishikawa, 2008). In other words, researchers have examined whether particular textbooks consist of enough representative words to achieve particular goals of language teaching.

Although corpus-based findings can determine the words to be learned, this approach does not seem interested in whether learners acquire knowledge of new words at what frequency of input. For example, Muraoka (2010) counted the frequency of recurring word types in the retired *SUNSHINE English Course 1-3*, showing that only 32% of words appeared five times or more, while 68% of words appeared less than five times. However, it was not reported whether Japanese junior high school students acquired the frequently repeated words from the textbooks more than the infrequently repeated ones. Because the highest frequency words reported in her study were function words such as articles, pronouns, and prepositions, which are difficult to acquire (Nation, 2013), a simple frequency analysis could not be indicative of the learners' vocabulary acquisition (Reynolds & Wible, 2014).

Synthesizing the results from all these studies, it is clear that both approaches have limitations specific to their research methodology when it comes to estimating the optimal input quantity for L2 vocabulary acquisition. The present study uses a different approach, providing a method for simulating L2 learners' vocabulary growth from textual input.

A Statistical Language Analysis Approach

Statistical modeling of word knowledge development. Landauer, Kireyev, and Panaccione (2011) have developed a computer simulation method to estimate the development of reading-oriented word knowledge. This method uses a model of statistical language analysis called *Latent Semantic Analysis* (LSA). It is an original mathematical and statistical technique for extracting semantic relations from the contextual usage of words in discourse, and is used to "create learning trajectories for each unique orthographic word-form" (Landauer et al., 2011, p. 92). A follow-up study proved that LSA computation could simulate, with high accuracy, the ages at which learners acquired word meanings from each age level corpus (Biemiller, Rosenstein, Sparks, Landauer, & Foltz, 2014).

The learning mechanism behind LSA is related to the usage-based model of language learning (Ellis, 2002; Inohara & Kusumi, 2011). At its core is the observation that our knowledge of a word emerges in memory from multiple exposures to a significant amount of information about its usage in different contexts. With respect to this, Landauer and Dumais (1997) demonstrated that, by using LSA computation, a large amount of vocabulary knowledge was derived from the effects of exposure to different texts, instead of from learning separate word meanings. In other words, the knowledge of a certain form-meaning pair is a reflection of the accumulated and abstracted experiences from repeated exposures to particular expressions.

The LSA theory is based on two assumptions about usage-based language learning, the principles of direct and indirect co-occurrence. The first principle states that words co-occurring in the same context share similar semantic properties (Landauer, Foltz, & Laham, 1998). For example, in the sentence *The dog jumped up and licked his face*, the target word *lick* co-occurs with other content words such as *dog*, *jump*, and *face*. The usage-based model suggests that, when learners frequently process the word *lick* in these kinds of contexts, they implicitly acquire its contextual-usage meaning (Ellis, 2002). The second principle arises to resolve a problem inherent in the first one, namely, that not all word meanings can be learned from direct input. For example, the target word *lick* is not in the sentence *Her little puppy grew up to be a big dog*. Although there must be a semantic relation between *lick* and *puppy*, the direct relations alone are insufficient to capture the mechanism of language learning. Landauer and Dumais (1997) claimed that learning of a certain word is elaborated from the indirect semantic relations between words by a process of induction. In LSA, the semantic similarity between *lick* and *puppy* is strengthened by the indirect co-occurrences through the mediation of *dog* in order to reflect the induction process of vocabulary learning.

The learning algorithm of LSA. To explain the learning algorithm of LSA based on Landauer et al. (1998), I will present an LSA example using a small corpus compiled from eight titles from *JACET Journal* on English vocabulary learning and writing, as listed in Figure 1. Content words used in at least two of the titles were analyzed. The main three phases of the analysis are word-by-document matrix creation, dimensionality reduction by singular value decomposition (SVD), and semantic similarity calculation (Inohara & Kusumi, 2011, 2012; Toyoda, 2008).

First, the corpus used as the textual input was transformed into word-by-document matrix {*X*}. Each cell denotes the frequency at which each word appears in each title. In this case, *learning* does not co-occur with either *English* or *test* in the same title. Therefore, Spearman's rank correlation coefficient is –.33 between *learning* and *English*. The value of the coefficient is the same between *learning* and *test*, which seems counterintuitive. According to Landauer et al. (1998), the present matrix is too rich with information to extract the true pattern of co-occurrences and semantic relations between the words in the titles.

Example of text data: Eight titles on English education in Japan

v1: Assessing the dimensionality of three hypothesized sub-skills of L2 *vocabulary* proficiency
v2: *Japanese* EFL learners' *vocabulary learning* strategies from the perspective of word frequency
v3: Estimating *vocabulary size*: Does *test* format make a difference?
v4: Development and validation of the PC version of the Mochizuki *vocabulary size test*
w1: A scaffolded *English writing course* for *Japanese university students*
w2: *English writing* in Japan: Toward integration
w3: The *learning* outcomes of an academic *writing course*: A study of *Japanese university students*
w4: Developing a *writing* rubric for classroom use in *Japanese* higher education

{*X*} =

	v1	v2	v3	v4	w1	w2	w3	w4
vocabulary	1	1	1	1	0	0	0	0
Japanese	0	1	0	0	1	0	1	1
learning	0	1	0	0	0	0	1	0
size	0	0	1	1	0	0	0	0
test	0	0	1	1	0	0	0	0
course	0	0	0	0	1	0	1	0
English	0	0	0	0	1	1	0	0
student	0	0	0	0	1	0	1	0
university	0	0	0	0	1	0	1	0
writing	0	0	0	0	1	1	1	1

Figure 1. A word-by-document matrix, {*X*}, formed from the titles of four articles about vocabulary (v1-v4) and four articles about writing (w1-w4) from *JACET Journal*. The words in italics are content words that are used in at least two of the titles.

To abstract superficial information and capture the latent semantics, SVD is applied to the matrix, "in which each cell frequency is weighted by a function that expresses both the word's importance in the particular passage and the degree to which the word type carries information in the domain of discourse" (Landauer et al., 1998, p. 263). In SVD, an $m \times n$ matrix is decomposed into three matrices, as shown in Figure 2. According to Quesada (2007), two matrices {*U*} and {*V**} describe the original row and column entities as orthogonal vectors, one representing the semantics of words (i.e., left singular vector) and one representing the contexts (i.e., right singular vector). The third is a diagonal matrix {*D*} of a singular value, the words' importance. This singular value is used to reduce dimensionality to an optional dimension, which creates an approximate matrix of the original. In this example, I used the first two values that explained 54% of the sum of the singular values, shown in the shaded columns of the three matrices in Figure 2, for dimensionality reduction from eight to two dimensions. Figure 3 displays the reconstructed two-dimensional approximate matrix {*X'*}. Every value in each cell represents how well the corresponding word can contribute to expressing its context like an eigenvalue in a factor analysis (Toyoda, 2008).

