Psycholinguistic word information in second language oral discourse
Tom Salsbury, Scott A. Crossley and Danielle S. McNamara
Second Language Research published online 13 April 2011
DOI: 10.1177/0267658310395851

The online version of this article can be found at:
http://slr.sagepub.com/content/early/2011/04/13/0267658310395851

Published by:

Second Language Research

Additional services and information for Second Language Research can be found at:

Email Alerts: http://slr.sagepub.com/cgi/alerts

Subscriptions: http://slr.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav
Psycholinguistic word information in second language oral discourse

Tom Salsbury
Washington State University, USA

Scott A. Crossley
Georgia State University, USA

Danielle S. McNamara
University of Memphis, USA

Abstract
This study uses word information scores from the Medical Research Council (MRC) Psycholinguistic Database to analyse word development in the spontaneous speech data of six adult learners of English as a second language (L2) in a one-year longitudinal study. In contrast to broad measures of lexical development, such as word frequency and lexical diversity, this study analyses L2 learners’ depth of word knowledge as measured by psycholinguistic values for concreteness, imagability, meaningfulness, and familiarity. Repeated measure ANOVAs yielded significant differences over time for concreteness, imagability, and meaningfulness, where the temporal intervals act as the independent variable, and the MRC values function as the dependent variables. Non-significant results were found for familiarity scores. The results provide evidence that learners’ productive vocabularies become more abstract, less context dependent, and more tightly associated over time. This indicates a deeper knowledge of second language vocabulary and has important implications for how vocabulary knowledge can be measured in future studies of L2 lexical development.

Keywords
second language acquisition, concreteness, imagability, familiarity, meaningfulness, word association, lexical, vocabulary

Corresponding author:
Tom Salsbury, Department of Teaching and Learning, College of Education, Washington State University, Pullman, WA 99164-2132, USA
Email: tsalsbury@wsu.edu
I Introduction

This article addresses word-based factors related to second language (L2) lexical acquisition and builds on recent work illustrating that L2 lexical proficiency can be investigated through the analysis of L2 learners’ knowledge and production of sense relations, lexical association models, and conceptual knowledge (Huckin and Coady, 1999; Haastrep and Henriksen, 2000; Crossley et al., 2009, 2010a, 2010b). An approach that investigates depth of lexical knowledge contrasts with past studies that have explored L2 lexical knowledge through broad measures of lexical development such as lexical accuracy, lexical frequency, and lexical diversity (Polio, 2001). Also, in contrast to most L2 lexical acquisition studies that explore learners’ written texts, the current study analyses the natural spoken output of learners.

Psycholinguistic word information relates to linguistic properties of words that affect word processing and learnability. The properties under investigation in the current article concern word meaning and are derived from the Medical Research Council (MRC) Psycholinguistic Database (Wilson, 1988). The properties were obtained from human judgments of a word’s:

- concreteness: how concrete or abstract a word is;
- imagability: how easy it is to construct a mental image of a word;
- meaningfulness: how associated a word is to other words; and
- familiarity: how commonly a word is experienced (Wilson, 1988).

The number of words available for analysis depends on the property being studied, but generally several thousand words are available.

Naturally, the words produced in the written and oral production of second language learners are found within the MRC Psycholinguistic Database. Although this should come as no surprise, what is surprising is that such a potentially valuable source of word information has not been more fully explored in past second language acquisition (SLA) research. By studying the emergence of specific words and their properties in relation to lexical development, researchers in SLA could gain valuable, new insights into how words are learned at a deeper level of meaning. Of course, not all knowledge about words is meaning based; important components of word learning that relate to form (i.e. word segment placement and prosodic features) are critical as well. However, it is the deeper-level meaning-based knowledge that is of concern in the present article.

The psycholinguistic scores for concreteness, imagability, meaningfulness, and familiarity used in this study come from an external source (i.e. the MRC Psycholinguistic Database) and are not derived from the texts themselves. In contrast, much research into L2 lexical development derives measures purely from the occurrence of word types and tokens in the target texts (Vermeer, 2000; Daller et al., 2003). These measures are referred to as indices of lexical diversity, also known as lexical variation or lexical variety (Jarvis, 2002). The majority of the indices related to lexical diversity concentrate on type token ratios (TTR), which are simple algorithms that divide the number of different word (types) by the total number of words (tokens) in a given text. There are various deviations of simple TTR
measures such as Corrected TTR (Carroll, 1964), Log TTR (Herdan, 1960), vocd (Malvern et al., 2004), Advanced TTR (Daller et al, 2003), Guiraud Advanced (Daller et al, 2003) and, most recently, the Measure of Textual Lexical Diversity (MTLD; McCarthy and Jarvis, 2010). The premise behind lexical diversity indices is that more diverse vocabularies indicate more proficient lexicons (Crossley et al., 2011, in press). Although lexical diversity indices are broad, versatile measures of lexical proficiency, they fail to account for important qualitative information about words, namely word difficulty (Vermeer, 2000; Daller et al., 2003), and, importantly, they do not measure sense relations, lexical associations, and conceptual development (Crossley et al., 2009, 2010a, 2010b).

