Using Latent Semantic Analysis to Explore Second Language Lexical Development

Scott A. Crossley, Tom Salsbury*, Philip McCarthy, and Danielle S. McNamara****

Department of English
Mississippi State University

*Department of Education
Washington State University

**Institute for Intelligent Systems
University of Memphis

Abstract

This study explores how Latent Semantic Analysis (LSA) can be used as a method to examine the lexical development of second language (L2) speakers. This year long longitudinal study with six English learners demonstrates that semantic similarity (using LSA) between utterances significantly increases as the L2 learners study English. The findings demonstrate that L2 learners begin to develop tighter semantic relations between utterances and words within a short period. The results have implications concerning the growth of lexical networks. This study also has important implications for inductive learning and contextualized vocabulary learning.

Introduction

The development of semantic knowledge is an important area of study in second language acquisition (SLA). This is not only because a lack of semantic knowledge can lead to global errors in language use, but also because errors based on semantic knowledge are the most common type of errors in second language (L2) production and are judged to be key elements in inhibiting communication (Ellis, R., 1995; Ellis, R., Tanaka, & Yamakazi, 1994). How L2 learners are able to quickly acquire words and word meanings relates to their ability to successfully make sense of and respond to language input as well as create coherent output. Thus, exploring how L2 learners develop semantic knowledge could lead to a better understanding of L2 language processing and support theories of L2 pedagogy.

In the last few decades, new theories of L2 vocabulary acquisition have evolved including lexical networks (Haastrup & Henriksen, 2000), lexical emergence (Meara, 2006), and lexical inference (Hucking & Coady, 1999). Indeed, these theories are redirecting traditional studies of lexical acquisition for L2 learners. However, while these theories are proving to be important in explaining how L2 learners develop robust vocabularies, systematic analyses of lexical L2 development that consider these theoretical perspectives are lacking. This paper begins to address this need by exploring how computational models of semantic acquisition can inform theories of L2 semantic knowledge. We do so by analyzing spontaneous spoken data collected from six L2 learners over the course of a year. We use Latent Semantic Analysis (LSA) to measure semantic similarity between utterances across the year and use the findings to examine the development of semantic relationships in L2 learners' speech over time.

Second Language Lexical Acquisition

Many traditional views of L2 lexical acquisition were constrained by the limited definition of a lexical entry as well as constrained by views that successful lexical acquisition was the result of explicit learning techniques and memorization strategies. While it is likely true that explicit vocabulary instruction concentrating on the first 2,000 to 3,000 words is valuable for the beginning learner (Nation, 2005), it is generally agreed that subsequent vocabulary acquisition results from inference strategies and the development of word connections (Hucking & Coady, 1999; Haastrup & Henriksen, 2000). This idea is premised on words being intertwined with one another forming word connections that are highly clustered and interconnected. In this way, L2 learners create and develop lexical networks through the accumulation of words. As new words emerge, L2 learners also create networks of links between the new words and already learned words (Haastrup & Henriksen, 2000). The assets of these interconnections are that no matter the number of the projections between words, the distance between each projection is relatively small (Ellis, N. 2007). Ellis argues that the condensed nature of these projections allows for the rapid creation of lexical networks and the efficient acquisition of lexical items.

Latent Semantic Analysis

The learning of words is a result of many processes. However, in this study, we simplify the examination of word learning by considering only one process: the semantic properties of words. Specifically, we examine the semantic properties of words using the LSA model with the understanding that LSA can be used as a model to approximate the development of semantic relations.

LSA works by determining the similarity of passage meaning through the analysis of large corpora. However, LSA does not depend on word frequency counts, word co-occurrences, or word correlations to measure semantic similarity between text samples. Nor does LSA depend on perceptual information, instinct, intentions, syntax, or pragmatics. In LSA, the similarity of words is based on topical and referential meanings. These meanings come from a large domain of knowledge where there are many direct and indirect relationships. Because there are too many relationships in language for each element to be introduced individually, most semantic knowledge is likely gained through induction (Landauer & Dumais, 1997). The induction of semantic knowledge is located contextually in LSA. Thus, if two words appear in the same context, and every other word in that context appears in many other contexts without them, the two will acquire semantic similarity to each other but not to the rest (Landauer & Dumais, 1997; Landauer, 2007). In this way, connections between related words develop. As an example, all component features related to *legs*, *tails*, *ears*, and *fur* are related to each other not only because of the occasions when they occur together, but, importantly, as the indirect result of the occasions when they occur with other elements (such as *animals*).