$\{X\} = \{U\}\{D\}\{V^*\}$

$\{U_{10\times8}\} =$

0.09	–0.71	0.17	–0.22	0.48	–0.22	–0.38	0.00
0.50	–0.05	0.36	–0.29	–0.30	–0.49	0.45	0.00
0.24	–0.10	0.56	0.01	0.07	0.77	0.15	0.00
0.02	–0.48	–0.28	0.20	–0.31	0.12	0.20	0.68
0.02	–0.48	–0.28	0.20	–0.31	0.12	0.20	–0.68
0.36	0.07	–0.04	0.41	0.13	–0.08	–0.08	–0.21
0.23	0.07	–0.51	–0.31	0.56	0.16	0.50	0.00
0.36	0.07	–0.04	0.41	0.13	–0.08	–0.08	0.10
0.36	0.07	–0.04	0.41	0.13	–0.08	–0.08	0.10
0.49	0.10	–0.32	–0.43	–0.35	0.24	–0.53	0.00

$\{D_{8\text{-dimension}}\} =$

3.59	2.63	–1.67	–1.32	–0.96	–0.68	–0.51	–0.00

$\{V^*_{8\times8}\} =$

0.03	–0.27	0.10	–0.17	0.50	–0.32	–0.78	0.00
0.23	–0.33	0.65	–0.38	0.26	0.09	0.44	0.00
0.04	–0.63	–0.24	0.13	–0.15	0.04	0.05	–0.71
0.04	–0.63	–0.24	0.13	–0.15	0.04	0.05	0.71
0.64	0.12	–0.35	0.15	0.31	–0.48	0.32	0.00
0.20	0.06	–0.50	–0.56	0.22	0.58	–0.06	0.00
0.64	0.06	0.28	0.39	–0.19	0.43	–0.36	0.00
0.28	0.02	0.02	–0.55	–0.68	–0.37	–0.17	0.00

Figure 2. Complete SVD of matrix {*X*}.

$\{X'\} =$

	v1	v2	v3	v4	w1	w2	w3	w4
vocabulary	0.51	0.68	1.19	1.19	–0.01	–0.05	0.11	0.06
Japanese	0.08	0.46	0.15	0.15	1.13	0.35	1.15	0.49
learning	0.10	0.29	0.20	0.20	0.53	0.16	0.55	0.24
size	0.34	0.43	0.80	0.80	–0.10	–0.07	–0.02	0.00
test	0.34	0.43	0.80	0.80	–0.10	–0.07	–0.02	0.00
course	–0.01	0.24	–0.06	–0.06	0.84	0.27	0.84	0.36
English	–0.03	0.14	–0.08	–0.08	0.56	0.18	0.55	0.24
student	–0.01	0.24	–0.06	–0.06	0.84	0.27	0.84	0.36
university	–0.01	0.24	–0.06	–0.06	0.84	0.27	0.84	0.36
writing	–0.02	0.33	–0.10	–0.10	1.16	0.37	1.15	0.49

Figure 3. Two-dimensional reconstructed matrix of its original, {*X*}.

Finally, the strength of semantic similarities between concepts described by words and the contexts in which they appear is calculated. It is represented as the cosine of the angle formed by two vectors (hereafter, *LSA value*). The LSA value can range from –1.00 to 1.00, and semantic similarity strengthens as the value approaches 1.00 because the angle formed by the two vectors approaches zero. For example, the LSA value between *learning* and *English* was .87, and it was .35 between *learning* and *test*[1] (see Figure 4). Thus, even though these words do not appear in the same context, LSA is able to determine that the terms *learning* and *English* can co-occur in the context of English education in Japan. This follows from the second principle of the inductive process of language learning in LSA (Inohara & Kusumi, 2011), according to which vocabulary gained from text comprehension consists of knowledge about words partly induced from words that do not appear in the same types of discourse (Landauer et al., 1998).

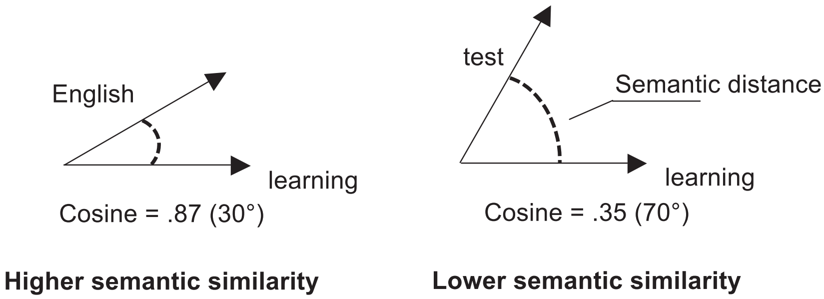

Figure 4. A schematic image of differences in semantic similarity between two pairs of words, as computed by LSA. Each arrow represents a vector of the corresponding word's meaning.

Estimation of input quantity for vocabulary acquisition. Using LSA to simulate learners' vocabulary growth has been an interesting approach in the field of vocabulary acquisition research. Although the focus of these studies has not been on determining the optimal quantity of input needed for vocabulary acquisition, they provide validity to applying LSA in the simulation of how incremental textual input advances learners' vocabulary growth. Crossley, Salsbury, McCarthy, and McNamara (2008), for example, used LSA to assess the quality of L2 lexical knowledge in terms of the ability to produce semantically coherent speech between text segments. In their study, spoken data from L2 learners were gathered over the course of a year. A significant increase in the LSA values between adjacent speech data reflected the participants' L2 lexical growth because of longer-term exposure to input. Moreover, Hamada (2014, 2015) demonstrated that LSA could manipulate the learning outcomes of contextualized L2 vocabulary learning. The participants in his experiments gained knowledge of new word meanings and usage based on information about how semantically similar target words were to contextual messages.

In the area of research on first language (L1) acquisition, Inohara and Kusumi (2012) applied the LSA theory to predict the effects of reading habits on L1 vocabulary growth. They tested whether the response patterns of a word association test can be simulated by LSA computation from either a newspaper-based corpus or a novel-based one. If the words input into a mental lexicon are classified as either newspaper-based or novel-based knowledge by the participants' reading habits, the word association patterns should differ according to co-occurrences of words stored in their mental lexicon (e.g., *refrigerator* → *appliances* in newspaper-based knowledge vs. *refrigerator → open* in novel-based knowledge). Their results showed that it is possible for LSA to simulate the causal relationship between reading habits and the word knowledge acquired from the types of text the participants read.

These prior studies demonstrated that vocabulary development and the representation of words in the mind were consistent with the LSA theory. However, their experiments were not designed to estimate the optimal input quantity for L2 learners to acquire word knowledge from text. Note that the present study does not aim to answer how many times a particular word is required to input for its acquisition because a research environment to achieve this goal was not arranged (see Conclusion for details). Instead, the present study was designed to demonstrate the applicability of the LSA theory to estimating the optimal input quantity for L2 vocabulary acquisition from a particular corpus.