A second type of lexical proficiency measure that does consider difficulty relates to word frequency. Generally, word difficulty is argued to correlate with word frequency counts for common words found in large corpora (Vermeer, 2000; Meara and Bell, 2001; Daller et al., 2003). The premise behind frequency measurements is that a higher lexical proficiency results in the use of less frequent and more difficult words (Nation, 1988; Meara and Bell, 2001). Two widely cited computational tools that use frequency scores to measure lexical proficiency are Lexical Frequency Profiles (LFP; Nation and Heatley, 1996) and P-Lex (Meara and Bell, 2001). The theory underlying these tools is that the texts produced by learners of different proficiencies will demonstrate different lexical profiles and that as learners’ vocabularies develop, lexical profiles will change in predictable ways. Meara and Bell (2001) contend that people with more proficient vocabularies will use more infrequent words. Thus, a greater incidence of difficult words in a text relates to a more proficient L2 vocabulary.

Although studies analysing L2 lexical development using measures of lexical diversity and frequency have been of critical importance, they generally focus on surface level lexical variables (i.e. types and tokens or the most frequent thousand words) and have not concentrated on deeper level indices related to conceptual knowledge, sense relations, and lexical association models. Recently, computer analyses of deeper level lexical knowledge have become possible because of advances in computational linguistics, corpus linguistics, and other related fields. Researchers are now able to perform computer analyses of texts beyond shallow or local components and instead investigate deeper, global components of a text (Graesser et al., 2004). Recent examples include Crossley et al. (2009) who investigated L2 conceptual development using WordNet (Miller et al., 1990), Crossley et al. (2010a) who investigated the growth of semantic co-referentiality using Latent Semantic Analysis (Landauer, 2007), and Crossley et al. (2010b) who investigated the development of polysemy networks in L2 learners.

The current article is meant to extend the findings of Crossley et al. (2009, 2010a, 2010b) by addressing depth of L2 lexical knowledge through an exploration of word meaning relations using psycholinguistic word information. Key to this study is the use of the lexical indices concreteness, imagability, meaningfulness, and familiarity taken from the MRC Psycholinguistic Database (Wilson, 1988). The purpose behind this study is simple: to demonstrate that there are significant changes in the psycholinguistic properties of L2 learners’ word use as they develop their lexical abilities over time. Such a demonstration would allow us to better understand second language acquisition and lexical processing. Additionally, work in this area is expected to stimulate future research concerning the creation of lexical meaning and the development of lexical proficiency.
In the current article, we begin by asking what psycholinguistic elements affect word learning. Previous work in first language (L1) research using psycholinguistic measures for word learnability have addressed similar questions, but such methods have not been applied to L2 research. For example, L1 researchers have long understood that words can be scored by native speakers for level of concreteness, which is the degree to which a word refers to here-and-now concepts, ideas, and things (Paivio et al., 1968; Toglia and Battig, 1978; Gilhooly and Logie, 1980). The concreteness of a word has implications for that word’s learnability: concrete words are easier to learn than abstract words. In the memory literature, the advantage that concrete words have over abstract words is referred to as the ‘concreteness effect’ (Paivio, 1991). Other dimensions of word learnability are explored in the psycholinguistic literature as well. Commonly studied dimensions are imagability, meaningfulness, and familiarity. We address each of these dimensions or attributes of words in the following sections of the article, and we ask whether such attributes impact the learnability and acquisition of words for L2 learners.

The most common repository for information on English word attributes is the MRC Psycholinguistic Database (Wilson, 1988). The database includes scores in 26 different linguistic and psycholinguistic attributes for as many as 150,837 English words although many of the attributes in the database contain far fewer words for analysis. This is the case for the four attributes chosen for the current article, which reports on scores for concreteness, imagability, meaningfulness, and familiarity. A total of 4,293 words are available for analyses of concreteness scores; 4,825 words are available for analyses of imagability scores; 2,627 words are available for analyses of meaningfulness scores (the Colorado norms); and 4,920 words are available for analyses of familiarity scores. The database has been updated to its current total of 150,837 words although the increase in word scores was not uniform across all 26 attributes. Thus the total number of words available to analyse for concreteness, imagability, meaningfulness, and familiarity has remained unchanged because the source files from which the scores are derived, as discussed below, have not changed.

**Concreteness:** The MRC database obtains its concreteness scores from earlier studies on word information by Paivio et al. (1968), Toglia and Battig (1978), and Gilhooly and Logie (1980). Researchers in these early studies asked participants to score words based on concreteness on an interval scale (from 1 to 7). A word that refers to an object, material, or person receives a higher concreteness score than an abstract word (Toglia and Battig, 1978). One of the most robust findings in memory literature relates to the advantage that concrete words have over abstract words in tasks involving recall, word recognition, lexical decision, pronunciation, and comprehension (Gee et al., 1999; for a review, see Paivio, 1991). Similarly for the learning of second language vocabulary, concrete words are learned earlier (Crossley et al., 2009) and more easily than abstract words (Ellis and Beaton, 1993).

