Studies in Second Language Acquisition, 2013, **35**, 727–755.
doi:10.1017/S0272263113000375

FREQUENCY EFFECTS OR CONTEXT EFFECTS IN SECOND LANGUAGE WORD LEARNING

What Predicts Early Lexical Production?

Scott A. Crossley and Nicholas Subtirelu
Georgia State University

Tom Salsbury
Washington State University

This study examines frequency, contextual diversity, and contextual distinctiveness effects in predicting produced versus not-produced frequent nouns and verbs by early second language (L2) learners of English. The study analyzes whether word frequency is the strongest predictor of early L2 word production independent of contextual diversity and distinctiveness and whether differences exist in the lexical properties of nouns and verbs that can help explain beginning-level L2 word production. The study uses machine learning algorithms to develop models that predict produced and unproduced words in L2 oral discourse. The results demonstrate that word frequency is the strongest classifier of whether a noun is produced or not produced in beginning L2 oral discourse, whereas contextual diversity is the strongest classifier of whether a verb is produced or not produced. Post hoc tests reveal that nouns are more concrete, meaningful, imageable, specific, and unambiguous than verbs, which indicates that lexical properties may explain differences in noun and verb production. Thus, whereas distributional

The authors would like to acknowledge the anonymous reviewers of this manuscript for their careful readings and ideas. The authors also wish to thank the editorial board at Studies in Second Language Acquisition along with Danielle S. McNamara at Arizona State University for their support.

Correspondence concerning this article should be addressed to Scott A. Crossley, Department of Applied Linguistics and ESL, Georgia State University, 34 Peachtree Street, Suite 1200, One Park Tower Building, Atlanta, GA 30303. E-mail: sacrossley@gmail.com

properties of nouns may allow lexical acquisition on the basis of association through exposure alone (i.e., nouns may adhere to frequency effects), the abstractness and ambiguity found in verbs make them difficult to acquire based solely on repetition. Therefore, verb acquisition may follow a principle of likely need characterized by contextual diversity effects.

The raw frequency of linguistic items in language input is a powerful predictor of language acquisition (Ellis, 2002). Frequent exposure to linguistic items can help explain syntactic, morphological, phonological, and lexical development in both a first language (L1) and a second language (L2). Approaches concerning the predictive power of frequency effects mainly rely on Zipfian distributions (Zipf, 1935) and implicit learning mechanisms (Ellis, 2002). More recently, these theories have expanded to include construction learning, which includes not only Zipfian distributions but also the importance of form (i.e., item saliency) and function (i.e., prototypicality of meaning and redundancy in cues; Boyd & Goldberg, 2009; Ellis & Collins, 2009; Ellis & Ferreira-Junior, 2009; Wulff, Ellis, Römer, Bardovi-Harlig, & LeBlanc, 2009). However, even in construction learning research, frequency effects are generally most predictive of language acquisition.

Recently, the primacy of frequency effects has been called into question in many L1 studies that examine word-naming and lexical decision tasks. Historically, research into such tasks has demonstrated that word frequency is a strong predictor of how quickly a word is (a) read (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Forster & Chambers, 1973; Frederiksen & Kroll, 1976), (b) judged to be a word versus a nonword (Forster, 1976), and (c) spelled accurately (Barry & Seymour, 1988). However, recent research indicates that other properties of words such as contextual diversity (i.e., the number of different contexts in which a word appears)[1] and contextual distinctiveness (i.e., the contextual constraints surrounding a word) are better predictors of lexical decision latencies (Adelman, Brown, & Quesada, 2006; Johns & Jones, 2008; McDonald & Shillcock, 2001) and word naming (Adelman et al., 2006; Johns & Jones, 2008) than word frequency. This line of research suggests that there are no facilitative effects for word frequency independent of both contextual diversity (Adelman et al., 2006) and contextual distinctiveness (McDonald & Shillcock, 2001) for L1 users.

The purpose of this study is to examine word frequency, contextual diversity, and contextual distinctiveness effects in explaining L2 lexical production. To investigate these lexical features, we conduct a corpus analysis of L1 and L2 spoken discourse to examine lexical differences between frequent words (both nouns and verbs) produced by beginning-level L2 learners and frequent words produced by L1 speakers but not produced by beginning-level L2 learners. We derive frequency and contextual diversity values from the SUBTLEXus (Brysbaert & New, 2009) and British National Corpora (BNC; 2001). We derive contextual distinctiveness values from the University of South Florida word association norms (USF; Nelson, McEvoy, & Schreiber, 1998) and from latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998). We use these values to classify whether a word is produced or not produced by beginning-level L2 learners. Such an approach was used by Crossley and Salsbury (2010) and affords an examination of word properties that are important in early lexical production and, by extension, early lexical acquisition. In selecting word properties related to lexical frequency, contextual diversity, and contextual distinctiveness, we are able to address the following research questions:

1. Is word frequency—independent of contextual diversity and distinctiveness—the strongest predictor of early L2 word production?
2. Do differences exist in the lexical properties of nouns and verbs that may explain early word production?