The Present Study

The statistical models of vocabulary knowledge constructed by LSA using the General Reading Space corpus were tested by the vocabulary sections of the Eiken Test in Practical English Proficiency, the multiple-choice format sentence completion items in particular. This format can assess the knowledge of words in context (Nation, 2013) or the contextualized vocabulary knowledge that can be quantified by LSA. In this test, test-takers select from four options the most appropriate word to complete a given text or dialogue, which requires them to use contextual information to provide a correct answer (Morimoto, 2006).

The use of Eiken vocabulary test items should be advantageous to evaluating the development of contextualized vocabulary knowledge because they are synchronized with the English teaching curriculum in Japan. In other words, the test grades (e.g., 2nd grade) are considered equal to corresponding school years (e.g., Japanese high school graduates). This helps combine the Japanese EFL learners' knowledge model constructed by LSA with test performances in future research. At this time, it should be noted that the knowledge models constructed from the General Reading Space corpus are inconsistent with Japanese EFL learners' knowledge because the contents of the corpus are not similar to the English learning environment in Japan. Nevertheless, it is important to examine how accurately the LSA knowledge model can select the correct answers in the Eiken vocabulary tests as a result of language acquisition from textual input.

The present study employed LSA computation when quantifying the strength of semantic similarity between item stems (e.g., *I usually get off the bus at the next* ––) and each alternative (1. *box* [–.03]; 2. *car* [–.04]; 3. *cup* [.00]; 4. *stop* [.28]). In this example, a corpus compiled specifically from texts for the 3rd grade readers was used for the calculation, resulting in the knowledge model regarding the word *stop* as the most appropriate one to complete the sentence from the viewpoint of the semantic similarity between the contextual message and the word meaning. In this way, answer patterns were simulated using LSA knowledge models constructed from 3rd, 6th, 9th, 12th, and College grade corpora and the pre-1st, 2nd, pre-2nd, 3rd, and 4th grade Eiken vocabulary test items. To assess how well the LSA knowledge models were able to explain incremental vocabulary acquisition from textual input, three simulations were conducted related to the following research questions (RQs):

RQ1: Does incremental textual input increase response accuracy in the Eiken vocabulary tests?

RQ2: Does incremental textual input increase semantic similarity between the item stems and correct responses?

RQ3: Does incremental textual input improve the lower graders' vocabulary knowledge to the level of the College graders' knowledge?

Method

Materials

Textual input. The corpora used for the current LSA knowledge models were "General Reading Space," constructed from a variety of texts, novels, newspapers, and other written information. These were randomly extracted from *The Educator's Word Frequency Guide* by the Touchstone Applied Science Associates, Inc. corpus, as described in Dennis (2007). The corpora were individually tagged with their grade level (3rd, 6th, 9th, 12th, or College grade), which was determined by the readability score of each document. It is important to note that they are cumulative corpora, i.e., that higher grade content includes all the lower grade content (see Dennis, 2007, p. 70, for details).[2] This compilation of the corpora aims to approximate the order of text encounters by the learners in each grade (Landauer et al., 2011).

Eiken vocabulary tests. The present study used the Eiken vocabulary test items administered from 2011 to 2015 (except for the third trial in 2015) from the pre-1st (GP1), 2nd (G2), pre-2nd (GP2), 3rd (G3), and 4th (G4) grades. In the test, 15 to 25 short texts (one or two sentences and dialogues long) are presented, in which a target word or phrase has been omitted. The word frequency in each item stem was almost equal across the levels of test difficulty because the text coverage of 3,000 high-frequency words based on the *JACET 8000* list of 8,000 basic words (JACET, 2003) ranged from 95% to 98%. In contrast, the frequency levels of target words become lower according to the test difficulty (e.g., 4,000-level words were 11% in GP1, 6% in G2, 3% in GP2, 1% in G3, and 0% in G4).[3]

Procedure

All applications of LSA were based on the LSA web site (http://lsa.colorado.edu/) built by the University of Colorado. The strength of semantic similarities between the item stems and each of the options was calculated separately for each of the five grade knowledge models. Following Landauer and Dumais (1997), the option with the highest LSA value was considered to be each knowledge model's answer. Let us consider the following example:

After the TV station bought new equipment for its weather department, the — of its forecasts improved. Now, it makes fewer mistakes when reporting the weather.

3rd	1. accuracy (n/a)	2. discovery (–.01)	3. gravity (.03)	4. prosperity (.14)
6th	1. accuracy (–.04)	2. discovery (.05)	3. gravity (–.03)	4. prosperity (.07)
9th	1. accuracy (.02)	2. discovery (.06)	3. gravity (.00)	4. prosperity (.03)
12th	1. accuracy (.07)	2. discovery (.05)	3. gravity (–.05)	4. prosperity (–.03)
College	1. accuracy (.16)	2. discovery (.04)	3. gravity (–.02)	4. prosperity (–.02)

The four alternatives (*accuracy*, *discovery*, *gravity*, and *prosperity*) had different degrees of semantic similarity with the item stem across different knowledge models. In this item, the word *prosperity* had the highest LSA value in the 3rd and 6th grade knowledge models; therefore, the answer simulated by LSA was option 4 (underlined). In the same way, the 9th grade knowledge model selected option 2, and the 12th and College grade knowledge models selected option 1.

The semantic similarities (i.e., LSA values) between the item stems and each option were also recorded. Even if the focal knowledge model cannot select the correct response, a significant increase in the LSA values indicates the development of L2 word knowledge because of incremental textual input (Crossley et al., 2008). In the above example, the LSA value for *accuracy* increased from –.04 in the 6th grade knowledge model to .16 in the College grade knowledge model. This suggests that the knowledge development resulting from incremental textual input sophisticated the accuracy of judging the semantic similarities in the test.

The mark *n/a* means that the option (in this case, *accuracy*) was not included in the focal corpus for each knowledge model. Additionally, LSA values can be unintended if any words in the stems are missing from the corpora (Dennis, 2007). In the above example, the word *reporting* was not included in the 3rd grade corpus. Nevertheless, the value was accepted because it is common for the item stem and the options in a vocabulary test item to include words unknown to test-takers. For practical reasons, contracted forms were manually restored to their originals (e.g., *isn't → is not*) because the LSA web program did not incorporate such automatic parsing.

Data Analysis

The mean correct response rates and LSA values were calculated from five items and counted as one data point in order to obtain a large enough sample size for Structural Equation Modeling (SEM). For example, the 25 vocabulary test items of Eiken GP1 were grouped as follows: items 1 to 5, 6 to 10, 11 to 15, 16 to 20, and 21 to 25 (GP1, k = 70; G2, k = 56; GP2, k = 56; G3, k = 42; G4, k = 42). Statistical analyses were conducted in R-3.2.4.

Latent growth curve modeling, a type of SEM, was used to examine whether the response accuracy and the strength of semantic similarities between stems and correct answers increased as the input quantity increased (RQs 1 and 2). Two latent variables, SLOPE and INTERCEPT, represent the latent growth of vocabulary knowledge as assessed by the Eiken test. The observed variables were the mean correct response rates, the LSA values, and the test grades (see Figure 6 for details). The results are interpreted as follows (see Toyoda, 2014 for review):

- *Unstandardized Estimates of SLOPE* show the mean growth of the response accuracy and the LSA values from 3rd to College grade knowledge models. Positive estimates mean that those scores increase as textual input is added to the knowledge models.
- *Variances of SLOPE/INTERCEPT* show the individual differences of the response accuracy and the LSA values at the growth rates and at the beginning of measurement (i.e., 3rd grade), respectively.