**Imagability:** The MRC database obtains imagability scores on words from Paivio et al. (1968), Toglia and Battig (1978), and Gilhooly and Logie (1980). For example, the database reports results from Toglia and Battig (1978) who asked native English-speaking...
participants to score 480 words on an interval scale from 1 to 7 according to the ease with which mental images were produced by each item. A high-imagery word such as *buffalo* evokes images easily and would thus be scored highly on the scale. A word such as *relevant* produces a mental image with difficulty and would thus be scored lower on the scale. Some words (but not all) can score highly in both concreteness and imagability. For instance, the words *accident* and *marriage* do not refer as strongly to specific objects, materials, or people (for the concreteness score), yet they evoke strong sensory experience or mental pictures for the participants who scored them. Thus, the words score higher on the imagery dimension than on the concreteness dimension. Imagability scores are important because a word or concept that arouses a mental image quickly and easily is more likely to be recalled. Ellis and Beaton (1993) found that high imagery words were good candidates for keyword techniques in second language vocabulary learning. Additionally, highly imagable words have more context availability (Schwanenflugel, 1991) because they are experienced and analysed visually (Ellis and Beaton, 1993). Context availability and visual analysis facilitate the learning of words for L2 learners.

c  **Meaningfulness:** The MRC database derives meaningfulness scores from Paivio et al. (1968) and Toglia and Battig (1978). The Paivio et al. (1968) norms and the Toglia and Battig (1978) norms for meaningfulness have a low correlation of .53 (Wilson, 1988). Because of this low correlation, the MRC database reports the Paivio et al. scores separately from the Toglia and Battig scores. The current article uses the Toglia and Battig norms because this set of scores reports on more words (2,627). The Paivio et al. norms report scores for only 760 words.

In the Toglia and Battig norms, participants were instructed to score words on an interval scale from 1 to 7 according to how strongly they felt the words were associated in meaning to other words. Words with high meaningfulness scores were highly associated with other words (e.g. *person*) whereas a low meaningfulness score indicated that the word was weakly associated to other words (e.g. *amorphous*). Ellis and Beaton (1993) explain that the semantic links between words further mediate the organization and memory of words. Nouns in general – but more specifically nouns with many internal relationships to each other – are acquired more easily in first languages than in second languages (Ellis and Beaton, 1993). This is assumedly because L2 learners tend to have fewer and more varied word associations than do native speakers (Meara, 1983; Schmitt and Meara, 1997; Meara and Schur, 2002; Zareva, 2007).

d  **Familiarity:** The MRC database derives familiarity scores from Toglia and Battig (1978) and Gilhooly and Logie (1980). Participants were instructed to score words on an interval scale from 1 to 7 based on how familiar they were with the given word. Higher scores indicate greater familiarity. For example, the word *adze* received a low mean familiarity score of only 2.12 while the familiar word *eating* had a mean score of 6.71 (Toglia and Battig; 1978). Word familiarity likely highlights the incremental nature of vocabulary learning (Schmitt and Meara, 1997) in that vocabulary proficiency develops over time as more familiar words are acquired. This trend results from L2 learners attending to a word they may hear in several different contexts before actually using that word. Laufer and Paribakht (1998) distinguish free-active knowledge of words for use in spontaneous oral
discourse from passive word knowledge, which is largely receptive knowledge. In other words, L2 speakers of a language may have receptive but not productive knowledge of many words based on the words’ familiarity (Schmitt and Meara, 1997). The question in this article is whether more familiar words are acquired earlier than less familiar words.

The MRC database also provides 22 additional indices of lexical attribute, but we did not investigate them for this study because they were highly associated with surface level features of the lexicon. For instance, the MRC database provides scores for letter incidence, phoneme incidence, syllable incidence, frequency, age of acquisition (AOA), word derivation, syntactic category, affixes and multi-word phrasal units, word status, variants of words with the same spelling but different pronunciations, capitalization, irregular plurals, word strings, phonetic transcription, and stress. Of these properties, frequency was the most relevant for this study; however, frequency scores alone, without further interpretation regarding word difficulty, fail to directly address depth of lexical knowledge (Vermeer, 2000). Age of acquisition (AOA) was also of interest initially. However, these scores are derived from guesses of participants and not the actual measured age at which words are acquired. Thus the attribute is not valid for this study. In sum, we chose four psycholinguistic properties that stood the greatest potential for measuring depth of adult L2 vocabulary knowledge.

2 Purpose

The purpose of the current study is to explore productive vocabularies in the oral discourse of L2 English learners using psycholinguistic indices of word knowledge taken from the MRC database. We subsequently use these indices to explore the development of word knowledge and acquisition in L2 learners. We measure word use along the four dimensions of concreteness, imagability, meaningfulness, and familiarity. Each of these four dimensions is studied in naturally occurring discourse in which learners negotiate the meanings of words and ideas with their interlocutors in regularly recorded sessions over a year of observation. From these sessions, we compiled a corpus of L2 oral data organized by learner. The longitudinal design with which we test our hypotheses responds to the call for more longitudinal, developmental work in the area of lexical development (Schmitt and Meara, 1997; Vermeer, 2000; Bogaards, 2001). Specifically, we test the following:

- In reference to concreteness, we predict that the oral discourse of L2 learners will exhibit a greater number of lexical items that are more abstract (less concrete) as time progresses and oral proficiency improves.
- The second hypothesis concerns imagability. We predict that words that score low for imagability (e.g. words that arouse mental images with difficulty) will become more frequent in the oral discourse of L2 learners as time progresses.
- In regards to meaningfulness, we predict that words that score low for meaningfulness (e.g. words that are less associated to other words) will become more frequent in the L2 oral discourse as time progresses.
- The fourth and final hypothesis to be tested relates to word familiarity. Our prediction is that less familiar words will occur with increasing frequency in L2 productive vocabularies.
II Methods

1 Participant selection

L2 English learners were interviewed every two weeks (not including program and university breaks) for a total of 18 sessions over a one-year period. The participants were all enrolled in an intensive English program, and the bi-weekly elicitation sessions were scheduled to coincide with the participants’ regular speaking class. Participants began the study during the first level of a 6-level program and continued their participation in the study as they progressed through the program. The current article 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 participant was given a pseudonym; this article 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.