BACKGROUND

Word Frequency Effects

Every level of language (i.e., phonology, morphology, lexicon, syntax, and discourse) is characterized by disproportionate frequencies of individual items within its inventory of features. These frequencies generally follow a Zipfian distribution (Zipf, 1935) in which the most frequently occurring types account for a majority of tokens. Theoretical accounts of these Zipfian distributions hold that they aid in the process of language learning due to the increased probability of form-function mapping for more frequent tokens (Ellis, 2006; MacWhinney, 1997). These more frequent tokens may then serve as exemplars for the acquisition of other less frequent tokens (Boyd & Goldberg, 2009; Ellis & Collins, 2009). For example, Ellis and Ferreira-Junior (2009) examined L2 learner production of verb-argument constructions. They found that learners first produced the most frequent, prototypical exemplar types from the inventories of verb-argument constructions (e.g., transitive or ditransitive constructions) in English from the input to which they were exposed (e.g., they first produced *give* for ditransitive constructions and *put in* for verb-locative constructions). In the same vein, Wulff and colleagues (2009) explored the effects of frequency on the acquisition of tense and aspect in English. They found that when the verbs that occurred most frequently in a specific tense-aspect occurred consistently in that tense-aspect—and not another—then there was a facilitative learning effect based on frequency because these items could serve as exemplars. The authors point to progressive aspect as an example because, unlike other tenses or aspects (e.g., past tense), the most frequently occurring verbs in progressive aspect were consistently found in the progressive.

Frequency effects are also thought to be important in L2 lexical acquisition (Ellis, 2002). Laufer and Nation (1995) found that the production of frequent words in L2 writing decreased as a function of time spent learning English, although studies examining spoken language have not supported this (Crossley, Salsbury, & McNamara, 2010). Corresponding evidence for the strength of word frequency indices to explain L2 lexical acquisition come from a series of studies conducted by Crossley and colleagues, which have demonstrated that word frequency is an important predictor of L2 lexical proficiency (Crossley, Salsbury, McNamara, & Jarvis, 2011a, 2011b) and that word frequency is a significant classifier of proficiency level for L2 writers (Crossley, Salsbury, & McNamara, 2012). In all cases, more advanced L2 learners produced less frequent words than beginning-level L2 learners. Other studies focusing on L1 processing have demonstrated the strength of word frequency in predicting linguistic performance. For example, studies have supported the notion that high-frequency words are recognized (Kirsner, 1994) and named (Balota et al., 2004; Forster & Chambers, 1973; Frederiksen & Kroll, 1976) more rapidly.

In theory, frequency effects can help explain language acquisition because the faster processing of high-frequency items testifies to associative learning from usage. With language experience, a learner's perceptual system becomes tuned to expect linguistic items according to their probability of occurrence in the input (Ellis, 2006). Thus, language learners acquire language patterns on the basis of association through exposure. This exposure follows Zipf's law in that the frequency of any word in the corpus will be inversely proportional to its rank in a frequency table. This law indicates that, although learners may be exposed to thousands of different words, the words they are regularly exposed to will be a much smaller set of highly frequent words, which signifies a decrease in input variability. Because learners hear and read the same sets of words consistently, these words are learned and produced more quickly (Mintz, Newport, & Bever, 2002).

Contextual Diversity

Some L1 researchers have begun to question the superiority of frequency effects over other conceptually and statistically related measures in explaining lexical processing (Adelman et al., 2006; Brysbaert & New, 2009; Johns & Jones, 2008). These researchers focus instead on contextual diversity indices (i.e., the number of unique contexts in which a word occurs). Whereas contextual diversity and word frequency are highly correlated, memory research has demonstrated that it is not the number of repeated exposures to an item that affects retrieval but rather the separation of these exposures in time and context (Glenberg, 1976, 1979). Specifically, if the time and context of exposures do not change substantially, there may be no benefit to repetition on subsequent retrieval (Verkoeijen, Rikers, & Schmidt, 2004) because the more contexts in which an item occurs, the more likely it is that the item will be needed in a new context; that is, the *principle of likely need* will increase (Anderson & Milson, 1989; Anderson & Schooler, 1991). Therefore, word frequency may be a poor indicator of need because words may cluster within a particular context and may not be shared among contexts. Adelman et al. (2006) argue that if item retrieval and lexical retrieval operate similarly, then the effect of word frequency will be reduced if repeated exposures to a word happen in a similar environment. Thus, word retrieval—and possibly word acquisition—may be less related to the number of exposures to a word that a learner receives and more related to the number of different contexts in which a word is experienced (i.e., a word's contextual diversity).