- *Unstandardized Path Coefficients (Test grade → SLOPE/INTERCEPT)* show whether individual differences in the latent growth of vocabulary knowledge can be explained in terms of test difficulty.
- *Correlations between SLOPE and INTERCEPT* show the relationship between the latent growth of vocabulary knowledge and its individual differences. In this study, positive correlation indicates that, when the response accuracy and the LSA values are high in 3rd grade, the latent growth rates of its vocabulary knowledge are also high.

To answer RQ3, Pearson product-moment correlation coefficients of the mean correct response rates and the LSA values between College grade and the lower grades were calculated to test how much the vocabulary knowledge approached the College grade knowledge model.

Results and Discussion

The Latent Growth of the Response Accuracy (RQ1)

Figure 5 summarizes the increase and distribution of the response accuracy for each test grade across the five knowledge models. Their numerical details are in Table 1. Five requisites for interpreting SEM results (normality, parameter estimation methods, model fit indices, missing data treatment, and sample size) were examined based on Toyoda (2014). First, univariate skewness (range = –0.16 to 0.25), kurtosis (range = –1.22 to –0.29), and multivariate kurtosis (0.62) were extremely close to zero (Zs < 1.96), ensuring the validity of the data normality assumption to use maximum likelihood estimation. Some model fit indices satisfied the guidelines: CFI = .95, TLI = .89, SRMR = .04; others did not: $\chi^2(7) = 45.84$, $p < .001$; RMSEA = .15, 90% CI [.11, .19]. The data do not include any missing values, and the sample size (k = 266) should be sufficiently large. No multicollinearities among the observed variables were found (r range = .18 to .79 [< .90]). Figure 6 depicts the results of the latent growth curve modeling for response accuracy.

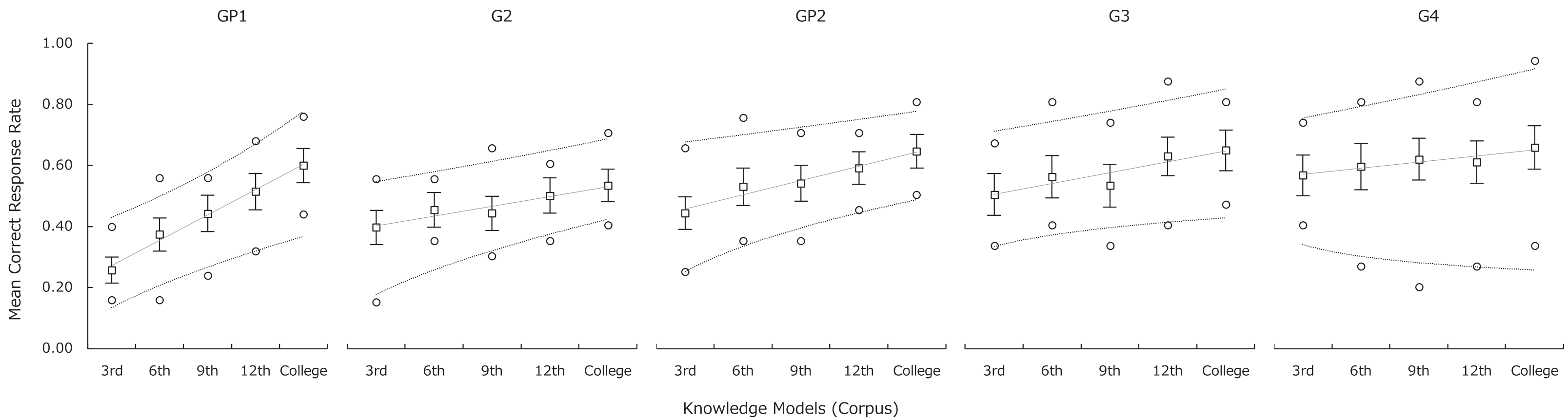

Figure 5. A growth curve of response accuracy. The squares and circles (upper/lower) represent the mean correct response rates and maximum/minimum values, respectively; the whiskers represent the 95% CIs of the mean correct response rates.

Table 1

Mean Correct Response Rates With 95% CIs and Standard Deviations of the Eiken Vocabulary Test

Knowledge	GP1 (k = 70)			G2 (k = 56)			GP2 (k = 56)			G3 (k = 42)			G4 (k = 42)		
models	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
3rd grade	.26	[.21, .30]	.18	.39	[.34, .45]	.21	.44	[.38, .49]	.20	.50	[.43, .57]	.23	.56	[.49, .63]	.22
6th grade	.37	[.32, .43]	.23	.45	[.39, .51]	.22	.53	[.46, .59]	.23	.56	[.49, .63]	.23	.59	[.51, .67]	.25
9th grade	.44	[.38, .50]	.25	.44	[.38, .50]	.21	.54	[.48, .60]	.22	.53	[.46, .60]	.23	.61	[.51, .67]	.23
12th grade	.51	[.45, .58]	.26	.50	[.44, .56]	.22	.59	[.53, .64]	.20	.62	[.56, .69]	.21	.60	[.53, .68]	.23
College grade	.60	[.54, .66]	.24	.53	[.47, .58]	.20	.64	[.58, .70]	.21	.64	[.57, .71]	.22	.65	[.58, .73]	.23

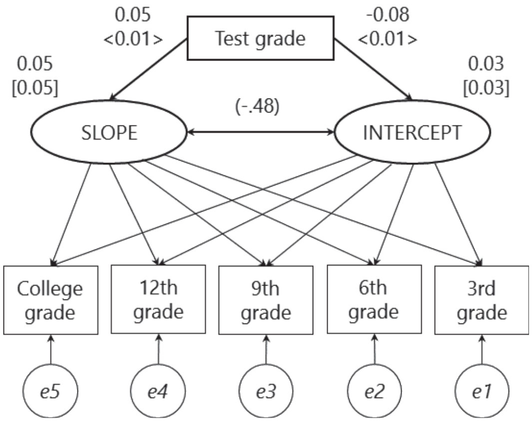

Figure 6. The latent growth curve model for response accuracy. Standard errors, variances, and a correlation coefficient are in angle, square, and round brackets, respectively; e = measurement error; all the numerics are significant at the level of .01.

The unstandardized estimate of SLOPE (0.05) shows that the mean correct response rates rose by 5% as the textual input increased from 3rd grade to College grade. However, the growth of response accuracy was unequal because the SLOPE variance was significant. The significant variance of INTERCEPT also indicates that accuracy variability was large at the time of 3rd grade. As visualized in Figure 5, these individual differences can be explained by the test difficulty, and the path coefficient from Test grade to INTERCEPT supports this observation. The value of –0.08 means that, as the test items grew in difficulty (i.e., G4 → GP1), the mean correct response rates dropped by 8%. Additionally, the significant correlation (–.48) indicates that, when the mean correct response rates were initially lower, their latent growth rates were higher, which is especially apparent in GP1.