Interviewers were recruited from a graduate-level course in second language acquisition and were provided with all of the elicitation materials. These included emotion cards (Rintell, 1989), picture description tasks, questions, and topics for discussion. The participants were also free to introduce their own spontaneous topics into the conversation. Language data was collected in naturalistic settings; that is, while interviewers came to the sessions prepared with various topics from which the learners could choose, the sessions were characterized by naturally occurring discourse. Elicitation sessions generally lasted between 30 and 45 minutes. The sessions were tape recorded and later transcribed.

2 Corpus

The transcribed data collected from the six learners forms the foundation for this analysis. Descriptive data for the corpora of each learner can be found in Table 1. 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, such as a word in the learner’s native language or an invented word. Non-target like forms of irregular past tense verbs were included (e.g. taked, slepted), as were forms such as wanna and gonna. In most cases proper nouns were also left in the data. All punctuation except the period and question mark was eliminated from the transcriptions. Periods were used to mark the end of thoughts or ideas. Each elicitation session was saved as a single text file containing the oral production of only the learner in focus, not the interviewer or other learners participating in the session. The text file was checked for spelling errors.

A longitudinal design was essential for this study because it allowed us to explore depth of word knowledge for individual learners and the learners as a group. Each learner corpus contains thousands of word tokens with a large number and variety of word types collected over a year of observation. The corpus provides a better picture of learner growth over time than a cross-sectional study, which would have collected substantially fewer word types and tokens per learner even though the sample size would have been
much greater. Although a longitudinal design presents more obstacles to obtaining sufficient number of participants, the statistical power of longitudinal studies comes from having a greater number of observations (data points) per participant, rather than large numbers of participants. In the current study, we collected a total of 99 oral texts corresponding to 99 observations of six learners over a period of one year.

3 Coh-metrix

To collect measurements from the MRC Psycholinguistic Database, each text file was analysed using the computational tool Coh-Metrix. Coh-Metrix measures cohesion and text difficulty at various levels of language, discourse, and conceptual analysis. The tool was designed with the goal of improving reading comprehension in classrooms by providing a means to improve textbook writing and to more appropriately match textbooks to the intended students (Graesser et al., 2004). The system integrates lexicons, pattern classifiers, part-of-speech taggers, syntactic parsers, shallow semantic interpreters, and other components that have been developed in the field of computational linguistics (Jurafsky and Martin, 2002).

4 Statistical analysis

One potential drawback to using the MRC Psycholinguistic Database to study L2 vocabulary development is the prospect of little overlap between the words shared in the L2 learners’ corpora and the MRC database (Ellis and Beaton, 1993). Thus, we first analysed the incidence of shared words between our L2 spoken corpus and the MRC database. Next, we conducted a series of repeated measure Analyses of Variance (ANOVAs) using the MRC database results from Coh-Metrix. We conducted these ANOVAs to test the assumption that as the learner spends time studying English, word sense relations develop as well. The ANOVA tests track the linear trend of the MRC database scores over the increasing temporal intervals. Thus, the temporal intervals act as the independent variable, and the MRC values function as the dependent variables. Because all participants did not share all the same temporal data points, the ANOVA test analysed 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

<table>
<thead>
<tr>
<th>Learner</th>
<th>Number of meetings</th>
<th>Average number of words per conversational turn</th>
<th>Average number of conversational turns</th>
<th>Average number of words per session</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eun Hui</td>
<td>18</td>
<td>21.31</td>
<td>52.17</td>
<td>1,120.67</td>
</tr>
<tr>
<td>Faisal</td>
<td>13</td>
<td>33.42</td>
<td>71.08</td>
<td>1,870.07</td>
</tr>
<tr>
<td>Takako</td>
<td>18</td>
<td>19.40</td>
<td>51.00</td>
<td>1,470.72</td>
</tr>
<tr>
<td>Kamal</td>
<td>15</td>
<td>23.75</td>
<td>50.27</td>
<td>1,216.20</td>
</tr>
<tr>
<td>Jalil</td>
<td>17</td>
<td>38.82</td>
<td>61.76</td>
<td>2,359.77</td>
</tr>
<tr>
<td>Marta</td>
<td>18</td>
<td>33.31</td>
<td>63.61</td>
<td>1,912.00</td>
</tr>
</tbody>
</table>
data was available for the first two weeks and the last two weeks for all six learners, they were included. These data points were analysed with data from the 16th week and the 32nd week as well. The ANOVA was supplemented with a post-hoc test of within-participants contrasts in order to examine where in the temporal progression significant differences in output could be identified. A total of 99 oral texts (corresponding to each meeting with the learners) were produced over the year of observation; this yielded the same number of data points from which we ran the analyses of variance. The first column of Table 1 shows the number of meetings (data points) by learner.