To determine the similarity of passage meaning, LSA depends on the mathematical technique known as singular value decomposition (SVD) which reduces thousands of dimensions and relationships between words to a more manageable number (usually around 300) in a manner similar to a factor analysis (Landauer, Foltz, & Laham, 1998). The data that SVD reduces in LSA are the raw, local associations between the words in a text and the context in which they occur. The dimensions reduced through SVD represent how often a word or words occur within a document (defined at the word, sentence, paragraph, or text level). These documents become weighted vectors and text selections are matched by comparing the cosine between two sets of vectors (receiving values between -1 and 1). This cosine relates to the similarity or dissimilarity between documents. In this way, LSA measures how likely two words will appear in similar discourse settings and then relates this inversely to their semantic distance, thus making word associations based on semantic similarity (Landauer & Dumais, 1997).

LSA as a model of human conceptual knowledge

LSA has been shown to model human conceptual knowledge in various ways. The most prominent of these that are of interest for the goals of this paper include the use of LSA to make word sorting and relatedness judgments, generate word synonymy judgments, and model vocabulary learning.

Word Sorting and Relatedness Judgments. As reported in Landauer et al. (1998), LSA has been successful in a replication of Anglin's (1970) study of word sorting and relatedness judgments. In Anglin's study, adults and children clustered words based on part of speech similarities, confirming that participants used abstract relations when grouping words. Landauer et al. (1998) conducted a similar study using LSA to replicate the grouping methods. The study found that LSA correlations with the grouping data rose as the number of documents included in the LSA semantic space rose. This led Landauer et al. to conclude that LSA sorted words in a similar manner to human participants.

Synonymy Judgments. To judge how accurately LSA recognized word synonymy, Landauer and Dumais (1997) tested LSA word scores on 80 test items from the synonym portion of the Test of English as a Foreign Language. The test items contained a stem word and four alternative words. The LSA-determined choices were made by computing cosines between the vector of each stem word and the four provided alternative words. The alternative word with the highest cosine was selected as the synonym. The LSA model scored 64.4% on the test set, which compared favorably to the 64.5% average of the L2 learners who had taken the same test. The results of this study imply that LSA can match the semantic knowledge of moderately proficient L2 English learners with respect to meaning similarity.

Vocabulary Learning. Children learn words at a phenomenal rate of about 10-15 words a day (Anglin, 1993). This has never been matched by adult vocabulary learning from word lists alone. Using LSA to replicate children's word learning rate, Landauer and Dumais (1997) trained an LSA model using reading texts which were equated on the number and variety of texts that introduce children to language. Using this method, the LSA model approximated the vocabulary learning of children and exceeded learning rates that had been achieved in controlled studies that taught children word meanings through context. It was estimated that three-fourths of the lexical knowledge acquired by LSA was through induction from data about other words.

L2 Vocabulary Learning and LSA

The studies above suggest that word learning is not the result of memorization techniques, but the result of words being learned implicitly with already known words helping to place new words in their proper semantic spaces. This approach to learning attempts to explain how children learn vocabulary so quickly: they do not learn thousands of individual words, but rather construct semantic spaces and embed related words and phrases into them (Kintsch, 2001). Considering that the learning of words is an inherently inductive process that allows for meaning to be induced through context, it is possible that most referential meaning is inferred from a speaker's experience with words alone (Landauer & Dumais, 1997).