We need to be careful in discussing the SEM results because the latent growth curve model did not fit the data perfectly. However, the model was able to explain the relationship between input quantity and vocabulary growth in the sense that (a) the incremental textual input heightened the response accuracy and (b) the growth of response accuracy was correlated with the test difficulty. The first result suggests that the corpus-derived vocabulary knowledge computed by LSA developed as the input quantity increased. As proposed by Landauer and his colleagues, this constitutes evidence that LSA is able to simulate the growth curve of reading-oriented vocabulary knowledge (Biemiller et al., 2014; Landauer et al., 1998, 2011). This level of vocabulary growth is consistent with the results from the experimental approach on incidental L2 vocabulary learning from texts (e.g., Chen & Truscott, 2010; Elgort & Warren, 2014; Horst et al., 1998; Waring & Takaki, 2003; Webb, 2008). Although this preliminary study did not compare the vocabulary growth of the knowledge models with the acquisition process of actual learners, Hamada (2014, 2015) provided evidence that LSA can predict Japanese EFL learners' outcomes of contextualized word learning.

The smaller growth of response accuracy in the easier test grades is not a surprising result. Given that the easier tests are designed to assess the vocabulary knowledge of beginner-level learners, the differences between the immature and mature knowledge models were predictably small (e.g., 3rd grade vs. College grade). In contrast, the modeling results demonstrated that those differences were large in the more difficult tests. Thus, the response accuracy provided by LSA computation was consistent with test difficulty, which further supports the validity of simulating the growth of vocabulary knowledge from incremental textual input.

Nevertheless, the College grade knowledge model achieved only 65% accuracy in the G4 test, which presumably does not reflect the actual test performance of Japanese EFL learners. Further research will be required to examine why there was no large growth of response accuracy in the easier test grades, in terms of the limitations of (a) LSA computation (Inohara & Kusumi, 2012), (b) incidental learning of words from reading (e.g., Waring & Takaki, 2003), and (c) the measurement methodology of vocabulary knowledge (e.g., Morimoto, 2006). The first two points cannot be answered in the present study; however, regarding the third point, Morimoto (2006) indicated that a sentence completion test has less validity in measuring the semantic knowledge of words than a synonym selection test even though the test items are contextualized. In contrast, Landauer and Dumais (1997) showed that an LSA knowledge model composed from a 61,000-word corpus achieved 65% accuracy in the retired TOEFL® synonym selection tests, and its response patterns were similar to actual L2 learners. Although the present study used Eiken vocabulary test items to assess the contextual usage of words, it is necessary to further consider the multifaceted nature of vocabulary knowledge.

The Latent Growth of the LSA Values (RQ2)

Figure 7 summarizes the increase and distribution of the LSA values between the item stems and the corresponding alternatives for each Eiken test grade across five knowledge models (see Table 2 for details). Robust maximum likelihood estimation was applied because the univariate and multivariate normality was violated (*Z*s > 1.96). The obtained data did not fit the latent growth curve model well: $\chi^2(10) = 72.58$, $p < .001$; CFI = .91; TLI = .86; RMSEA = .15, 90% CI [.13, .18]; SRMR = .15. No multicollinearities were found (r range = .43 to .86 [< .90]). Figure 8 displays the results of the latent growth curve modeling for the LSA values.

Growth of the LSA values is not apparent from the unstandardized estimate of SLOPE (0.00). This significant estimate suggests that the overall LSA values did not increase even though textual input was incrementally added to the knowledge models. In contrast, the significant correlation (–.51) indicates that, when the initial LSA values were lower, their latent growth rates were higher, which is especially apparent in GP1. This is supported by the significant variance of INTERCEPT, which indicates that the LSA values were widely dispersed at the time of 3rd grade. Instead of growing due to incremental textual input, the LSA values dropped by 0.01357 when applied to difficult test items because the path coefficient from Test grade to INTERCEPT was significant.

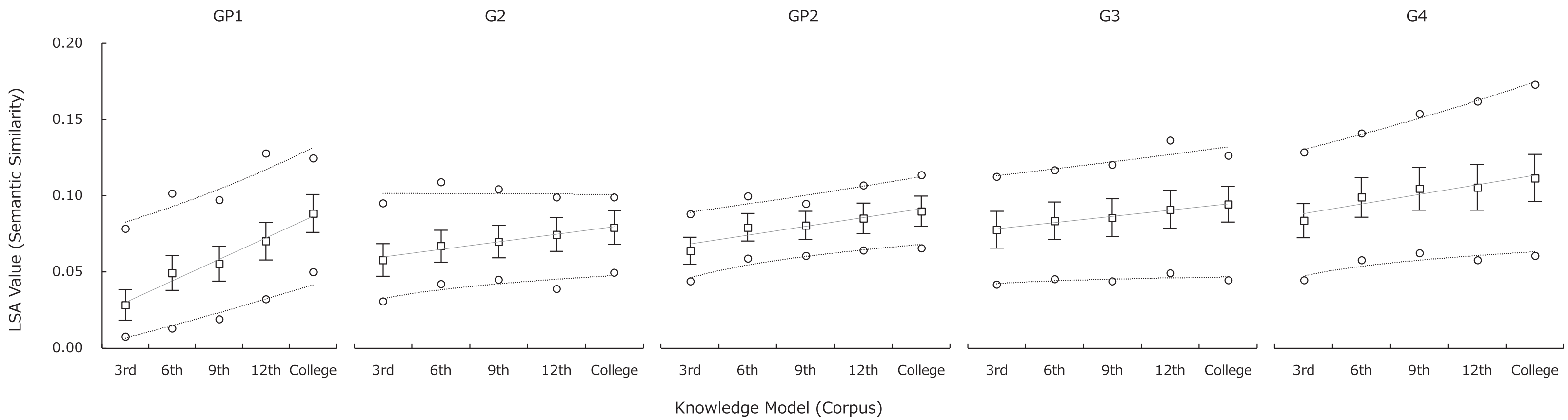

Figure 7. A growth curve of the LSA values. The squares and circles (upper/lower) represent the mean LSA values and maximum/minimum values, respectively; the whiskers represent the 95% CIs of the mean LSA values.