### Results

#### MRC database and L2 corpus word comparisons

To address the concern that there would be little overlap between the words contained in the L2 corpora and the MRC database, we compared the words for each of the four attributes that we studied in the MRC database to the words produced by the L2 learners. We accomplished this by examining the L2 corpora and the MRC database for instances of shared words. We then computed the percentage of MRC words contained in the L2 corpus as well as the percentage of L2 words contained in the MRC corpus. Lastly, we normalized the incidence of all words found in the L2 corpus and then computed a frequency of word use for all MRC words (type and token) found in the L2 corpus. This was accomplished by calculating the frequency of each word in the L2 corpus and dividing that frequency by the total number of word tokens in the corpus. The normed frequencies for each word in the L2 corpus that were contained in the MRC database were then added together to provide a calculation of the overall incidence of MRC words in the L2 corpus. Table 2 summarizes these comparisons.

This analysis demonstrated a surprisingly high overlap between the two data sets. Overall, between 30%–38% of the words appearing in the L2 corpus also appeared in the MRC database. In addition, between 27%–39% of the 3,690 word types in the L2 corpus were analysed by the MRC database. For example, the concreteness analysis examined 1,310 words from the L2 corpus and accounted for 35.5% of the total word types in the corpus. Most importantly, however, were the normed frequencies for the overall incidence of MRC words found in the L2 corpus. These values ranged from 80% to 85%.

<table>
<thead>
<tr>
<th></th>
<th>Total word types in MRC database</th>
<th>Total word types in L2 corpus</th>
<th>Total MRC words contained in L2 corpus</th>
<th>L2 words contained in MRC database (percentage)</th>
<th>MRC words contained in L2 corpus (percentage)</th>
<th>Normed frequency for MRC word use in L2 corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concreteness</td>
<td>4,293</td>
<td>3,690</td>
<td>1,310</td>
<td>30.5</td>
<td>35.5</td>
<td>83.3</td>
</tr>
<tr>
<td>Imagability</td>
<td>4,825</td>
<td>3,690</td>
<td>1,441</td>
<td>30.0</td>
<td>39.1</td>
<td>85.1</td>
</tr>
<tr>
<td>Meaningfulness</td>
<td>2,627</td>
<td>3,690</td>
<td>995</td>
<td>37.9</td>
<td>27.0</td>
<td>79.8</td>
</tr>
<tr>
<td>Familiarity</td>
<td>4,920</td>
<td>3,690</td>
<td>1,454</td>
<td>30.0</td>
<td>39.4</td>
<td>85.1</td>
</tr>
</tbody>
</table>

Table 2 Shared words between L2 corpus and MRC database
This means that the MRC analysis in this study represents well over 80% of the total oral output of the learners. This percentage, as well as the other percentages, seems sufficiently high to be representative of the corpus as a whole.

2 Concreteness

ANOVA results comparing the three trimesters of learning to the MRC findings for concreteness show that MRC concreteness scores decreased as a function of time, $F(5, 25) = 9.42, p < .01$ (for more details, see Table 3).\(^3\) Within-participants contrasts demonstrated that the concreteness scores of the 52nd week were significantly different from the 2nd week ($F(5, 25) = 20.16, p < .01$). Significant differences were also noted between the 2nd week and the 32nd week ($F(5, 25) = 19.79, p < .01$) and the 2nd week and the 50th week ($F(5, 25) = 30.90, p < .01$) (for details, see Table 4). These findings provide evidence that concreteness scores decrease with time spent learning English.

3 Imagability

ANOVA results comparing the three trimesters of learning to the MRC findings for imagability show that MRC imagability scores significantly decreased as a function of time ($F(5, 25) = 11.20, p < .01$) (for more details, see Table 5). Within-participants contrasts demonstrated that the imagability scores of the 52nd week were significantly different from the 2nd week ($F(5, 25) = 14.22, p = .01$). Significant differences were also noted between the 2nd week and the 16th week ($F(5, 25) = 6.55, p = .05$), the 2nd week and the 32nd week ($F(5, 25) = 22.75, p < .01$), and the 2nd week and the 50th week ($F(5,

---

Table 3 Mean and standard deviations for concreteness scores

<table>
<thead>
<tr>
<th>Week</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>388.30</td>
<td>19.23</td>
</tr>
<tr>
<td>4</td>
<td>384.19</td>
<td>19.85</td>
</tr>
<tr>
<td>16</td>
<td>371.45</td>
<td>15.13</td>
</tr>
<tr>
<td>32</td>
<td>363.30</td>
<td>21.47</td>
</tr>
<tr>
<td>50</td>
<td>349.61</td>
<td>15.76</td>
</tr>
<tr>
<td>52</td>
<td>349.51</td>
<td>20.54</td>
</tr>
</tbody>
</table>

Table 4 Tests of within-participants contrasts (measure: concreteness)

<table>
<thead>
<tr>
<th>Week distinction</th>
<th>$F(5,25)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 to 4</td>
<td>0.39</td>
<td>0.56</td>
</tr>
<tr>
<td>2 to 16</td>
<td>4.91</td>
<td>0.07</td>
</tr>
<tr>
<td>2 to 32</td>
<td>19.79</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>2 to 50</td>
<td>30.90</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>2 to 52</td>
<td>20.16</td>
<td>&lt; 0.01</td>
</tr>
</tbody>
</table>

Downloaded from sir.sagepub.com at GEORGIA STATE UNIVERSITY on May 25, 2011
These findings provide evidence that imagability scores decrease with time spent learning English.