To test this theory, Adelman et al. (2006) examined response times in six different word-naming and lexical decision tasks using indices of word frequency and contextual diversity. To obtain their contextual diversity index, Adelman and colleagues counted the number of documents in a corpus that contained each word in their analysis under the assumption that a greater number of unique contexts would directly relate to a word's contextual diversity. Contextual diversity was found to predict word processing times independently of word frequency. Additionally, there was no evidence of a facilitative effect for word frequency independent of contextual diversity. That is, when both contextual diversity and word frequency indices were used to measure the same data simultaneously, contextual diversity was a significant predictor of the variance in the data (for both word-naming and lexical decision tasks), but word frequency indices were not. These results suggest that contextual diversity indices are better predictors of lexical decision and word-naming tasks and that word frequency indices do not contribute to explaining human responses to these tasks when investigated concurrently with contextual diversity indices. Similar results have been found in other studies comparing the strengths of word frequency and contextual diversity indices in explaining human word response times (Johns & Jones, 2008; Steyvers & Malmberg, 2003).

Contextual Distinctiveness

Another approach to examining how context affects the principle of likely need is to examine a word's contextual distinctiveness. The underlying assumption of contextual distinctiveness is that the immediate context of a word has priority in explaining language processing effects (McDonald & Shillcock, 2001). Like indices of contextual diversity, indices of contextual distinctiveness can be derived from language corpora. However, they are operationalized as co-occurrence counts of words that are located in the immediate environment. According to this view, it is thus not the number of different texts that a word occurs in (i.e., distant context) but rather the number of words that strongly associate with a word (i.e., local context) that best predicts language processing. Context, in this sense, involves semantic representations of a word that are influenced by the immediate environment in which the word is encountered. Co-occurring words are argued to be the best indicators of this environment (McDonald & Shillcock, 2001). For instance, the word *amok* is extremely constrained by its immediate environment (usually only occurring in the bigram *run amok*). However, *run* has many fewer constraints than *amok* and occurs with a variety of different words (McDonald & Shillcock, 2001). Consequently, *amok* has high contextual distinctiveness, whereas *run* has low contextual distinctiveness.

Although it has not received the same amount of attention as contextual diversity, the effects of contextual distinctiveness have also been compared to word frequency effects in explaining lexical decision latencies. McDonald and Shillcock (2001) implemented indices of contextual distinctiveness by developing co-occurrence vectors between words in a lemmatized version of the BNC (2001) spoken subcorpus. In two different experiments that assessed lexical decision response time latencies, the index of contextual distinctiveness was a stronger predictor of human responses than the index of word frequency. These findings indicate that words that appear in relatively constrained linguistic contexts (i.e., have high distinctiveness) have longer lexical decision latencies, whereas words in unconstrained contexts (i.e., words that have low distinctiveness) have shorter response times. McDonald and Shillcock concluded that, although word frequency and contextual distinctiveness indices are highly correlated, word frequency likely approximates contextual distinctiveness. As such, contextual distinctiveness is not only a more powerful predictor of lexical response latencies but also a more theoretically compelling variable than word frequency because it is based on a word's lexical environment.

METHOD

The purpose of this study is to assess the predictive ability of computational indices related to word frequency, contextual diversity, and contextual distinctiveness to categorize nouns and verbs as either being produced or not produced by beginning-level L2 learners. Such an approach enables us to examine the effects of these lexical properties in explaining lexical production, which is unlike most previous studies that have examined word frequency, contextual diversity, and contextual distinctiveness in relation to word-naming and lexical decision tasks. Additionally, our focus is on beginning-level L2 learners and not fluent L1 speakers. Thus, we are able to examine the effects of these lexical properties on lexical production as compared to lexical processing.

We associate lexical production with lexical acquisition under the hypothesis that productive vocabulary taps into word knowledge at the processing, storage, and, most importantly, word retrieval level. For our study, the word frequency and contextual diversity indices are collected from the SUBTLEXus (Brysbaert & New, 2009) and the BNC (2001). Our contextual distinctiveness indices are taken from the USF word association norms (Nelson et al., 1998) and from LSA (Landauer et al., 1998) near-neighbor counts. Our goal is to better understand the role of word frequency, contextual diversity, and contextual distinctiveness in explaining lexical production by beginning L2 learners.

Word Lists

We used the lists of produced and unproduced words reported in Crossley and Salsbury (2010). The lists comprise (a) frequent words produced by beginning L2 learners and (b) words that are not produced by L2 learners, but that are frequently produced by native speakers of English. The basic assumption behind these word lists was that words produced by beginning-level L2 learners are easier to acquire and that words not produced by L2 learners but commonly produced by L1 speakers are more difficult to acquire.