Table 2
Mean LSA Values With 95% CIs and Standard Deviations of the Eiken Vocabulary Test

Knowledge	GP1 (*k* = 70)			G2 (*k* = 56)			GP2 (*k* = 56)			G3 (*k* = 42)			G4 (*k* = 42)		
models	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*	*M*	95% CI	*SD*
3rd grade	.03	[.02, .04]	.04	.06	[.05, .07]	.04	.06	[.05, .07]	.03	.07	[.06, .09]	.04	.08	[.07, .09]	.04
6th grade	.05	[.04, .06]	.05	.06	[.05, .08]	.04	.08	[.07, .08]	.03	.08	[.07, .09]	.04	.10	[.08, .11]	.04
9th grade	.06	[.04, .07]	.05	.07	[.06, .08]	.04	.08	[.07, .09]	.04	.08	[.07, .10]	.04	.10	[.09, .11]	.05
12th grade	.07	[.06, .08]	.05	.07	[.06, .08]	.05	.08	[.07, .09]	.04	.09	[.07, .10]	.05	.10	[.09, .12]	.05
College grade	.09	[.08, .10]	.05	.08	[.07, .09]	.04	.09	[.08, .10]	.04	.09	[.08, .10]	.04	.11	[.09, .12]	.05

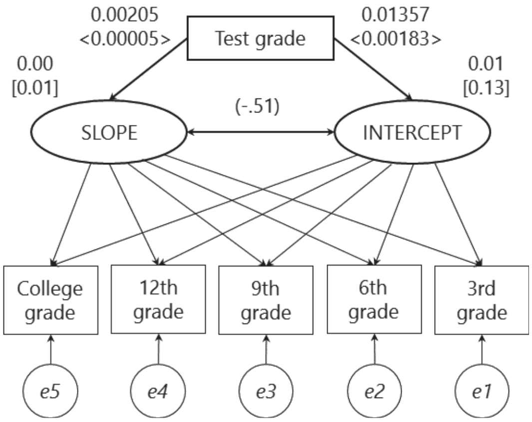

Figure 8. The latent growth curve model for the LSA values. Standard errors, variances, and a correlation coefficient are in angle, square, and round brackets, respectively; e = measurement error; all the numerics are significant at the level of .01.

The current discussion must be tentative because the SEM results were not robust. Whereas Figures 5 and 7 show a similar growth of vocabulary knowledge as assessed by the response accuracy and LSA values, the modeling suggests a different result, namely, no latent growth of the LSA values.[4] The other results were fully consistent with the growth of response accuracy.

There are many aspects to evaluating the results of this LSA simulation. Probably the most important step is to compare the patterns of answering the Eiken vocabulary test items between Japanese EFL learners and the LSA knowledge models. In particular, it is necessary to examine whether the responses provided by the learners follow the imperceptible differences in LSA values among alternative options. For example, the 3rd grade model was able to select *seat* (.07) as a correct answer instead of *case* (.06) in a test dialogue (*A: Is this — taken? B: No. No one is sitting there*), but the two options' LSA values were actually close. In contrast, it is assumed that Japanese EFL learners who know both target words would easily determine which word semantically fits into the contextual message.

LSA is specialized for extracting and representing word meanings depending on their contextual usage by computing an LSA value (Landauer et al., 1998); therefore, the particular characteristics of Eiken vocabulary test items seem to affect the present results. The strength of stem-option semantic similarities determines the vocabulary test difficulty because it improves the success in deriving the word meanings from context (Hamada, 2015). To avoid the learners easily discriminating between a correct answer and the distractors, the stem-option semantic similarities might be empirically adjusted in designing test items. For example, Nakagawa (2007) demonstrated that semantic stem-option linkage was not related to the accuracy of Japanese EFL learners' performance in the Eiken vocabulary test.

In addition to understanding how sensitive the simulation results provided by LSA are to a particular corpus (Biemiller et al., 2014), there are a number of issues to be resolved in using LSA values to estimate the growth of vocabulary knowledge from textual input. Because the response to RQ2 was not robust, detailed and systematic investigation into the validation of LSA computation and measurements that reflect the processes and outcomes of vocabulary acquisition is required.

Patterns of Word Knowledge Development (RQ3)

Table 3 shows the results of correlation analyses for the response accuracy and the LSA values; the College grade model was used as a benchmark for the comparisons. The analysis demonstrated that 3rd grade response patterns were significantly different from College grade ones in terms of both response accuracy (.18) and LSA values (.43). However, the correlation coefficients increased as textual input was added to the knowledge models, and strong correlations ultimately surfaced between the 12th grade and College grade models in both response accuracy (.79) and LSA values (.86). This shows that the cumulative textual input improved the lower graders' word knowledge to the level of adult word knowledge. Figure 9 exemplifies this trend; whereas the dots on the scatterplot of the 3rd grade × College grade comparison were distributed widely, they gradually converged upon a diagonal line.

Although an increase in the correlation coefficients was found in each test grade, the growth speed differed according to test difficulty. Specifically, response accuracy from 3rd grade to 6th grade grew more rapidly in the beginner test grades (e.g., G3; .11 → .63) than in the advanced test grades (e.g., GP2; .08 → .33). Similar results were found in terms of the LSA values. In particular, the level of correlations in the LSA values was almost sequential across the test grades in each knowledge model (e.g., 9th grade: .56 [GP1] → .78 [G2] → .80 [GP2] → .89 [G3] → .92 [G4]; 12th grade: .74 [GP1] → .93 [G2] → .93 [GP2] → .94 [G3] → .95 [G4]).

These results are relevant to the discussion on RQs 1 and 2. The result that the knowledge models constructed by LSA corresponded well with the Eiken test difficulty suggests the validity of computation simulation using LSA being used to estimate the input quantity required for the knowledge models to achieve sophisticated word knowledge. As with the experimental approach (e.g., Horst et al., 1998), the LSA-based simulation replicated the development of reading-oriented vocabulary knowledge from textual input. However, unless future research compares the present results with the test performance of Japanese EFL learners, careful consideration is required when assessing the validity of LSA in estimating the optimal input quantity for L2 vocabulary acquisition.

Some limitations have been mentioned in an earlier section. Another possible issue related to RQ3 is how well the LSA simulation is able to account for individual differences in vocabulary growth. Although LSA could simulate incremental vocabulary acquisition from textual input, immature word knowledge will not always grow to be mature when we look at individuals (e.g., Nation, 2013). The results of the latent growth curve modeling show some individual differences in vocabulary growth (RQs 1 and 2); however, it is necessary to determine whether the results reflect actual L2 learners' traits in vocabulary acquisition. Biemiller et al. (2014) have claimed that, "[b]ecause each individual is exposed to text in their own way, the best that can be achieved is comparing average paths, with all their inherent differences" (p. 146). Whereas this claim rings true, there is large variability in L2 vocabulary acquisition from texts (e.g., Elgort & Warren, 2014). According to Landauer et al. (2011), this

