The development of semantic relations in second language speakers: A case for Latent Semantic Analysis

Scott A. Crossley
Mississippi State University
sc544@msstate.edu

Tom Salsbury
Washington State University
tsalsbury@wsu.edu

Danielle S. McNamara
University of Memphis
d.mcnamara@mail.psyc.memphis.edu

Abstract

This study explores how a Latent Semantic Analysis (LSA) index provided by Coh-Metrix can be used as a method to examine the development of lexical networks in second language (L2) speakers. Using the LSA index along with an index of lexical development (lexical diversity), a measure of general linguistic development (TOEFL scores) and indices of lexical overlap in a yearlong longitudinal study, this study demonstrates that semantic similarity along with TOEFL and lexical diversity values significantly increase as learners study a second language. Measures of word overlap did not demonstrate significant growth. The findings demonstrate that the LSA value reported by Coh-Metrix can be used to approximate lexical growth and that L2 learners likely begin to develop tighter semantic relations between utterances and words within a short period of time. This finding implies the growth of lexical networks. This study also has important implications for the development of coherent speech patterns in L2 learners.

Key Words: Latent Semantic Analysis, Lexical Networks, Corpus Linguistics, Lexical Proficiency, Computational Linguistics, MTLD.

Resumen

En este estudio exploramos cómo puede usarse el índice de Latent Semantic Analysis (LSA) de Coh-Metrix como un método de examinar el desarrollo de redes léxicas en estudiantes de una lengua extranjera (L2). Al usar el índice LSA junto con un índice de desarrollo léxico (diversidad léxica), una medida de desarrollo lingüística general (TOEFL) e índices de superposición léxica en un estudio longitudinal a lo largo de un año, este estudio demuestra que la similaridad semántica, junto

con valores de TOEFL y de diversidad léxica, aumenta de modo significativo durante el aprendizaje de una segunda lengua. Las mediciones de superposición de palabras no aumentaron de modo significativo. Los resultados demuestran que el valor LSA de Coh-Metrix puede usarse para aproximar el crecimiento léxico y que los estudiantes de L2 probablemente empiezan a desarrollar relaciones semánticas más estrechas entre los enunciados y las palabras en un plazo corto. Este resultado implica el crecimiento de redes léxicas. Este estudio tiene también importantes implicaciones para el desarrollo de pautas de habla coherente de los estudiantes de una L2.

Palabras clave: Ánalisis semantico latente, redes léxicas, lingüística de corpus, conocimiento del léxico, lingüística computacional, MTLD.

1. Introduction

Words are not simply referents to objects and phenomena and word learning is not simply learning its definition. Learning a word requires the recognition of a word's auditory sound patterns, orthography, syntactic properties, lexical structure and the word's lexical and semantic relations (Bogaards, 2001; Nation, 2005). Many traditional views of lexical acquisition in second language (L2) learners have been constrained by the limited definition of a lexical entry as well as constrained by views that successful lexical acquisition is the result of explicit learning techniques and memorization strategies. While it may be 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 (Huckin & Coady, 1999; Haastrup & Henriksen, 2000). This notion is premised on words being intertwined with one another, forming word connections that are highly clustered and interconnected (Ferrer i Cancho & Solé, 2001; Ferrer i Cancho, Solé, & Köhler, 2004; Haastrup & Henriksen, 2000). In this way, L2 learners create and develop lexical networks as word knowledge accumulates. As new words emerge, L2 learners also create networks of links between the new words and already learned words (R. Ellis, 1994; Haastrup & Henriksen, 2000). The assets of these interconnections are that no matter the number of the nodes between words, the distance between each node is relatively small (Ellis, 2007). It is the condensed nature of these nodes that allow for the rapid creation of lexical networks and the more efficient acquisition of L2 lexical items.

Recent studies (e.g. Crossley, Salsbury, & McNamara, 2009; 2010) have demonstrated the success of computational models in accounting for the development of lexical networks in L2 learners. Our current study further examines

how a computational model of semantic acquisition can be used to test theories of L2 lexical development. We do so by analyzing spontaneous spoken data collected from six L2 learners over the course of a year. We use a Latent Semantic Analysis (LSA) index taken from the computational tool Coh-Metrix (Graesser, McNamara, Louwerse, & Cai, 2004) to measure semantic similarity between utterances across the year and contrast the LSA index with a measure of language growth (TOEFL) and an index of lexical diversity (Measure of Text and Lexical Diversity; McCarthy, 2005) along with indices of simple word overlap. We use the findings to examine the growth of semantic relationships in L2 learners' speech develops.