**Table 5** Mean and standard deviations for imagability scores

<table>
<thead>
<tr>
<th>Week</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>423.77</td>
<td>20.71</td>
</tr>
<tr>
<td>4</td>
<td>422.72</td>
<td>21.18</td>
</tr>
<tr>
<td>16</td>
<td>403.25</td>
<td>16.40</td>
</tr>
<tr>
<td>32</td>
<td>393.97</td>
<td>19.00</td>
</tr>
<tr>
<td>50</td>
<td>381.21</td>
<td>14.03</td>
</tr>
<tr>
<td>52</td>
<td>382.45</td>
<td>19.25</td>
</tr>
</tbody>
</table>

**Table 6** Tests of within-participants contrasts (measure: imagability)

<table>
<thead>
<tr>
<th>Week distinction</th>
<th>$F(5,25)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 to 4</td>
<td>0.24</td>
<td>0.88</td>
</tr>
<tr>
<td>2 to 16</td>
<td>6.55</td>
<td>0.05</td>
</tr>
<tr>
<td>2 to 32</td>
<td>22.75</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>2 to 50</td>
<td>23.13</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>2 to 52</td>
<td>14.22</td>
<td>0.01</td>
</tr>
</tbody>
</table>

**Table 7** Mean and standard deviations for meaningfulness scores

<table>
<thead>
<tr>
<th>Week</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>452.43</td>
<td>13.78</td>
</tr>
<tr>
<td>4</td>
<td>456.09</td>
<td>7.71</td>
</tr>
<tr>
<td>16</td>
<td>445.85</td>
<td>13.02</td>
</tr>
<tr>
<td>32</td>
<td>436.07</td>
<td>12.54</td>
</tr>
<tr>
<td>50</td>
<td>426.09</td>
<td>11.44</td>
</tr>
<tr>
<td>52</td>
<td>424.43</td>
<td>11.53</td>
</tr>
</tbody>
</table>

$25) = 23.13, p < .01$ (for details, see Table 6). These findings provide evidence that imagability scores decrease with time spent learning English.

## 4 Meaningfulness

ANOVA results comparing the three trimesters of learning to the MRC findings for word meaningfulness show that MRC meaningfulness scores decreased as a function of time ($F(5, 25) = 14.09, p < .01$) (for details, see Table 7). Within-participants contrasts demonstrated that the meaningfulness scores of the 52nd week were significantly different from the 2nd week ($F(5, 25) = 19.40, p < .01$). Significant differences were also noted between the 2nd week and the 50th week ($F(5, 25) = 11.62, p = .02$) (for details, see Table 8). These findings provide evidence that meaningfulness scores decrease with time spent learning English.
In comparing MRC familiarity scores to the three trimesters of learning, a Mauchly’s test indicated that the assumption of sphericity in the ANOVA had been violated ($\chi^2(5) = 29.89, p = .02$); therefore, the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($E = .41$) and a Bonferroni confidence adjustment. The results showed that familiarity scores demonstrated no significant differences as a function of time, $F(2.045, 10.26) = 2.31, p = .15$ (for details, see Table 9). A test of within-participants contrast demonstrated that the familiarity scores of the 52nd week were not significantly different from the 2nd week, $F(1, 5) = 2.13, p = .20$ (for details, see Table 10). These findings demonstrate that as L2 students study English their word familiarity scores do not decrease.

### Table 8 Tests of within-participants contrasts (measure: meaningfulness)

<table>
<thead>
<tr>
<th>Week distinction</th>
<th>$F(5,25)$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 to 4</td>
<td>1.27</td>
<td>0.31</td>
</tr>
<tr>
<td>2 to 16</td>
<td>1.08</td>
<td>0.35</td>
</tr>
<tr>
<td>2 to 32</td>
<td>6.24</td>
<td>0.06</td>
</tr>
<tr>
<td>2 to 50</td>
<td>11.62</td>
<td>0.02</td>
</tr>
<tr>
<td>2 to 52</td>
<td>19.40</td>
<td>&lt; .01</td>
</tr>
</tbody>
</table>

### Table 9 Mean and standard deviations for familiarity scores

<table>
<thead>
<tr>
<th>Week</th>
<th>Mean</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>593.06</td>
<td>3.85</td>
</tr>
<tr>
<td>4</td>
<td>594.15</td>
<td>2.72</td>
</tr>
<tr>
<td>16</td>
<td>591.12</td>
<td>4.60</td>
</tr>
<tr>
<td>32</td>
<td>590.50</td>
<td>6.13</td>
</tr>
<tr>
<td>50</td>
<td>588.83</td>
<td>3.68</td>
</tr>
<tr>
<td>52</td>
<td>589.37</td>
<td>4.12</td>
</tr>
</tbody>
</table>

### Table 10 Tests of within-participants contrasts (measure: familiarity)

<table>
<thead>
<tr>
<th>Week distinction</th>
<th>$F$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 to 4</td>
<td>0.46</td>
<td>0.53</td>
</tr>
<tr>
<td>2 to 16</td>
<td>0.52</td>
<td>0.50</td>
</tr>
<tr>
<td>2 to 32</td>
<td>0.66</td>
<td>0.45</td>
</tr>
<tr>
<td>2 to 50</td>
<td>2.98</td>
<td>0.15</td>
</tr>
<tr>
<td>2 to 52</td>
<td>2.13</td>
<td>0.20</td>
</tr>
</tbody>
</table>

5 Familiarity

In comparing MRC familiarity scores to the three trimesters of learning, a Mauchly’s test indicated that the assumption of sphericity in the ANOVA had been violated ($\chi^2(5) = 29.89, p = .02$); therefore, the degrees of freedom were corrected using Greenhouse–Geisser estimates of sphericity ($E = .41$) and a Bonferroni confidence adjustment. The results showed that familiarity scores demonstrated no significant differences as a function of time, $F(2.045, 10.26) = 2.31, p = .15$ (for details, see Table 9). A test of within-participants contrast demonstrated that the familiarity scores of the 52nd week were not significantly different from the 2nd week, $F(1, 5) = 2.13, p = .20$ (for details, see Table 10). These findings demonstrate that as L2 students study English their word familiarity scores do not decrease.