Table 3

Pearson Product-Moment Correlation Coefficients With 95% CIs

Knowledge models	GP1 (k = 70) r	GP1 95% CI	GP1 p	G2 (k = 56) r	G2 95% CI	G2 p	GP2 (k = 56) r	GP2 95% CI	GP2 p	G3 (k = 42) r	G3 95% CI	G3 p	G4 (k = 42) r	G4 95% CI	G4 p	Total (k = 266) r	Total 95% CI	Total p
Mean correct response rates																		
3rd grade	.18	[–.05, .40]	.129	.21	[–.06, .45]	.126	.08	[–.19, .34]	.547	.11	[–.20, .40]	.495	.19	[–.12, .47]	.229	.18	[.06, .29]	.003
6th grade	.33	[.10, .53]	.005	.59	[.38, .74]	.000	.33	[.07, .54]	.014	.63	[.40, .78]	.000	.59	[.35, .76]	.000	.47	[.37, .56]	.000
9th grade	.55	[.36, .69]	.000	.67	[.49, .79]	.000	.69	[.52, .81]	.000	.73	[.55, .85]	.000	.72	[.53, .84]	.000	.66	[.59, .72]	.000
12th grade	.73	[.59, .82]	.000	.88	[.80, .93]	.000	.79	[.66, .87]	.000	.79	[.64, .88]	.000	.82	[.69, .90]	.000	.79	[.74, .83]	.000
LSA value (Semantic similarity)																		
3rd grade	.39	[.17, .57]	.001	.35	[.09, .56]	.009	.30	[.04, .52]	.026	.56	[.31, .74]	.000	.70	[.51, .83]	.000	.43	[.32, .52]	.000
6th grade	.44	[.23, .61]	.000	.67	[.50, .79]	.000	.63	[.44, .77]	.000	.87	[.76, .93]	.000	.85	[.74, .92]	.000	.63	[.55, .70]	.000
9th grade	.56	[.37, .70]	.000	.78	[.65, .86]	.000	.80	[.68, .88]	.000	.89	[.80, .94]	.000	.92	[.86, .96]	.000	.73	[.67, .78]	.000
12th grade	.74	[.61, .83]	.000	.93	[.89, .96]	.000	.93	[.88, .96]	.000	.94	[.90, .97]	.000	.95	[.91, .97]	.000	.86	[.83, .89]	.000

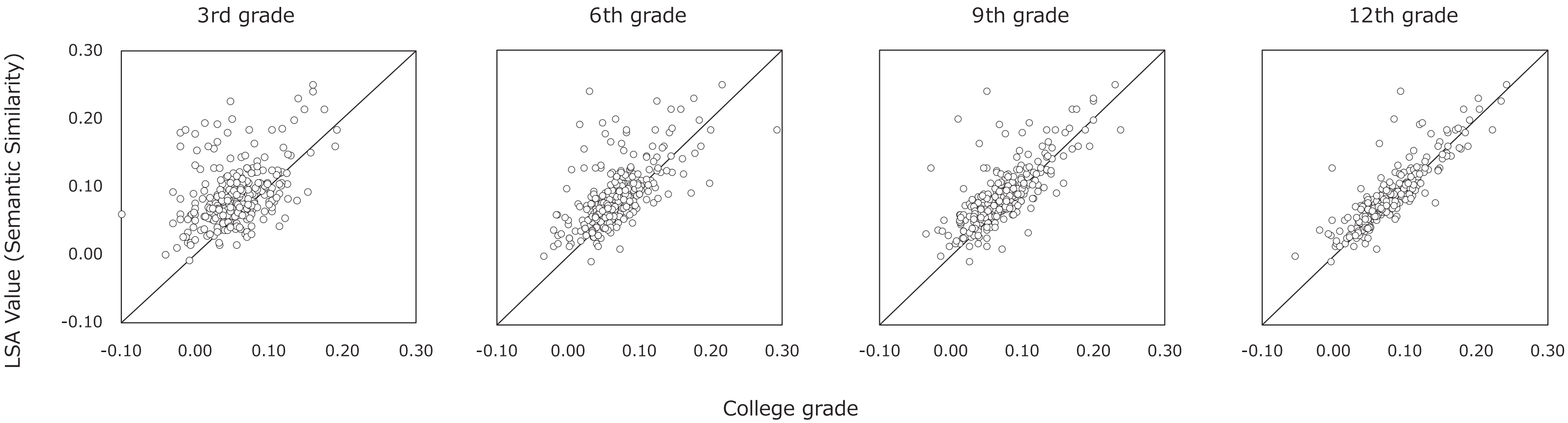

Figure 9. Scatterplots of LSA values between the College grade knowledge model and the other grades' knowledge models (k = 266).

problem can be solved by modifying the text input to knowledge models in order to correct and improve the actual rate of the incidental learning of word meanings.

Conclusion

This preliminary study demonstrated that the model of statistical language analysis was able to simulate the development of L2 vocabulary knowledge from incremental textual input. The answers to the three RQs showed the applicability of LSA to estimating the input quantity needed for L2 learners to acquire knowledge of word meanings and their usage in context. Specifically, incremental textual input increased the response accuracy in the Eiken vocabulary test, and the rate of knowledge development differed according to test difficulty (RQ 1). Additionally, word knowledge approached College level as a result of incremental textual input (RQ3). Nevertheless, sophisticated follow-up studies are required to answer why LSA values were not consistent with the initial prediction (RQ2).

The current results suggest the course of future research that simulates L2 vocabulary acquisition from texts. First, it is necessary to compile corpora of English textbooks used in Japan because the corpus used in this study originally reflected L1 students' knowledge. Unfortunately, in terms of estimating the optimal input quantity, how many times a certain word appeared in a particular knowledge model was not reported in any LSA research. For example, we do not know how often the word *accuracy* (see the Procedure section) occurred in the different types of documents. Although one might expect that a variety of textbook corpora have been compiled so far (Chujo, 2015), LSA requires a special kind of corpus editing due to the need for word-by-document matrices (Quesada, 2007). Moreover, we need to define what exactly language input in the context of English education in Japan is. Textbooks used in regular classes should be a major source of language input for Japanese EFL learners (Ishikawa, 2008); however, a variety of communication sources such as teachers' instructions must be considered for an accurate simulation.

The possible methodology for the estimation of input quantity in the vocabulary growth of Japanese EFL learners includes (a) compiling a cumulative corpus from textbooks, (b) the construction of knowledge models that reflect the vocabulary knowledge of Japanese EFL learners, and (c) the validation of the knowledge models using multicomponent testing. Imagine that a 7th grade knowledge model gives a correct answer in a vocabulary test; we can then go back to the corpus to compare how many times the target word appeared in textbooks and teacher instructions with how contextualized it was. At the same time, it is also important to determine what types of vocabulary tests should be used in assessing a simulation's result and the way it compares to actual learners' performance. Although this simulation will not capture the full picture of L2 vocabulary acquisition, the conclusion I hope to draw is that it helps estimate how much input is required for L2 learners to be able to use their reading-oriented vocabulary knowledge in different situations of language use.

All the possible improvements described here must be implemented to take a further step into conducting research that compares LSA simulations with the process and outcomes of L2 vocabulary acquisition of Japanese EFL learners. This could help reveal any individual differences in L2 vocabulary acquisition and estimate the optimal input quantity for each individual to acquire new words from textual input. Although determining which words to teach and in what texts is a complicated task because of various extraneous factors such as learners' motivation and strategy, the approach of statistical language analysis has enormous potential for identifying the optimal input quantities and types of texts for the development of L2 vocabulary knowledge.