Theories of lexical learning that have key elements similar to those used to support the LSA model are common in L1 and L2 learning. For instance, Landauer and Dumais (1997) contend that vocabulary learning is the result of implicit associations made between words and not the explicit learning of their meanings. In L2 vocabulary learning, researchers argue that learners use previous word knowledge to build associations with the new words they encounter (Ellis, R., 1994; Haastrup & Henriksen, 2000). This approach to vocabulary acquisition, referred to as network building, states that learners are able to integrate new vocabulary into their mental lexicon only through comparison with previously learned words (R. Ellis, 1994). In this way, lexical networks develop as words gain associations with one another (Haastrup & Henriksen, 2000). Thus, lexical acquisition results from simple learning processes, applied over an extended period of time, producing complex knowledge systems. LSA is identified as one computational approach that could provide supporting evidence of how language data linked with simplistic learning mechanisms can lead to the emergence of complicated language representations in L2 learners (N. Ellis, 1998, 1999).

Methods

Our purpose in this study is to explore how a computational model of semantic knowledge might measure the lexical growth of L2 learners. To accomplish this, we test whether LSA measurements of semantic co-referentiality increase as learners study an L2 and whether a common measurement of lexical proficiency (in this case lexical diversity) demonstrates growth as well. A significant increase in lexical diversity measures would provide evidence that another aspect of lexical proficiency is increasing, thus supporting the notion that L2 learners' lexical proficiency is developing. A significant increase in LSA values would provide additional support for the growth of L2 learners' lexical proficiency. However, more importantly, an increase in LSA values might suggest that developing L2 vocabularies exploit the strengths of semantic networks and create stronger associations and interconnections between words and utterances. This result could also give additional credence to theories of inductive and contextualized learning. For this study, we chose to look at a small set of learners over a long period of time (e.g., rather than a cross-sectional study of a large group of learners). A longitudinal approach is necessary when analyzing the process of lexical development because the process requires long-term language analysis to capture gradual changes over time (Haastrup & Henriksen, 2000).

Participant Selection

To gather the language data for this study, a group of L2 English learners enrolled in an intensive English program at a university in the United States were interviewed in natural settings every 2 weeks (not including university breaks) over a 1-year period. While interviewers came prepared with a variety of elicitation methods, the sessions contained naturalistic discourse. To control for familiarity between the student and interviewer, each L2 learner had at least four different interviewers over the course of the year. Learners' proficiency levels were tested upon arrival to the program, and all participants in the study tested into the lowest proficiency level, Level 1, of a 6-level program. The current paper reports on six of the learners in the original cohort of students. Other learners were dropped from the analysis because of large gaps in the elicitation data. The participants ranged in age from 18 to 29 years old and were from varied language backgrounds. They had all studied English in their native secondary schools and had successfully completed high school in their country of origin.

Corpus

The spoken data collected from the six learners was transcribed and forms the foundation for this analysis. For the six learners, the average number of meetings was 16.5, (SD = 2.07) and the average length of the transcript was 1658.29 words (SD = 473.48). In preparation for the analysis of the learner corpus, transcriptions of each elicitation session were modified in the following ways: Interjections such as *ah*, *uhm*, and *yea* were deleted as were any words that were clearly non-English words. Non-target like forms of irregular past tense verbs were included (e.g. *taked*, *sleeped*); however, these were quite rare. Proper nouns were also left in the data. All punctuation except the period and question mark was eliminated from the transcriptions. Each elicitation session was saved as a single text file containing the oral production of only the learner in focus. The text file was manually and electronically checked for spelling errors.

Word Measurements

To collect LSA measurements, each text file was analyzed using the computational tool *Coh-Metrix*, which measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis (Graesser et al., 2004). LSA values from Coh-Metrix are taken from the college level TASA corpus. Coh-Metrix LSA values were used to measure lexical development because they operate at the text level and provide measurements that examine similarity in meaning and conceptual relatedness between text segments. In Coh-Metrix, sentences, paragraphs, and texts are measured as weighted vectors and LSA values are computed as geometric cosines between these vectors with values ranging between -1 to 1 (-1 being low similarity in meaning and conceptual relatedness and 1 being high). Because the data used in this study was based on spoken utterances and not written text, only LSA paragraph to

Table 1
Mean and Standard Deviations (SD) for LSA values and Measure of Textual Lexical Diversity (MTLD) Values