The L2-produced word list was developed using a language learner corpus. The corpus was collected during a yearlong longitudinal study of L2 learners as reported by Salsbury (2000). The corpus has been used in a variety of lexical development studies (Crossley, Salsbury, & McNamara, 2009, 2010; Salsbury, Crossley, & McNamara, 2011) and is characterized by naturalistic speech (i.e., spoken, unprepared speech). The corpus consists of oral data collected in bimonthly (approximately every two weeks) elicitation sessions with L2 English learners enrolled in an intensive English program at a large American university. All learners were tested on arrival to the program and were placed into the lowest proficiency level of the program. The learners were also assessed through the institutional TOEFL examination. The portion of the corpus used by Crossley and Salsbury (2010) focused on the six learners who completed the yearlong study, focusing specifically on the first 3 months of the study (i.e., when standardized tests indicated the learners were at the beginning level). The learners came from various L1 backgrounds: Spanish (n = 1), Japanese (n = 1), Korean (n = 1), and Arabic (n = 3).

The average number of word tokens per elicitation session ranged from 1,100 to 2,100, which varied by learner and by session. In the first 3 months of the study (from which data for the present article were collected), learners were provided with a total of 50 unique topics about which to talk. The topics were prepared in advance and included emotion cards (e.g., *happy, sad, lonely*), topic cards (e.g., *wishes, travel, movies*), photograph descriptions (from *Life: Faces*), questions related to attitudes toward learning English, and questions about imaginary scenarios (e.g., *Imagine you have a wish for any talent or skill. What is that talent or skill?*). Learners selected the topics they wanted to discuss but typically chose all of the sessions' prepared topics and added in their own topics. Elicitation sessions lasted 30–45 min, with an average number of learner conversational turns ranging from 50 to 70. The interviewers (all native English-speaking university students) were instructed to ask follow-up questions and comment on the learners' content. Thus, the elicitation sessions were dominated by learner talk. Additionally, the elicitation sessions were informal, and participants were on a first-name basis.

To select the words for the L2 unproduced word list, Crossley and Salsbury (2010) used the SUBTLEXus corpus (Brysbaert & New, 2009). The corpus consists of 51 million words from subtitles of American television shows and films. The SUBTLEXus corpus was selected because it focuses on everyday interactions and contains oral language similar in register to the interactions recorded in the L2 corpus.

The word lists did not contain all produced and unproduced words found in the corpora but, instead, included only words that met three criteria. The first criterion was word frequency: Only those words that were within the first 3,000 most common words in the English language (as reported by SUBTLEXus) were included. This criterion was selected to ensure that the participants in the study would have most likely been exposed to the words in either instructional or naturalistic settings. The second criterion was that the words in the L2-produced list had to be accurately used by at least half of the L2 participants. The third criterion was that the words had to fit clearly into a noun or verb categorization. If questions arose, Crossley and Salsbury (2010) analyzed word use in context to ensure its part-of-speech category. The 45 most frequent verbs and nouns from each group that fit these criteria were then selected, which produced a word list of 90 nouns (i.e., 45 produced and 45 unproduced) and 90 verbs (i.e., 45 produced and 45 unproduced). Close examination revealed a repetition in the use of the word *steal* in the unproduced verbs. This extra instance of the word *steal* was removed, which left 89 verbs (45 produced and 44 unproduced).

Selected Indices

We report on a variety of indices related to word frequency, contextual diversity, and contextual distinctiveness, which will be discussed in more detail in the following section. The selected nouns and verbs along with their respective word frequency, contextual diversity, and contextual distinctiveness values are presented in Appendices A–D.

Contextual Diversity Indices

SUBTLEXus contextual diversity. We selected a contextual diversity index from the SUBTLEXus corpus (Brysbaert & New, 2009). The foundations of the SUBTLEXus corpus are subtitles taken from 2,046 U.S. films from the years 1900–1990, 3,218 U.S. films from the years 1990–2007, and 4,575 U.S. television series. The total size of the corpus is 51 million words: 16.1 million words from television shows, 14.3 million words from films before 1990, and 20.6 million words from films after 1990. The SUBTLEXus database derives a contextual diversity score for each word in the corpus by calculating the number of different television shows and films in which the words appear. We predict that words that have greater SUBTLEXus contextual diversity values will be learned and produced earlier.

British National Corpus contextual diversity. The BNC consists of 100 million words of British English collected from the latter part of the 20th century. The BNC contains roughly 90% written texts and 10% spoken texts. We divided the BNC into its spoken and written components and then lemmatized the samples by replacing each inflected token with its canonical form to ensure that all root forms of the words in the word lists were captured. We then conducted a text count (i.e., the number of text samples in which a word occurs) using the corpus linguistics tool WordSmith (Scott, 2004). We used the percentage of text samples in which a word occurs in the BNC as our BNC contextual diversity measure. From this measure, we computed three indices of contextual diversity: (a) for the entire BNC, (b) for the written subsection, and (c) for the spoken subsection. We investigated both written and spoken texts to assess potential differences in L2 input. We predict that the words that have greater BNC contextual diversity values will be acquired and produced earlier than words that do not.