1.1. Latent Semantic Analysis

While the learning of a word is a collection of processes, the examination of how a word is learned is less dynamic because multiple cognitive processes can generally not be evaluated at once. In this paper, we consider only one of the many processes described earlier: the semantic relations between words. Specifically, we examine the semantic properties of words using an LSA model. We do not use LSA as a model of language production (Landauer, 2002), but rather as an approximation of human verbal knowledge. We are thus interested in exploring how LSA can be used to model the development of semantic relations in L2 learners.

LSA works by estimating the similarity of passage meaning through the analysis of large corpora. 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 both direct and indirect relationships. Because there are too many relationships in language for each element to be introduced individually, the semantic knowledge in LSA is gained through induction (Landauer & Dumais, 1997). The induction of semantic knowledge is also located contextually in LSA within the corpora. In LSA, if two words appear in the same context, and every other word in that context appears in many other contexts without them, the two words will acquire semantic similarity to each other but not to the other words (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*). Since the LSA model argues for semantic growth based on inference through weak interrelations between domains of knowledge, it is similar to connectionist models of semantic knowledge (Landauer & Dumais, 1997)

To determine the similarity of passage meaning, LSA depends on a mathematical technique known as singular value decomposition (SVD) which

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

57

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; 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).

1.2. 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 relevant to the goals of this study include the use of LSA to make word sorting and relatedness judgments, to generate word synonymy judgments, to model vocabulary learning, and to measure textual and spoken coherence.

Word Sorting and Relatedness Judgments. As reported in Landauer et al. (1998), LSA has been successful in a replication study of classic word sorting and relatedness judgments (i.e., Anglin, 1970). Anglin's study found that participants clustered words based on abstract parts-of-speech similarities. A similar study was conducted using LSA to replicate the grouping methods that humans employ (Landauer et al., 1998). The study found that LSA correlations with human grouping data rose as the number of documents included in the LSA semantic space rose. Thus, Landauer and Dumais (1997) concluded that LSA sorted words in a similar manner as humans.

Synonymy Judgments. Landauer and Dumais (1997) tested LSA word values on 80 test items from the synonym portion of the Test of English as a Foreign Language (TOEFL). The LSA-determined choices were made by computing cosines between the vector of each stem word and the four provided alternative words. The word with the highest cosine was selected as the synonym. The LSA model scored 64.4% on the test set and this 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 closely mimics the semantic knowledge of moderately proficient L2 English learners with respect to meaning similarity.

Vocabulary Learning. LSA has also been used to replicate children's word learning rate (Landauer & Dumais, 1997). To simulate children's vocabulary, LSA

was trained using reading texts that equated to 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.

Coherence. LSA has also been used to effectively measure the amount of textual and spoken coherence and predict the effect of text coherence on comprehension. In a replication of two classic coherence studies (Britton & Gulgoz, 1991; McNamara, E. Kintsch, Butler, Songer, & W. Kintsch, 1996), Foltz, W. Kintsch, and Landauer (1998) demonstrated that LSA was able to predict the readability and cohesiveness of texts that had been modified to make them more or less coherent. Foltz et al. (1998) contended that these findings support the notion that LSA provides an accurate measure of text comprehensibility by measuring the semantic relatedness of adjacent text and topics (see also, McNamara, Cai, & Louwerse, 2007). Additionally, LSA values have been linked to the incidence of negotiations for meaning in the spoken interactions of L2 learners and native speakers (Crossley, Salsbury, McCarthy, & McNamara, 2008).

The results of these studies have several implications concerning LSA and the acquisition of word knowledge. First, it appears that LSA models abstract word categories in a manner similar to human participants. Second, it appears that LSA acquires knowledge about word similarities from text alone and that this word similarity equals that learned by L2 English learners. Third, the LSA model seems capable of acquiring word knowledge at a rate that is similar to that attained by children. Fourth, LSA seems to be able to measure cohesive relationships between language samples. Lastly, most of the word knowledge attained by the LSA model is the result of inductive methods of learning based on exposure to text alone. This suggests 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 respective semantic spaces.