Week	Mean LSA Value	SD LSA Value	Mean MTLD value	SD MTLD value	*Week Comparison*	*F*	*p*	*F*	*P*
2	0.16	0.01	28.43	7.27		LSA	LSA	MTLD	MTLD
4	0.20	0.04	25.37	4.55	2 to 4	4.85	<.01	1.21	0.32
16	0.23	0.04	32.26	7.27	2 to 16	36.68	<.01	2.12	0.21
32	0.30	0.17	31.12	3.78	2 to 32	3.99	0.01	1.36	0.30
50	0.26	0.06	34.88	4.25	2 to 50	20.29	<.01	4.40	0.09
52	0.32	0.12	35.43	2.92	2 to 52	10.88	0.02	7.93	0.04

paragraph values were analyzed. This is because sentence punctuation for the spoken utterances would be artificial and many spoken utterances were too short to provide proper lexical coverage. However, complete propositions were easily broken up based on turn-taking. This analysis thus measured the LSA values of adjacent utterances spoken by L2 learners to identify growths in semantic and conceptual similarity.

To investigate whether other features of the L2 learners' lexical proficiency increased along with LSA values, we used the *Measure of Textual and Lexical Diversity* (MTLD: McCarthy, 2005). MTLD is similar to other measures of lexical diversity (LD) such as *vocd* (Malvern, et al., 2004) or *TTR* (Templin, 1957), but, unlike all other LD measures, it does not correlate with text length (McCarthy, 2005; McCarthy & Jarvis, 2007). It does, however, strongly correlate with other LD measures. Thus, MTLD is able to assess differences of lexical diversity between spoken and written texts, even while those texts may be considerably different in terms of text length. We use MLTD in this study to assess whether L2 learners exhibit signs of lexical growth over the course of a year.

Results

We conducted a repeated measures Analysis of Variance (ANOVA) using the LSA results from Coh-Metrix to test the assumption that as the learner acquires lexical proficiency, semantic relations develop as well. The ANOVA was used to track the linear trend of the LSA values across the temporal intervals and to test the assumption that as learners' time spent studying English increased, their word similarity patterns would increase. Because all participants did not share all the same temporal data points, the ANOVA test analyzed development on a quarterly basis. This allowed for breaks in the data related to winter and spring holidays to be considered as well as missing data points resulting from absences. Because data was available for the first two sessions and the last two sessions for all six learners, it was included. These data points were also analyzed with data from the 16th week and the 32nd week. The ANOVA was supplemented with a post-hoc test of within-subjects contrasts to examine where in the temporal progression significant differences in output could be identified. To support the ANOVA findings, the correlation between time spent learning English and LSA values was also computed.

The results showed that L2 learners' LSA values increase as a function of time, defined as the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning, $F(5, 25) = 3.95$, $p < .01$. Within-subjects contrasts demonstrated that the LSA values from the last meeting on the 52nd week were significantly different from the first meeting on the second week, $F(5, 25) = 10.878$, $p = .02$. Additionally, significant differences in LSA values were found between the second week and the 16th week, $F(5, 25) = 36.68$, $p < .01$, and the second week and the 50th week, $F(5, 25) = 20.29$, $p < .01$ (see Table 1). In addition, there was a significant correlation between the L2 learners' time spent studying English and their LSA values ($r = 0.49$, $p < 0.001$, $N = 37$). These findings provide evidence that as learners spend time studying the English language semantic relatedness, increases between utterances.

To test the assumption that other aspects of learners' lexical proficiencies grow as time spent learning English increases, a repeated measures ANOVA was conducted using the MTLD results from Coh-Metrix. In a similar fashion to the LSA analysis, this ANOVA was used to track the linear trend of the MTLD values over the increasing temporal intervals. Like the LSA data, participants did not share all the same temporal data points so the ANOVA analyzed development on a quarterly basis, but included the second and fourth weeks and the 50th and 52nd weeks. To support the findings from the MTLD ANOVA, the correlation between time spent learning English and MTLD values was also calculated.

The results showed that the L2 learners' MTLD values increase as a function of time, defined as the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning, $F(5, 25) = 7.41$, $p < .001$. Within-subjects contrasts indicate that MTLD values from the 52nd week were significantly greater than those of the second week, $F(5, 25) = 7.93$, $p < .05$ (see Table 1).