Contextual Distinctiveness Indices

University of South Florida word association norms. Our first index of contextual distinctiveness was derived from the word association scores reported in the USF word association norms (Nelson et al., 1998). The USF database contains nearly 750,000 association responses for more than 5,000 words generated by 6,000 participants. In the database, word associations were collected by giving native English-speaking participants a prime word—that is, a word meant to stimulate a response—and by asking them to write the first word that came to their mind that was meaningfully related to that prime word. For our index of contextual distinctiveness, we calculated the average number of associations for each word in our word lists. We used these associations to identify the distributional information contained in the prime word under the hypothesis that the associations reflect the environment in which the word is encountered. We predict that words with many associations will be less semantically distinct (i.e., the words will occur in a greater number of local contexts) and thus lead to quicker acquisition and production.

Latent semantic analysis. Latent semantic analysis measures the semantic associations between words through a mathematical technique known as singular value decomposition (SVD), which reduces thousands of dimensions and relationships between words in large corpora to a smaller number. This procedure promotes quick calculations of associations between words and the context in which they occur (Landauer et al., 1998). These associations represent how often a word or words occur within documents. To compare words, LSA computes the cosine between the two sets of vectors reported by the SVD for each word. These vectors report values between –1 (no similarity) and 1 (high similarity). Thus, two words will acquire semantic similarity to each other—but not to the other words—if (a) the words appear in the same context and (b) other words in that context appear in many other contexts without them (Landauer & Dumais, 1997). As an example, the LSA cosine between *dog* and *cat* is .36. The LSA cosine between *dog* and *fur* is .24. The LSA cosine between *dog* and *table* is .04.

For our LSA index of contextual distinctiveness, we computed the number of associations reported for each word in our word lists that occurred at least 20 times in the Touchstone Applied Science Association (TASA) corpus[2] and had a LSA value greater than .26 using logarithmic entropy values.[3] We predict that words with more word associations will be learned and produced earlier by L2 learners.

Word Frequency Indices

SUBTLEXus word frequency. Frequency values were taken from the SUBTLEXus corpus (Brysbaert & New, 2009). As previously discussed, the SUBTLEXus corpus provides not only a contextual diversity count but also a frequency count for the number of times a word appears in the 51-million-word corpus (i.e., frequency per million words on the basis of word form). The word frequency counts taken from SUBTLEXus should better reflect the frequency of words in spoken language than more commonly available corpora of written language. We predict that L2 learners will acquire and produce words occurring more frequently in SUBTLEXus first.

British National Corpus word frequency. We derived word frequency values for each of the words in the word lists using frequency counts (i.e., frequency per thousand words on the basis of word forms) calculated from the spoken and written BNC subcorpora and from the entire BNC corpus. We used WordSmith (Scott, 2004) to calculate the frequency values using a lemmatized version of the BNC. Like the SUBTLEXus word frequency values, we predict that L2 learners will acquire and produce words that occur more frequently in the BNC corpus and subcorpora first.

Word familiarity. Word familiarity refers to how commonly a word is experienced in natural discourse (Wilson, 1988). We collected word familiarity values from the Medical Research Council (MRC) database, which reports familiarity scores for 9,392 words taken from Toglia and Battig (1978) and Gilhooly and Logie (1980). Participants in these studies were instructed to score words on an interval scale from 1 to 7 on the basis of how familiar they were with the given word; higher scores indicated greater familiarity. For example, the word *adze* received a low mean familiarity score of only 2.12, whereas the word *eat* had a mean score of 6.71. Researchers have interpreted word familiarity scores as measures of word exposure because the scores reflect word distribution patterns in spoken data (demonstrating a bias toward natural exposure; Stadthagen-Gonzalez & Davis, 2006). Thus, word familiarity is often associated with indices of word frequency (e.g., Schmitt & Meara, 1997), and some research indicates that frequency effects are actually the result of spurious correlations between frequency and familiarity (Gernsbacher, 1984). We predict that more familiar words should be learned and produced first by L2 learners.

Statistical Analysis

For each word list (nouns or verbs), we first conducted a MANOVA to examine if the contextual diversity, contextual distinctiveness, and word frequency indices demonstrated significant differences between the produced and unproduced words. We next conducted a stepwise discriminant function analysis (DFA) using selected indices for each word list that demonstrated significant differences between the produced and unproduced words but did not exhibit multicollinearity with other indices in the set.[4] The DFA generates a discriminant function, which provides an algorithm to predict group membership (i.e., whether or not the words were produced).

We first used the DFA on the entire word list and then used the DFA model reported for the entire set to predict group membership of the words in the word list using leave-one-out cross-validation (LOOCV). For LOOCV, we selected a fixed number of folds that equaled the number of observations (i.e., words). In the cross-validation, one observation is left out, and the remaining instances are used as the training set (in this case either the 89 remaining words in the noun word list or the 88 remaining words in the verb word list). We then tested the accuracy of the model by examining the model's ability to predict the proficiency classification of the omitted instance. Thus, we were able to test the accuracy of the model on an independent data set (i.e., on data that is not used to train the model). If the results of the discriminant analysis in both the entire set and the *n* fold cross-validation set are similar, then we have increased confidence that the findings support the extension of the analysis to external data sets.