2. Methods

Our purpose in this study is to evaluate how a computational model of semantic knowledge (LSA) compares to measures of lexical growth and language. If LSA values demonstrate growth in conjunction with other measures of lexical and linguistic growth, LSA might add explanatory power to theories of L2 lexical development. To accomplish this, we test the L2 learner's spoken language demonstrated growth in LSA measurements of semantic co-referentiality as the learners acquire a second language. To ensure that the L2 learners in this study demonstrated lexical growth in their L2, we also examined whether their spoken language exhibited growth in

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

————— 59

lexical diversity values over the course of the year. We also tested general language growth through the use of TOEFL scores. While we realize that the TOEFL measures general language proficiency by assessing reading, grammar, and listening skills, we also realize that TOEFL scores will broadly correlate to measures of vocabulary. Thus, our use of TOEFL scores is also premised on the notion that a single, higher-order factor related to language proficiency underlies individual first order language ability factors such as lexical ability (Oller, 1979). Such a position has been supported in the literature (Carroll, 1983; Shin, 2005) and specifically in reference to TOEFL scores (Bachman, Davidson, Ryan, & Choi, 1995). We use lexical diversity scores as a proxy for lexical growth in a similar fashion of many past researchers (Crossley et al., 2009; Jarvis, 2002; Malvern & Richards, 2002; Polio, 2001). We also report on whether or not indices of lexical overlap increase as well. Lexical overlap values are included to assess whether growth in LSA is simply a result of lexical repetition or, in contrast, related to developing semantic relations.

A finding that demonstrates growth in TOEFL scores, lexical diversity values, and LSA values, but not lexical overlap indices would suggest that as L2 learners demonstrate general language and specific lexical growth, they exploit the strengths of semantic networks and create stronger associations and interconnections between words and utterances. This result would give credence to theories of lexical acquisition and provide evidence that simple lexical constructions could lead to more complex lexical networks.

2.1. Participant Selection

For this study, we chose to look at a small set of learners over a long period of time instead of a cross-sectional study of a large group of learners. A longitudinal approach is advantageous when analyzing the process of lexical development because the process requires long-term language analysis to capture its gradual nature (Haastrup & Henriksen, 2000). To gather the language data needed for this study, a group of L2 English learners enrolled in an intensive English program at a large university in the United States were interviewed every 2 weeks (not including program and university breaks) over a 1-year period. The bi-weekly elicitation sessions were scheduled to coincide with the students' regular speaking class. 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. Learners' language growth was also assessed every other month through TOEFL scores. 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 during the year of observation or because they did not complete the year. Each learner in the study was given a pseudonym; this paper

reports on data from Marta (Spanish L1), Takako (Japanese L1), Eun Hui (Korean L1), Faisal (Arabic L1), Kamal (Arabic L1) and Jalil (Arabic L1). The participants ranged in age from 18 to 29 years old and had all successfully completed high school in their country of origin. None of the learners had lived in the United States for more than three weeks prior to the start of the study. All learners had studied English in their native secondary schools.

Although interviewers came to the sessions prepared with various topics from which the L2 learners could discuss, the sessions were characterized by naturally occurring discourse. Procedurally, this meant that interviewers were encouraged to engage with the learners. They responded to the learners' comments and opinions and were free to agree or disagree openly. Interviewers were recruited from a graduate course in second language acquisition, and they used the data (transcripts) collected in the sessions for their own graduate work for the course. From the learner's point of view, the graduate student interviewers were similar to conversation partners. Interviewers and learners were paired together for a semester (6 data collecting sessions). New interviewers were recruited for subsequent semesters of the study although several of the interviewers chose to continue working with the learners as volunteers. The learners and interviewers were invested in the data collection process, and a positive rapport developed between the two groups of participants. This promoted a comfortable and safe environment in which the learners could talk freely and openly. It could be argued that the elicitation sessions were more structured than natural conversation because the interviewers were given specific materials to use and instructed on techniques to encourage the learners to speak as much as possible while keeping their own oral output to a minimum. On the other hand, the elicitation sessions were much less structured than a formal interview. The learners also asked questions of the interviewers and engaged in natural conversational turns.

Eighteen sets of elicitation materials were used in the data collection. The materials were designed to give the participants some control over the discussion topics in order to promote conversation. Table 1 presents a list of the 18 sets of materials and includes the average number of topics discussed overall during each session. The table also presents the average number of utterances that occurred across the learners for each session. We define an utterance as the entire conversational turn for the learner. An utterance is followed by the interviewer's conversational turn. Overlaps between interviewers and learners were transcribed sequentially with the interviewer's utterance following the learner's. Interviewers came with topics printed on cards; the learners then chose which topics they wanted to talk about. Changes in topic, either to another topic card or to a related topic, were initiated by the learner. The conversation topics were not limited to the printed materials

that were given to the participants. Learners often spontaneously introduced topics related to the elicitation materials or introduced topics tangential to a particular conversation. Elicitation sessions generally lasted between 30-45 minutes. The sessions were tape recorded and later transcribed. This method of data collection has successfully been employed in past studies of L2 lexical development (Crossley et al., 2009; 2010; Salsbury, Crossley, & McNamara, 2010).