There was a significant correlation between the L2 learners' time spent learning English and the MTLD values (r = 0.60, p < 0.001, N = 37). These findings provide evidence that as learners spend time studying the English language, lexical diversity increases as well semantic relatedness.

Discussion

The statistical analyses provided evidence that the L2 learners in this study displayed increasing lexical proficiency over time. Importantly, the results also showed that the L2 learners' utterances began to reflect closer semantic similarity as indicators of lexical proficiency increased. The LSA ANOVA results demonstrated that learners' utterances develop significant semantic links over the course of a year and thus supported the notion that as learners acquire lexical proficiency, the semantic properties of their utterances become more closely interrelated and conceptually similar. These results also appear to support the notion that as learners progress, they make use of words that are more semantically associated to one another. The semantic similarity of utterances used by the L2 speakers in this study seem to develop quite rapidly with significant differences noted within the first 16 weeks of learning English in a second language environment. This trend continued to the 52nd week.

One important implication of closer semantic similarity is that it likely helps in maintaining semantic links between utterances and ensuring that interlocutors are able to co-reference new information with past information. Thus, it appears that as L2 learners acquire lexical ability, they are also able to create more cohesive relationships between utterances leading to more cohesive speech patterns. As these relationships progress over time, links between words and utterances that are more conceptually similar seem to develop as well. These links likely assist in the learning of new words by providing associations to connect new words with old words based on semantic similarity. This finding helps provides evidence that acquiring vocabulary consists of creating lexical networks and locating words and phrases within those networks (N. Ellis, 2007; Haastrup & Henriksen, 2000; Kintsch, 2001).

These findings have important implications for research in the development of increased L2 lexical fluency. Research conducted by Meara and Schur (2002) has demonstrated that L2 lexical association networks have more associations per word than L1 networks. This finding was based on the results of an experiment which showed that L2 learners produce more varied responses than L1 speakers on word association tasks. The authors argued that L2 networks consist of a few large components and are characterized by more random associations than L1 networks. Meara and Schur argued that the less connected and less predictable lexical networks exhibited by L2 speakers would not allow L2 learners to link vocabulary to smaller, more constrained, and tightly connected lexical sets. While native speakers of English may have tighter lexical networks than L2 English learners, our study provides evidence that L2 learners appear to develop semantic networks that become more interconnected over time. This supports the contention of Haastrup and Henriksen (2000) that L2 lexical networks are constantly developing as learners restructure their semantic networks based on new input and new associations.

Conclusions

The results of this study provide important implications for the development of lexical networks in L2 learners as well as provide evidence that L2 learners begin to develop closer semantic similarities between speech segments as they progress in acquiring a second language. The use of an LSA model of language acquisition is important because the model itself mimics aspects of language learning that are important in L2 studies, such as the use of inductive and contextualized learning. Learning vocabulary through techniques that embrace interaction with language and the development of connections between words is likely a more valuable approach than those that promote rote memorization because word meanings are likely individualistic and based on word experience. The more experience one has with a word within context, the better the chance that connections between that word and other words will develop, assisting in the formation of semantic spaces and lexical networks.

This study has a number of limitations that should be considered. While LSA is a powerful tool that acquires language through contextualized induction and appears to match human judgments on word synonymy and word sorting, it provides only an approximation of verbal knowledge (Landauer, 2002; Landauer, 2007). Thus, links between LSA measurements and human lexical development are sufficient (but not exhaustive) for explaining lexical acquisition. Additionally, some limitations are located within the dataset used in this study. For instance, unlike the LSA model, which is based on written texts, this study analyzed semantic similarity in spontaneous spoken utterances. However, considering that LSA does not consider syntax or pragmatics, we do not consider this to be a serious limitation. In addition, LSA has been successfully used to analyze spoken data in the past (e.g. Foltz, 2007). The sample population in this study was also small (six learners), but sufficient in terms of statistical power.

We contend that while real world experiments are important, a hybrid approach that considers both real world data and computational tools can also provide significant insights into how L2 learners develop lexical networks. The use of computational tools based on cognitive and psycho-linguistic approaches, corpus linguistic techniques, connectionist theory, and real world data as found in this

study result in an inter-disciplinary approach as called for by N. Ellis (1999).