IV Discussion

Results from the repeated measures ANOVA tests confirm three predictions: the learners’ meaningfulness, concreteness, and imagability scores decreased significantly over the
course of a year. The prediction that familiarity scores would decrease was not supported. We address each of these results in turn.

1. **Concreteness**

The statistical analysis demonstrated that as L2 learners progress they exhibit a greater use of abstract words – or less concreteness – over time. The oral discourse of the learners showed a significant decrease in concreteness scores by the 32nd week of the year-long observation. The results of the current study show that learners are better able to use abstract language as their proficiency increases. These results parallel findings by Crossley et al. (2009) with hypernymy relations. Their study found a positive correlation between hypernymy values and concreteness scores, suggesting that the L2 learners’ oral discourse displays a greater range in hypernymic relations while at the same time evidencing a greater amount of less specific language. These findings demonstrate that as L2 learners’ oral proficiency develops, they are less dependent in their oral discourse on the inherent context created by words with high concreteness scores. Examples from the data are provided below. The examples were produced by Takako in her final elicitation session. Her utterances contained more abstract words such as *weird* (found in example 1) and *bothered* (found in example 2). The concreteness scores in the MRC database for *weird* and *bothered* are 253 and 267 respectively. Although these are only two words, they illustrate a general pattern across learners of discourse that exhibits a greater number of abstract words.

1. I stopped talking, and I just walk slow, but he still like saying something, like, where you from, and ah, can I call you, like, so weird
2. She kind of bothered with this relationship

The ability to use abstract language is fundamental. Our data reveal that the movement longitudinally from the concrete to the abstract in word development is clearly important in understanding how learners acquire second language vocabularies. In addition, our analysis reveals that concreteness in L2 oral data is largely independent of topic. Early elicitation sessions, when learners used significantly more concrete words, were not characterized by topics that would naturally elicit more concrete language. For example, in the second elicitation session, still in the first month of the study, learners were given a set of topic cards containing single words such as *politics*, *love* and *wishes* and a set of emotion cards with single words such as *lonely*, *confused* and *scared*. These topics are quite abstract and thus likely to elicit abstract language if the speaker were sufficiently proficient in the language. However, this was not the case. It seems apparent that the increase in abstract words over time is unrelated to the task.

2. **Imagability**

The statistical analysis demonstrated that as L2 learners developed lexical proficiency, they showed significant decreases, as predicted, in imagability scores. These scores relate to the image that a word evokes in the mind. Words with high imagability scores evoke a mental image with ease and are thought to be more contextual (Ellis and Beaton, 1993) relative to words with low imagability scores. That L2 learners’ imagability scores
decrease over time spent learning English indicates that their oral discourse becomes less context dependent with an increase in proficiency.

Examples of words scored for imagability later in the year of observation are provided below. The examples were produced by Jalil in his penultimate elicitation session. His utterances contained words that are less imagable such as control (found in example 3) and theory (illustrated in example 4). In the MRC database, the imagability score for control is 347 and the score for theory is 317.

3. Like we can control what you be like
4. Like I I trying to make a theory now

3 Meaningfulness

As learners become more proficient in the L2, they use words that have fewer associations to other words. At first glance this appears counterintuitive because larger vocabularies lead to a greater variety and number of word associations (Zareva, 2007). We maintain that meaningfulness is related to word difficulty; less meaningful words are less associated to other words and therefore more difficult to learn. More proficient learners are able to use words with fewer associations even though the overall network of word associations in the mental lexicon is larger and more interconnected. Thus, meaningfulness scores highlight the process by which lexical networks develop from mostly random associations with a few large components (lexical items) to networks that are smaller, more constrained and more tightly connected (Meara and Schur, 2002).

Support for this interpretation comes from a recent study by Zareva (2007) who found quantitative differences between intermediate L2 learners of English and native English speakers in the total number of word associations they provided. Advanced L2 learners displayed patterns that were similar to native speakers and both groups provided significantly more word associations than intermediate L2 learners. Zareva argued that L2 learners with smaller vocabulary sizes (6,000 words on average) exhibit mental lexicons that are loosely connected (fewer links among words) in comparison to native speakers, whose larger vocabularies permit them to associate all of the words that they know to many other words. In sum, larger vocabularies are better connected in terms of the number and variety of connections (Zareva, 2007).

Lower meaningfulness scores are illustrated in the following examples. In (5), produced by Eun Hui in the first elicitation session, the word pets appears.

5. pets yes. I have very very dog.

The meaningfulness score in the MRC database for pet is 582. In her last elicitation session in Week 52, Eun Hui produces words with lower meaningfulness scores, such as link, illustrated in (6). The meaningfulness score for link is 324.