Table 1. Materials and topics

Material	Average number of topics		Average number of utterances	
	N	StDev	N	StDev
1. General information	13	8.79	83	22.24
2. Emotion and topic cards	12	3.19	70	17.37
3. Photograph descriptions	5	1.1	51	4.44
4. Emotion and topic cards	8	2.64	46	27.07
5. Attitudes towards learning English	9	2.58	40	10.60
6. Imagination cards	7	2.04	44	10.50
7. General information (new interviewer)	12	.98	86	26.37
8. Past, present, future chart	8	2.17	70	16.72
9. Photograph descriptions	5	3.16	50	16.66
10. Attitudes towards host community	9	1.76	65	38.24
11. Important personal events, people and talents	6	2.28	47	26.85
12. Hopes and plans for summer vacation	8	2.00	64	26.07
13. Memorable experiences, dreams, regrets and fears	8	1.53	91	8.96
14. Making a difference in another's life	7	2.05	73	17.25
15. Wishes	9	2.83	66	16.73
16. Attitudes towards learning English	9	2.94	74	30.81
17. Goals, careers, regrets and fantasies	7	2.61	74	35.55
18. Greatest achievements in life	7	2.8	83	34.3

2.2. Corpus

The spoken data collected from the six learners was transcribed and forms the foundation for this analysis. Descriptive data for the corpora of each learner can be found in Table 2. 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. Proper nouns were 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

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

oral production of only the learner in focus, not the interviewer or other learners participating in the session. The text file was manually and electronically checked for spelling errors.

Table 2. Descriptive statistics for language participant data

Learner	Number of Meetings	Average Number of Words per Utterance	Average Number of Utterances	Average Number of Words per Text
Eun Hui	18	21.31	52.17	1120.67
Faisal	13	33.42	71.08	1870.07
Takako	18	19.40	51.00	1470.72
Kamal	15	23.75	50.27	1216.20
Jalil	17	38.82	61.76	2359.77
Marta	18	33.31	63.61	1912.00

2.3. Coh-Metrix

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 were used to measure lexical development because they work at the text level and not the word level. Thus they 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). Coh-Metrix reports LSA values based on the college level TASA space, which is a common LSA space that covers reading knowledge up to the first year of college..

Because the data used in this study was based on spoken utterances and not written text, only LSA paragraph to paragraph values were analyzed. This is because sentence punctuation for the spoken utterances would be artificial and many spoken utterances are too short to provide proper lexical coverage. However, complete propositions are easily broken up based on turn-taking. These propositions are analyzed at the paragraph level by Coh-Metrix. This analysis thus measured the LSA values of adjacent utterances spoken by L2 learners to identify growths in semantic and conceptual similarity. This method of comparing adjacent utterances is similar to methods successfully employed by Crossley et al. (2009; 2010) and Salsbury et al. (2010).

Lexical co-reference indices were also collected using Coh-Metrix. Coh-Metrix measures lexical co-reference between adjacent sentences in four ways: noun overlap between adjacent sentences, argument overlap between adjacent sentences, stem overlap between adjacent sentences, and content word overlap between adjacent sentences. Noun overlap measures how often a common noun exists between two sentences. Argument overlap measures how often two sentences share common arguments (nouns, pronouns and noun phrases), while stem overlap measures how often a noun in one sentence shares a common stem with other word types in another sentence. Content word overlap measures the incidence of similar content words between sentences (McNamara, Louwerse, McCarthy, & Graesser, in press). In each of these indices, Coh-Metrix compares adjacent sentences in a text and computes the number or nouns, arguments, stems, and content words that overlap (i.e. are shared).