We also report the results of the LOOCV using recall, precision, and F1 scores. Recall scores are computed by tallying the number of hits (i.e., correct predictions) over the number of hits + false negatives (i.e., the number of produced words that were misclassified as unproduced words). Precision is the number of hits divided by the number of hits + false positives (i.e., the number of unproduced words that were classified as produced words). The precision and recall scores allow us to better understand the accuracy of the mode, whereas the F1 scores give us a weighted average of the precision and recall results.

RESULTS

Nouns

MANOVA. A MANOVA was conducted using the contextual diversity, contextual distinctiveness, and word frequency indices as the dependent variables and the produced and unproduced nouns as the independent variables. All variables that reported significant differences after a Bonferroni correction (in which $p < .005$) were then assessed using Pearson correlations for multicollinearity (with a threshold of $r > .70$). All the variables except the LSA index demonstrated significant difference between the produced and unproduced nouns. The word frequency index taken from SUBTLEXus demonstrated the highest effect size and demonstrated multicollinearity with the contextual diversity indices from SUBTLEXus and the BNC spoken corpus. For this reason, the SUBTLEXus word frequency index was kept in the analysis, and the other two indices were removed. The word frequency index taken from the entire BNC corpus demonstrated multicollinearity with the word frequency indices taken from the BNC spoken and written corpora and the contextual diversity index from the entire BNC corpus and the BNC written corpus. Because the word frequency for the entire BNC index demonstrated a greater effect size with the independent variable, it was retained in the analysis, and the other four indices were removed (see Table 1 for the descriptive statistics and the MANOVA results for the variables selected for the DFA analysis).

Discriminant Function Analysis. The stepwise DFA selected variables from column 1 in Table 1 on the basis of a statistical criterion that retains the variables that best classify the grouping variable (produced or unproduced nouns) and helps control for potential multicollinearity. For our analysis, the significance level for a variable to be entered or to be removed from the model was set at the norm generally adopted in applied linguistics: $p \leq .05$. The stepwise DFA retained two variables as significant predictors of whether a noun was produced or not produced by beginning-level L2 learners (word frequency SUBTLEXus and word familiarity) and removed the remaining two variables as nonsignificant predictors.

The results demonstrate that the DFA using the two indices correctly allocated 77 of the 90 nouns in the total set, χ^2 ($df = 1$, $n = 90$) = 45.534, $p < .001$, for an accuracy of 86.0% (the chance level for this analysis and all analyses is 50%). For the LOOCV, the discriminant analysis correctly allocated 75 of the 90 nouns for an accuracy of 83.3% (see the confusion matrix reported in Table 2 for results).[5] The measure of agreement between the actual text type and that assigned by the model produced a weighted Cohen's kappa of .711, which demonstrated a substantial agreement.[6]

The precision scores (i.e., the number of correct results divided by the number of all retrieved results) and recall scores (i.e., the number of correct results divided by the number of all relevant results) from the model for predicting the nouns as either produced or not produced are presented in Table 3. The model performed equally well regardless of

Table 1. Mean (standard deviations), *f* value, *p* value, and effect size for nouns

Index	Produced	Not produced	*f*	*p*	η^2
Word frequency SUBTLEXus	4.254 (0.357)	3.633 (0.336)	72.345	< .001	.451
Word familiarity	600.267 (25.511)	568.067 (26.443)	34.560	< .001	.282
Word frequency total BNC	0.485 (0.422)	0.153 (0.148)	24.639	< .001	.219
USF word association norms	89.622 (75.086)	36.733 (31.792)	18.932	< .001	.177

Note. BNC = British National Corpus; USF = University of South Florida.

Table 2. Predicted word type (produced or not produced) from total set and LOOVC set for nouns

Actual word type	Predicted word type	
	Produced	Not produced
Total set		
Produced	38	7
Not produced	6	39
LOOCV set		
Produced	38	7
Not produced	9	36

Note. LOOCV = leave-one-out cross-validation.

whether the noun was produced or unproduced. The overall accuracy of the model for the total set was .855 (the average F1 score) and .822 for the cross-validated set. The results demonstrate that the combination of two indices related to word frequency can substantially discriminate between produced and unproduced nouns.

Verbs

MANOVA. A MANOVA was conducted using the contextual diversity, contextual distinctiveness, and word frequency indices as the dependent variables and the produced versus unproduced conditions as the independent variables. All variables that reported significant differences after a Bonferroni correction were then assessed using Pearson correlations for multicollinearity (with a threshold of $r > .70$). Like the noun analysis, all the variables except the LSA index demonstrated significant difference between the produced and unproduced verbs. The contextual diversity

Table 3. Precision and recall finding (total set and LOOCV set) for nouns

Condition	Precision	Recall	F1
Total set			
Produced	.844	.864	.854
Not produced	.867	.848	.857
LOOCV set			
Produced	.844	.809	.826
Not produced	.800	.837	.818

Note. LOOCV = leave-one-out cross-validation.