6. so he tried he tried to link with another office person

The use of less meaningful (less associated) words indicates that lexical items are becoming more specialized as the learners’ lexical networks develop over time. New
input constrains the meanings and associations of already learned forms, and new words are acquired to fill the gaps.

4  Familiarity

The prediction that learners would use significantly fewer familiar words, as scored by native speakers in the MRC database, was not borne out in the data. Our data consisted of words that L2 learners use, not words for which they may have receptive but not productive knowledge. Thus, the L2 learners in this study were unlikely to use words which native speakers would find unfamiliar. The small changes (decreases) in familiarity scores for word choices over time would not be detected and the analysis would tend to yield non-significant results, unlike the larger differences observed in the measures for concreteness, imagability, and meaningfulness. Perhaps a longer period of observation, more than one year, is necessary to see significant changes in familiarity measures with L2 learners. In addition, the fact that the data was oral, not written, is a potential second reason for the lack of significance in the analysis for familiarity scores. Oral data would naturally tend to yield a higher number of the most frequent, familiar words in English, which learners must learn early in their language development and for which teachers put the most emphasis (Nation, 2005).

V  Conclusions

This study has demonstrated that L2 oral discourse appears to be characterized by a decrease in words with a high number of semantic links (meaningfulness or word association scores), a decrease in the number of concrete words, and a decrease in words with high imagability scores. Taken together, these results suggest that L2 learners’ vocabulary use in oral discourse becomes less dependent on physical, visual, and semantic contexts. Similar processes are found in the emergence and development of L2 learners’ grammars in what Giacalone Ramat (1992) has termed ‘acquisitional grammaticalization’, where learners’ grammars move from lexical strategies that are highly context dependent to a gradual acquisition of the morphological devices of the target language. It is noteworthy that development of the lexicon parallels the movement away from context dependence found in studies on L2 grammatical development. If indeed language emerges through discourse – the negotiating of ideas and language with other native and non-native speakers – then understanding the factors that play into word knowledge and how that knowledge is manifested in real communication is a critical piece to understanding second language acquisition.

While the study provides support for notions of L2 lexical development, it also raises many interesting questions and future avenues of research. The first, and most important, is the notion of word learnability. The results from the current article provide strong evidence that psycholinguistic properties of words impact the learnability of those words. Specifically, words with higher scores for concreteness, imagability, and meaningfulness appear to carry lower learning burdens in L2 lexical acquisition and therefore tend to emerge earlier than words with lower scores on these same indices. Future research should attempt to replicate these results with larger populations of learners as well as with younger L2 learners. Such research would also have an important impact on pedagogy.
Another follow-up to the current study might involve qualitative analysis of individual word associations to study the process of learning new meanings for already known forms or combinations of forms, semantic relations, morphological relations, correct use of words at the level of grammar, collocational use and appropriate use of words at the level of pragmatics and discourse (Bogaards, 2001). Previous research in L2 lexical acquisition has tended toward quantitative analysis in studies on breadth of lexical knowledge and qualitative analysis in studies on depth of lexical knowledge (Read, 1993; Daller et al., 2003). The current study using psycholinguistic measures that affect word learning, however, shows that it is possible to explore depth of knowledge on very large samples of words using quantitative analyses.

This study, however, is not without limitations. A drawback to data of this sort is that we have limited control over the topics that learners talk about, even with prepared topic cards from which learners choose. We have argued that topic was not a factor in the scores obtained for the learners; however, we acknowledge that some topics will lend themselves to more or less concrete language and imagability, irrespective of the learner’s level of proficiency or development in L2 vocabulary. Future studies might seek to better control for topic and thus more thoroughly test our hypotheses. For example, we are currently conducting research with child English language learners in elementary math discussion groups. Topics in this study center on mathematical reasoning and are thus more uniform than the study presented in this article.

In conclusion, however, we argue that quantifying lexical development through actual language use in a longitudinal design provides valuable insights into the acquisition of second language lexicons. By focusing on the productive vocabularies of learners, we free ourselves from the problem of separating what learners know from what they use. Our results suggest that psycholinguistic word information, namely concreteness, imagability, and meaningfulness play a role in what learners know about words and how they are able to use them. They provide us with sought-after evidence about the development of depth of lexical knowledge in L2 learners. Such findings, when combined with studies of surface based lexical features (Jarvis, 2002; Malvern et al., 2004; McCarthy and Jarvis, 2007) and depth of lexical knowledge (Crossley et al., 2009, 2010a, 2010b) provide us with a better understanding of the complex features that comprise lexical proficiency. Not only that, but they also lead us closer to a comprehensive definition and understanding of L2 lexical proficiency.

Acknowledgments

The research was supported in part by the Institute for Education Sciences (IES R3056020018-02; R305A080589). 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 Institute for Education Sciences. The authors are indebted to Dr Philip McCarthy of the Institute for Intelligent Systems for his assistance with the statistical analyses presented in this article.

Notes

1. The database uses unpublished norms that expand upon the norms in Paivio et al. (1968).
2. Toglia and Battig (1978) report a positive correlation of .883 between imagability and concreteness scores.
3. Note that the MRC Psycholinguistic Database reports scores from 100 to 700. This differs from the original source files, which asked participants to score words on a scale from 1 to 7.
References