We also collected lexical diversity (LD) values from Coh-Metrix to examine lexical growth in the L2 learners. We used the *Measure of Textual and Lexical Diversity* (MTLD: McCarthy, 2005). MTLD is similar to other indices of LD such as D (Malvern, Richards, Chipere, & Duran, 2004) or *TTR* (Templin, 1957), but, unlike other LD indices, MTLD does not vary as a function of text length for text segments whose length is in the 100-2000 word range. The algorithm used in MTLD separates texts into segments that have a TTR value of .71. The algorithm then divides the total segments in the text by the total words in the text. The results of this division come to produce a forward value. The algorithm then repeats the entire process, but starts from the end of the text instead of the beginning (the reverse value). The forward value and the reverse value are then summed and divided by two, creating the MTLD values. Unlike most other lexical diversity indices, MTLD also allows for comparisons between text segments of largely different lengths (up to 2000 words) and has proven effective for both spoken and written texts. Additionally, it produces reliable results over a wide range of genres while strongly correlating with other LD indices (McCarthy, 2005). Thus, MTLD is able to assess differences of lexical diversity between different texts even while those texts may be considerably different in terms of text length. MTLD has been successfully deployed in second language studies that have examined lexical differences between L1 and L2 writers (McCarthy, Hall, Duran, Doiuchi, Duncan, Fujiwara, & McNamara, in press) and in studies that have examined the development of lexical proficiency in L2 spoken output (Crossley et al., 2009).

2.4. Statistical Analysis

To test the assumption that as the learner acquires lexical proficiency, semantic relations develop as well, a series of repeated measures Analyses of Variances

(ANOVAs) were conducted using the LSA, lexical co-reference, and lexical diversity results from Coh-Metrix and the learners' TOEFL scores. These ANOVAs were used to track the linear trend of the selected measures over the increasing temporal intervals and to test the assumption that as learners' time spent learning English increased, their word similarity patterns, their TOEFL scores, and their lexical diversity would increase, while their lexical co-reference values would not increase. Because all participants did not share all the same temporal data points, the LSA, lexical c-reference, and lexical diversity ANOVA tests analyzed development on a trimester basis. This allowed for breaks in the data related to winter and spring breaks to be considered as well as missing data points resulting from participant absences. Because data was available for the first two weeks and the last two weeks for all six learners, it was included. These data points were analyzed with data from the 16th week and the 32nd week as well. This ANOVA was supplemented with a post-hoc test of within-subjects contrasts in order to examine where in the temporal progression significant differences in output could be identified. In reference to the TOEFL scores, only four learners completed all six of the TOEFL examinations given over the course of the year. One learner missed the second TOEFL examination and a different learner missed the fifth. As a result, we had TOEFL scores for all six learners from the 6th, 22nd, 42nd, and 52nd weeks of learning only. This is in contrast with the temporal data points used to analyze the LSA, lexical co-reference, and lexical diversity values which were analyzed on the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning.

3. Results

3.1. TOEFL Scores

ANOVA results show that the L2 learners' TOEFL scores increased as a function of time, defined as the 6th, 22nd, 42nd, and 52nd weeks of learning, $F(5, 15) = 22.78$, $p < .001$ (see Table 3 for details). A test of within-subjects contrasts demonstrated that the TOEFL scores from the last examination on the 52nd week were significantly different from the first examination on the sixth week, $F(5, 15) = 33.98$, $p < .01$. Additionally, significant differences in TOEFL scores were found between the sixth week and the 32nd week, $F(5, 15) = 17.23$, $p < .01$, and the sixth week and the 42nd week, $F(5, 15) = 35.80$, $p < .01$. There was also a significant linear trend, $F(1, 25) = 40.07$, $p < .001$. These findings provide evidence that TOEFL scores increase with time spent learning English and implicate that significant language development occurred.

Table 3. Mean and Standard Deviations for TOEFL Scores

Week	Mean	Standard Deviation
6	358.33	49.79
22	418.83	33.04
42	450.66	30.12
52	458.83	29.25

3.2. Lexical Diversity Index

ANOVA results demonstrated that the L2 learners' MTLD values increased as a function of time, defined as the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning, $F(5, 25) = 7.41$, $p < .001$ (see Table 4 for details). A test of within-subjects contrasts demonstrated significant differences between the second week and the 52nd. There was also a significant linear trend, $F(1, 25) = 22.83$, $p < .01$. These findings indicated that as time spent learning English increased, MTLD values increased.

Table 4. Mean and Standard Deviations for MTLD values

Week	Mean	Standard Deviation
2	28.43	7.27
4	25.37	4.55
16	32.26	7.27
32	31.12	3.78
50	34.88	4.25
52	35.43	2.92

3.3. Latent Semantic Analysis Values

ANOVA results show that the L2 learners' LSA values increased as a function of time, defined as the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning, $F(5, 25) = 3.95$, $p < .01$ (see Table 5 for details). A test of 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 first week and the 16th week, $F(5, 25) = 36.68$, $p < .01$, and the first week and the 50th week, $F(5, 25) = 20.29$, $p < .01$. There was also a significant linear trend, $F(1,$

25) = 10.99, $p < .05$. These findings provide evidence that LSA semantic similarity values increase with time spent learning English.