Table 4. Mean (standard deviations), *f* value, *p* value, and effect size for verbs

Index	Produced	Not produced	*f*	*p*	η^2
Contextual diversity SUBTLEXus	3.725 (0.228)	3.357 (0.299)	42.808	< .001	.330
Word familiarity	586.044 (26.731)	550.977 (29.557)	34.492	< .001	.284

index from SUBTLEXus reported the highest effect size and also demonstrated multicollinearity with the contextual diversity index from the BNC spoken corpus, which was removed. The word frequency index from SUBTLEXus had a larger effect size and demonstrated multicollinearity with all three of the word frequency indices calculated from the BNC corpus, which were also removed. Lastly, the contextual diversity index calculated from the total BNC corpus had a larger effect size and demonstrated multicollinearity with the contextual diversity index from the BNC written corpus, which was removed. Descriptive statistics and MANOVA results for the final indices used in the verb DFA are presented in Table 4.

Discriminant Function Analysis. The stepwise DFA retained the contextual diversity SUBTLEXus index and the word familiarity index. The results demonstrate that the DFA using the two indices correctly allocated 70 of the 89 verbs in the total set, χ^2 ($df = 1, n = 89$) = 29.223, $p < .001$, for an accuracy of 78.6%. For the cross-validated set, the discriminant analysis correctly allocated 69 of the 89 verbs for an accuracy of 77.5% (see Table 5 for results). The measure of agreement between the verb type (produced and unproduced) and that assigned by the model produced a weighted Cohen's kappa of .573, which demonstrated a moderate agreement.

Table 5. Predicted word type (produced or not produced) from total set and LOOCV set for verbs

Actual word type	Predicted word type	
	Produced	Not produced
Total set		
Produced	36	9
Not produced	10	34
LOOCV set		
Produced	36	9
Not produced	11	33

Note. LOOCV = leave-one-out cross-validation.

Table 6 provides the index precision and recall scores for predicting whether the verbs were produced or unproduced. The model performed similarly for predicting both produced and unproduced verbs. The overall accuracy of the model for the total set was .786 (the average F1 score). The accuracy for the cross-validated set was .775. The results demonstrate that the combination of two indices related to contextual diversity and word familiarity can moderately discriminate between produced and unproduced nouns.

Post Hoc Analysis

Noting that the leading predictor of whether a noun was produced or not produced was SUBTLEXus word frequency and that the leading predictor of whether a verb was produced or not produced was SUBTLEXus contextual diversity, we ran a post hoc analysis to investigate differences in the lexical properties of nouns and verbs under the premise that lexical differences between nouns and verbs may explain why frequency is predictive of nouns and contextual diversity is predictive of verbs. We predicted that verbs would be more ambiguous and more abstract than nouns, and, thus, the learning of verbs would benefit from greater textual diversity because context would assist with developing stronger lexical associations. Because nouns already contain richer lexical properties (i.e., they are less ambiguous and more concrete), we predicted that their production would not strongly rely on contextual diversity. For this analysis we focused on lexical properties related to depth of knowledge such as word hypernymy (i.e., word specificity), word polysemy (i.e., word ambiguity), and word meaningfulness (i.e., number of word associations) and core lexical items such as word concreteness and word imageability. These indices will be discussed briefly.

Table 6. Precision and recall finding (total set and LOOCV set) for verbs

Condition	Precision	Recall	F1
Total set			
Produced	.800	.783	.791
Not produced	.773	.791	.782
LOOCV Set			
Produced	.800	.766	.783
Not produced	.750	.786	.767

Note. LOOCV = leave-one-out cross-validation.

Depth of Knowledge Indices. Depth of knowledge indices tap into lexical associations among words and are important indicators of word knowledge (Crossley et al., 2009, 2010). We calculated depth of lexical knowledge using WordNet (Fellbaum, 1998; Miller, Beckwith, Fellbaum, Gross, & Miller, 1993) and the MRC Psycholinguistic Database (Wilson, 1988). From WordNet, we calculated an approximate polysemy value (i.e., the number of senses a word has) and a word hypernymy value (i.e., the number of levels a word has in a conceptual, taxonomic hierarchy).[7] These indices relate to word ambiguity and specificity, respectively. Using the MRC Psycholinguistic Database, we calculated word meaningfulness (i.e., the number of associations a word has according to human raters; Gilhooly & Logie, 1980; Paivio, Yuille, & Madigan, 1968; Toglia & Battig, 1978) for each word.

Core Lexical Items. Core lexical items are more likely to be basic category words, which are words that are generally learned first and that are characteristic of emerging lexicons (R. Brown, 1958; Murphy, 2004). One method to measure core lexical items is through reporting word concreteness and imageability scores as found in the MRC Psycholinguistic Database. These scores measure lexical constructs such as word abstractness (i.e., concreteness) and the evocation of mental and sensory images (i.e., imageability).