Research into L2 lexical networks is far from complete, though. This study, while providing support for the development of lexical networks in L2 learners, is only an initial investigation that begins to tap into the possibilities of computational models. Future studies considering the LSA model should examine larger L2 populations and the growth of semantic spaces in written as well as spoken language. Also, future research in L2 lexical development would benefit from the construction of an LSA space specifically designed from L2 input. Such a space would consider L2 reading texts and listening samples and would allow for further investigations of L2 lexical networks and lexical development. In addition, the development of spoken and textual coherence in L2 learners should be further investigated. It is our hope that such studies will be conducted as they could provide valuable information into L2 lexical network development and inform L2 pedagogy. This study provides a step in that direction.

Acknowledgments

The research was supported in part by the Institute for Education Sciences (IES R305G020018-02). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES.

References

Anglin, J.M. 1970. *The growth of word meaning.* Cambridge, Massachusetts: MIT Press.

Anglin, J.M. 1993. Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development*, 58 (10, Serial No. 238).

Ellis, N.C. 1998. Emergentism, connectionism and language learning. *Language Learning*, 48, 631-664.

Ellis, N.C. 1999. The ways of words. In D. Albrechtsen, B. Henriksen, & I. Mees (Eds.), *Festschrift for Kirsten Haastrup* (pp. 15-28). Odense: Odense University Press.

Ellis, N.C. 2007. Optimizing the input: Frequency and sampling in Usage-based and Form-focused Learning. In M. H. Long & C. Doughty (Eds.), *Handbook of Second and Foreign Language Teaching*. Oxford: Blackwell.

Ellis, R. 1994. A theory of instructed second language acquisition. In N.C. Ellis (Ed.), *Implicit and Explicit Learning of Languages* (pp. 79-114). London: Academic Press.

Ellis, R. 1995. Modified oral input and the acquisition of word meanings. Applied Linguistics, 16, 409-435.

Ellis, R., Tanaka, Y., & Yamakazi, A. 1994. Classroom interaction, comprehension, and L2 vocabulary acquisition. Language Learning, 44, 449-491.

Foltz, P. 2007. Discourse coherence and LSA. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 379-400). Mahwah, NJ: Erlbaum.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., & Cai, Z. 2004. Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193-202.

Haastrup, K., & Henriksen, B. 2000. Vocabulary acquisition: Acquiring depth of knowledge through network building. *International Journal of Linguistics*. *10 (2)*, 221-240.

Hucking, T., & Coady, J. 1999. Incidental vocabulary acquisition in a second Language. *Studies in Second Language Acquisition, 21* (2), 181-93.

Kintsch, W. 2001. Predication. *Cognitive Science, 25*, 173-202.

Landauer, T.K. 2002. On the computational basis of learning and cognition: Arguments from LSA. In N. Ross (Ed.), *The Psychology of Learning and Motivation* (43-84), 41.

Landauer, T.K. 2007. LSA as a theory of meaning. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 379-400). Mahwah, NJ: Erlbaum.

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.

Landauer, T.K., Foltz, P.W., & Laham, D. 1998. Introduction to Latent Semantic Analysis. *Discourse Processes, 25*, 259-284.

Malvern, D.D., Richards, B.J., Chipere, N., & Duran, P. 2004. *Lexical diversity and language development: Quantification and assessment*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.

McCarthy, P. M. 2005. An assessment of the range and usefulness of lexical diversity measures and the potential of the measure of textual, lexical diversity (MTLD). *Dissertation Abstracts International, 66* (12), (UMI No. 3199485)

McCarthy, P. M., & Jarvis, S. 2007. A theoretical and empirical evaluation of *vocd*. *Language Testing*. 24: 459-488.

Meara, P. 2006. Emergent properties of multilingual lexicons. *Applied Linguistics, 27* (4), 620-644.

Meara, P., & Schur, E. 2002. Random association networks: A baseline measure of lexical complexity. *British Association for Applied Linguistics, 34*, 169-182.

Nation, I.S.P. 2005. Teaching and learning vocabulary. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 581-595). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Templin, M. 1957. *Certain language skills in children*. Minneapolis: University of Minnesota Press.