Table 5. Mean and Standard Deviations for LSA Values

Week	Mean	Standard Deviation
2	0.16	0.01
4	0.20	0.04
16	0.23	0.04
32	0.30	0.17
50	0.26	0.06
52	0.32	0.12

3.4. Lexical Coreference Values

ANOVA results demonstrate that the L2 learners' argument overlap values did not increase as a function of time, defined as the 2nd, 4th, 16th, 32nd, 50th, and 52nd weeks of learning, $F(5, 25) = 1.63$, $p > .05$, nor did the learners noun overlap values, $F(5, 25) = .69$, $p > .05$, stem overlap values, $F(5, 25) = .48$, $p > .05$, and content word overlap values, $F(5, 25) = 1.11$, $p > .05$. In addition, tests of within-subjects contrasts demonstrated that the no differences were noted between the argument, noun, stem, and content word overlap values between any temporal indices. These findings show that co-reference indices related to lexical repetition did not demonstrate growth over the course of the yearlong study.

4. Discussion

The statistical analyses in this study provide evidence that as the language skills of second language learners examined in this study developed (as reported by the TOEFL scores), our two indices related to lexical development demonstrate significant development as well. Our traditional index of lexical growth (lexical diversity) exhibited growth over the yearlong study as did the Coh-Metrix LSA index. These findings support the hypothesis that as the L2 interlanguage grows, the learners' spoken production also begins to reflect closer semantic similarity. This implies that as the L2 learners advanced, they began to make more accurate semantic connections and use more semantically similar terms.

Excerpts from the oral elicitation sessions illustrate the change in spoken production toward more semantically similar terms. The first set of examples is taken from Marta's data. In the first excerpt (1), taken from the first trimester,

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

67

Marta is discussing happiness. In the second example (2), taken from the last trimester, Marta is discussing her family and her limited time. The LSA value for the first example is .14. The LSA value for the second example is .63.

(1) Happy, oh happy, difference times I happy. For example, the evaluation, happy. Speak my family happy. Different times happy. Confused, lonely, I don't know. My boyfriend, I call me. I speak a one hour. Excited.

(2) I feel worried about my son or my daughter because I do not have time. I want to spend more time with my husband and my child. But I think my works very important.

The increased LSA scores from sample 1 to sample 2 likely result from the production of more conceptually similar words. Sample 2, for instance, has three well-defined semantic spaces where conceptually similar words are used to link the spaces. These spaces include cognition (with words such as *feel, worry,* and *think*), time (with words such as *time* and *spend*), and family (with words such as *son, daughter, husband,* and *child*). Sample 1, on the other hand, has less defined semantic spaces and more of them.

The second set of examples is taken from Eun Hui's data. In the first excerpt (3), taken from the first trimester, Eun Hui is discussing what she did the previous weekend. In the second example (4), taken from the last trimester, Eun Hui is discussing violence and an ex-boyfriend. The LSA value for the first example is .09. The LSA value for the second example is .37.

(3) We picnic. My coordinator and classmate. Shopping and park. Dinner. Chicken on the grill. Very delicious.

(4) Finally I met him. I cannot meet. I could not meet you. Even though you kill me. We argued about two hours. And then he first hit me.

In a similar manner to the first set, this second set of examples demonstrates that the L2 learner uses more defined semantic spaces in the last trimester than in the first trimester. Specifically, in example 3, the learner produces many semantic spaces with few semantic links. In contrast, example 4 has three well-defined semantic spaces: a contact space (constructed with words such as *meet*), a time space (constructed with words such as *finally, hours,* and *first*), and a conflict space (constructed with words such as *kill, argue,* and *hit*).

The LSA ANOVA results demonstrate that the L2 learners' utterances in this study developed significant semantic links over the course of a year. This finding supports the notion that as learners acquire English vocabulary, 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 related 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 all the way to the 52nd week.

These findings seem to support the idea that as L2 learners began to expand their vocabularies they created semantic networks that are more closely connected. These networks, over time, appear to develop networks of links between words and utterances that are more conceptually similar. These links likely assist in the learning of new words by providing association networks to connect new words with old words based on semantic similarity. These findings thus provide support to the notion that acquiring vocabulary consists of creating lexical networks and locating words and phrases within those networks (N. Ellis, 1997, 2007; R. Ellis, 1994; Ferrer i Cancho & Solé, 2001; Ferrer i Cancho, Solé, & Köhler, 2004; Haastrup & Henriksen, 2000; Kintsch, 2001). The findings also provide support for theories of emergent lexical properties in that lexical networks can emerge and develop from simplistic initial conditions that eventually advance into more complex networks.