Notes

1. A cosine value is calculated using the following formula (Martin & Berry, 2007):

$$\cos(word_a, word_b) = \frac{\sum word_a word_b}{||word_a|| \cdot ||word_b||}.$$

Therefore, the calculation of the LSA value $_{\text{learning} \times \text{English}}$ is as follows:

- $\sum word_a word_b =$ (0.29 × 0.14) + (0.20 × –0.08) + (0.20 × –0.08) + (0.53 × 0.56) + (0.16 × 0.18) + (0.55 × 0.55) + (0.24 × 0.24) = 0.69
- $||word_a|| \cdot ||word_b|| = \sqrt{(0.29^2 + 0.20^2 + 0.20^2 + 0.53^2 + 0.16^2 + 0.55^2 + 0.24^2)} \times \sqrt{(0.14^2 + -0.08^2 + -0.08^2 + 0.56^2 + 0.18^2 + 0.55^2 + 0.24^2)} = 0.79$
- 0.69/0.79 = 0.87

2. Except for word types included in the General Reading Space corpus, no other specification (especially the total frequency of each word) was described.
3. The frequency levels were calculated in a web-based program (v8an; http://www.tcp-ip.or.jp/~shim/j8web/j8web.cgi).
4. Multiple comparisons with the Bonferroni correction showed significant differences in the LSA values among the five knowledge models in GP1. Similar differences were found in the other test grades, although some adjacent comparisons were insignificant. The different results between SEM and this analysis of variance might be attributed to the instability of the present SEM results.

Acknowledgements

I wish to thank three anonymous reviewers for their useful comments to improve an earlier version of this article.

References

Biemiller, A., Rosenstein, M., Sparks, R., Landauer, T. K., & Foltz, P. W. (2014). Models of vocabulary acquisition: Direct tests and text-derived simulations of vocabulary growth. *Scientific Studies of Reading*, *18*, 130–154. doi:10.1080/10888438.2013.821992

Chen, C., & Truscott, J. (2010). The effects of repetition and L1 lexicalization on incidental vocabulary acquisition. *Applied Linguistics*, *31*, 693–713. doi:10.1093/applin/amq031

Chujo, K. (2015). English textbook corpora and applications: A brief history of development. *English Corpus Studies*, *22*, 77–85.

Crossley, S. A., Salsbury, T., McCarthy, P., & McNamara, D. S. (2008). Using latent semantic analysis to explore second language lexical development. *Proceedings of the twenty-first international FLAIRS conference*, *12*, 136–141. Retrieved from http://www.aaai.org/

Papers/FLAIRS/2008/FLAIRS08-040.pdf
Dennis, S. (2007). How to use the LSA web site. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 57–70). Mahwah, NJ: Lawrence Erlbaum Associates.
Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*, 365–414. doi:10.1111/lang.12052
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 225–241. doi:10.1017.S0272263102002024
Hamada, A. (2014). Using latent semantic analysis to promote the effectiveness of contextualized vocabulary learning. *JACET Journal*, *58*, 1–20. Retrieved from http://ci.nii.ac.jp/naid/110009807050
Hamada, A. (2015). Improving incidental L2 vocabulary learning with latent semantic analysis. *ARELE: annual review of English language education in Japan*, *26*, 61–75. Retrieved from http://ci.nii.ac.jp/naid/110010006581
Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223. Retrieved from http://nflrc.hawaii.edu/rfl/PastIssues/rfl112horst.pdf
Inohara, K., & Kusumi, T. (2011). Psychological validity of conceptual similarities based on latent semantic analysis: Merits and limitations in statistical analysis of corpuses. *Japanese Psychological Review*, *54*, 101–122.
Inohara, K., & Kusumi, T. (2012). Effects of reading habits on vocabulary: Examination by latent semantic analysis. *Cognitive Studies*, *19*, 100–121. Retrieved from https://www.jstage.jst.go.jp/article/jcss/19/1/19_100/_pdf
Ishikawa, S. (2008). *Eigo corpus to gengo kyoiku: Data to shiteno text* [English corpus and language education: Text as data]. Tokyo, Japan: Taishukan Shoten.
Japan Association of College English Teachers (JACET) Basic Word Revision Committee (Ed.). (2003). *JACET list of 8000 basic words*. Tokyo, Japan: Author.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240. doi:10.1037/0033-295X.104.2.211
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284. doi:10.1080/01638539809545028
Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for word knowledge. *Scientific Studies of Reading*, *15*, 92–108. doi:10.1080/10888438.2011.536130
Martin, D. I., & Berry, M. W. (2007). Mathematical foundations behind latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 35–56). Mahwah, NJ: Lawrence Erlbaum Associates.
Morimoto, Y. (2006). Goi-test no keishiki ga goi-chishiki to dokkai-noryoku no sokutei ni oyobosu eikyo [The effects of vocabulary-test formats on the assessment of vocabulary knowledge and reading ability]. *STEP Bulletin*, *18*, 77–91. Retrieved from https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol18/vol_18_p77-p91.pdf

Muraoka, R. (2010). Chugakkou-kentei-kyokasho de gakushu-sareru goi, gakushu-sarenai goi: Nobegosu, kotonarigosu, goi-range no kantenkara [Words to be learned and not to be learned from the authorized textbooks used in Japanese junior high school: Investigation from tokens, types, and vocabulary range]. *STEP Bulletin*, *22*, 182–203. Retrieved from https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol22/vol_22_p182-p203.pdf

Nakagawa, C. (2007). Kikanbu to sentakushi no kanren-kyodo ga goi-test performance ni oyobosu eikyo [The effects of the relatedness between stems and alternatives on the performance in vocabulary tests]. *STEP Bulletin*, *19*, 41–56. Retrieved from https://www.eiken.or.jp/center_for_research/pdf/bulletin/vol19/vol_19_p41-p56.pdf

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Pellicer-Sánchez, A., & Schmitt, N. (2010). Incidental vocabulary acquisition from an authentic novel: Do things fall apart? *Reading in a Foreign Language*, *22*, 31–55. Retrieved from http://nflrc.hawaii.edu/rfl/April2010/articles/pellicersanchez.pdf

Quesada, J. (2007). Creating your own LSA spaces. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 71–88). Mahwah, NJ: Lawrence Erlbaum Associates.

Reynolds, B. L., & Wible, D. (2014). Frequency in incidental vocabulary acquisition research: An undefined concept and some consequences. *TESOL Quarterly*, *48*, 843–861. doi:10.1002/tesq.197

Toyoda, H. (Ed). (2008). *Data mining nyumon: R de manabu saishin data kaiseki* [Introduction to data mining: Learning the up-to-date analysis with R]. Tokyo, Japan: Tokyo Tosho.

Toyoda, H. (Ed.). (2014). *Kyobunsan kozo bunseki R-hen: Kozo hoteishiki modeling* [Structural equation modeling in R]. Tokyo, Japan: Tokyo Tosho.

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*, 130–162. Retrieved from http://nflrc.hawaii.edu/rfl/October2003/waring/waring.pdf

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*, 232–245. Retrieved from http://nflrc.hawaii.edu/rfl/October2008/webb/webb.pdf

View publication stats