These findings might also have important implications for the development of increased lexical fluency. Research conducted by Meara and Schur (2002) demonstrated that L2 lexical association networks have more associations per word than L1 networks. This finding was based on the results of an experiment that showed that L2 learners produce more varied responses than L1 speakers on word association tasks. This led Meara and Schur to argue that stimulus words in the L2 network generate more associations and thus had looser semantic spaces. Meara and Schur (2002) described L2 lexical networks, in comparison to L1 lexical networks, as having more random associations than L1 networks and consisting of a few large components. In reviewing their findings, they argued that the less connected and less predictable lexical networks exhibited by L2 speakers do not allow L2 learners to link vocabulary to smaller, more constrained, and tightly connected lexical sets. While native speakers of English obviously have tighter lexical networks than L2 English learners (Meara & Schur, 2002), our research on semantic growth using LSA measurement of L2 learners utterances appears to demonstrate that L2 learners develop semantic networks that become more interconnected over time. This supports the contention of Haastrup and Henriksen (2000) that the lexical networks of L2 learners are constantly developing as learners restructure their semantic networks based on new input and new associations.

Another important interpretation of these findings stems from the implication that increased LSA values may provide links to the development of cohesive

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

———— 69

speech patterns in second language learners (e.g., Crossley et al., 2008). The findings from this study demonstrate that as learners develop their lexicon, their utterances become more semantically similar. This semantic similarity is important in maintaining semantic links between utterances and ensuring that interlocutors are able to coreference new information with past information. Thus, the findings suggest that as L2 learners acquire lexical ability, they also are able to create more cohesive relationships between utterances leading to more cohesive speech patterns.

5. Conclusion

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 study also bases its findings on the Coh-Metrix computational tool, which is freely available and user friendly. Such a tool can be useful for measuring lexical growth in L2 users. Lastly, the use of an LSA model of language acquisition is significant as the model itself mimics aspects of language learning that are important in L2 studies. This includes the use of inductive learning and contextualized learning. However, while LSA is a powerful tool that acquires language through contextualized induction and appears to match human judgments on word synonymy and word sorting, its learning mechanisms are not completely understood and it is only an approximator of verbal knowledge. Thus, any links between LSA measurements and human lexical development are theoretical.

This study, while providing support for the development of lexical networks in L2 learners, is only an exploratory investigation that begins to tap into the possibilities of the LSA model. Future studies using LSA should consider larger sample populations so that the findings can be extended past the 6 learners in this study. Additional research should also consider the growth of semantic spaces in written as well as spoken language. Also, future research in L2 lexical development that employs an LSA model might benefit from the construction of an L2 LSA space specifically designed using input that is frequent for L2 learners. A space such as this would consider simplified language input, listening samples, and grammar texts. Such studies might prove valuable in further clarifying how L2 lexical networks develop and how L2 lexical acquisition might benefit from contextual language and inductive learning approaches. This study provides a step in that direction.

The finding that the L2 learners in this study develop closer semantic links between utterances as their language skills progress over the course of a year is

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

important. Nevertheless, it should be noted that the findings presented here are open for interpretation as they are a simulation of L2 lexical acquisition. The findings are also exploratory in nature and based on computational models. However, since lexical acquisition is a phenomenon that is generally too complex to be analyzed based on human intuition, computational approaches are beneficial. Such simulations of lexical networks are needed to move the field forward (Meara, 2006) and provide additional research into the development of L2 lexical networks. Such explorations are virtually impossible without the assistance of computational tools such as those used in this study.

Acknowledgments

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.

Bachman, L. F., F. G. Davidson, K. Ryan, and I. C. Choi. 1995. *An investigation into the Comparability of Two Tests of English as a Foreign Language: The Cambridge TOEFL Comparability Study.* Cambridge: UCLES.

Bogaards, P. 2001. "Lexical units and the learning of foreign language vocabulary". *Studies in Second Language Acquisition, 23*: 321-343.

Britton, B. K., and S. Gulgoz. 1991. "Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures". *Journal of Educational Psychology*, 8(3): 329-345.

Carroll, J. B. 1983. "Psychometric theory and language testing". In *Issues in Language Testing Research*, Oller, J. W. (ed.), 80-107. Rowley, MA: Newbury House.

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

71

Crossley, S. A., T. Salsbury, P. M. McCarthy, and D. S. McNamara. 2008. "LSA as a measure of second language natural discourse". In Sloutsky, V., B. Love, and K. McRae (Eds.), *Proceedings of the 30th annual conference of the Cognitive Science Society* 1906-1911. Washington, D.C.: Cognitive Science Society.

Crossley, S. A, T. Salsbury, and D. S. McNamara. 2009. "Measuring L2 Lexical Growth using Hypernymic Relationships". *Language Learning, 59* (2), 307-334.

Crossley, S. A., T. Salsbury, and D. S. McNamara. 2010. The development of polysemy and frequency use in English second language speakers. *Language Learning, 60* (3).

Ellis, N.C. 1997. "Vocabulary acquisition: Word structure, collocation, grammar, and meaning". In McCarthy, M and N. Schmidt (Eds.) 1997, *Vocabulary: Description, Acquisition and Pedagogy* (pp. 122-139). Cambridge: Cambridge University Press.

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

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

Ferrer i Cancho, R., and R. V. Solé. 2001. "The small world of human language". *Proceedings of the Royal Society of London, B., 268*: 2261-2265.

Ferrer i Cancho, R., R. V. Solé, and R. Köhler. 2004. "Patterns in syntactic dependency networks". *Physical Review, E69*: 0519151-0519158.

Foltz, P.W., W. Kintsch, and T. K. Landauer. 1998. "The measurement of textual coherence with Latent Semantic Analysis". *Discourse Processes*, 25: 285-307.

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

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

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

Jarvis, S. 2002. "Short texts, best-fitting curves and new measures of lexical diversity". *Language Testing, 19*(1): 57-84.

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 Ross, N. (Ed.), *The Psychology of Learning and Motivation* (43-84), 41.

_____. 2007. "LSA as a theory of meaning". In Landauer, T., D. S. McNamara, D. Simon, and W. Kintsch (eds.): *Handbook of Latent Semantic Analysis*. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Landauer, T.K., and S. T. Dumais. 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., P. W. Foltz, and D. Laham. 1998. "Introduction to Latent Semantic Analysis". *Discourse Processes, 25*: 259-284.

Malvern, D., and B. Richards. 2002. "Investigating accommodation in language proficiency interviews using a new measure of lexical diversity". *Language Testing, 19*(1): 85-104.

Malvern, D. D., B. J. Richards, N. Chipere, and P. Duran. 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., Hall, C., Duran N.D., Doiuchi, M., Duncan, B., Fujiwara, Y., & McNamara, D.S., (in press). A computational analysis of journal abstracts written by Japanese, American, and British scientists. The ESPecialist.

McNamara, D.S., Z. Cai, and M. M. Louwerse. 2007. "Comparing latent and non-latent measures of cohesion". In Landauer, T., D. McNamara, D. Simon, and W. Kintsch (Eds.) *Handbook of Latent Semantic Analysis*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

McNamara, D.S., Z. Cai, and M. M. Louwerse. 2007. "Optimizing LSA measures of cohesion". In Landauer, T., D.S. McNamara, S. Dennis, and W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis,* pp. 379-400. Mahwah, NJ: Erlbaum.

McNamara, D. S., E. Kintsch, Butler, N. Songer, and W. Kintsch. 1996. "Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text". *Cognition and Instruction*, 14(1): 1-43.

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.

73

McNamara, D. S., M. M. Louwerse, P. M. McCarthy, and A. C. Graesser. (in press). "Coh-Metrix: Capturing linguistic features of cohesion". *Discourse Processes.*

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

Meara, P., and E. Schur. 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 Hinkel, E. (Ed.), *Handbook of research in second language teaching and learning,* pp. 581-595. Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Oller, J. W. (1979). Language tests at school: A pragmatic approach. London: Longman.

Polio, C. 2001. "Research methodology in second language writing research: The case text-based studies". In Silva, T., and P. K. Matsuda (Eds.) *On Second Language Writing,* pp. 91-115. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Salsbury, T., Crossley, S. A, & McNamara, D. S. 2010. Psycholinguistic word information in second language oral discourse. *Second Language Research, 26* (2).

Shin, S. K. 2005. "Did they take the same test? Examinee language proficiency and the structure of language tests". *Language Testing, 22* (1): 31–57.

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

Crossley, S.A., Salsbury, T. & McNamara, D.S. 2010. "The development of semantic relations in second language speakers: A case for Latent Semantic Analysis". *Vigo International Journal of Applied Linguistics* 7: 55-74.