MANOVA. A MANOVA was conducted using the depth of knowledge and core lexical indices as the dependent variables and the noun and verb conditions as the independent variables. The analysis demonstrated significant differences between nouns and verbs for all indices. Overall, nouns contained much richer lexical properties such that nouns were significantly more specific, concrete, imageable, meaningful, and unambiguous. Verbs, however, were more abstract and ambiguous and, thus, less concrete, imageable, and meaningful. Descriptive statistics and MANOVA results for this analysis are located in Table 7.

Table 7. Mean (standard deviations), *f* value, *p* value, and effect sizes between nouns and verbs

Index	Nouns	Verbs	*f*	*p*	η^2
Hypernymy	8.444 (2.029)	1.933 (0.986)	374.886	< .001	.810
Concreteness	515.511 (91.271)	380.867 (75.551)	58.113	< .001	.398
Imageability	545.467 (78.872)	422.956 (90.576)	46.822	< .001	.347
Meaningfulness	524.644 (51.712)	460.044 (52.617)	34.503	< .001	.282
Polysemy	6.244 (3.569)	12.311 (11.931)	10.679	< .001	.108

DISCUSSION

This study has demonstrated that indices related to both word frequency and contextual diversity are strong classifiers of whether a word is produced or not produced by beginning-level L2 learners. Indices related to contextual distinctiveness were not strong predictors of whether words were produced or not produced. However, word frequency and contextual diversity indices appear to act independently of each other. Word frequency indices are strong classifiers of whether or not a noun is produced, whereas contextual diversity indices are strong classifiers of whether or not a verb is produced. A post hoc analysis showed that nouns and verbs significantly differed in a variety of lexical properties such that nouns were more specific, concrete, imageable, meaningful, and unambiguous; this finding provides potential evidence for the differences in classification variables between nouns and verbs.

Our noun analysis demonstrated that noun production was best explained using two indices related to word frequency and word familiarity. These two indices categorized 83% of the nouns correctly as being either produced or not produced by the L2 learners in our study. The MANOVA demonstrated that both of these indices had large effect sizes (Cohen, 1992). Additionally, the standardized coefficients reported by the DFA, which indicate the strength of the contribution that each lexical index makes in predicting the dependent variable, demonstrated that word frequency was a stronger classifier of word production than familiarity was (coefficients of .803 and .431, respectively). Contextual diversity indices were also strong predictors of noun production, but they were highly collinear with word frequency indices and did not report stronger effect sizes than our word frequency indices. This finding indicates that word frequency is a stronger predictor of noun production than contextual diversity. The findings also indicate that word familiarity contributes additional classification power independent of word frequency and provides evidence that word frequency and word familiarity may measure different lexical properties.

Our verb analysis—unlike our noun analysis—did not retain a word frequency index but, instead, retained a contextual diversity index along with a word familiarity index. These two indices correctly classified 78% of the verbs as being produced by beginning-level L2 learners. The two indices demonstrated large effect sizes, which indicate a strong relationship between the lexical properties and the dependent variable. The standardized coefficients suggested that the contextual diversity index was a stronger classifier than the word familiarity index (reporting coefficients of .668 and .502, respectively). Word frequency indices, although strongly predictive of verb production, did not report stronger effect sizes than contextual diversity indices, even though they were highly collinear. These findings indicate that contextual diversity is the strongest predictor of verb production for the sampled population, with word familiarity providing additive power to the analysis.

A comparison of the two analyses demonstrates important differences in the effects of word frequency and contextual diversity on lexical production. It appears that one lexical property is not superior to the other in explaining lexical production but that both are important qualities of word production that differ depending on the syntactic category of the word. Nouns—which are more concrete, more imageable, more specific, more meaningful, and less ambiguous—are best predicted by word frequency indices. This is likely due to the fact that the lexical properties of nouns allow them to be learned on the basis of repetition alone. Thus, the distributional properties of nouns allow lexical acquisition that is based on association through exposure. This is in contrast to verbs, which are more abstract and ambiguous. The abstractness and ambiguity found in verbs likely make them more difficult to acquire through repetition alone, and, thus, acquisition is likely aided by exposure to the words in a variety of contexts, following a principle of likely need. It is these contexts that may provide the lexical properties that afford acquisition. Therefore, context is an important element of abstract and ambiguous word learning, but words that have richer innate properties (i.e., concreteness, imageability, specificity) can be acquired through frequency alone.

In addition to word frequency and contextual diversity, word familiarity is also an important indicator of lexical acquisition and provides additive classification accuracy to both noun and verb production. Unlike word frequency and contextual diversity, the exact lexical properties that underlie human judgments of word familiarity are difficult to pinpoint. Word familiarity has often been interpreted as a measure of word exposure because words that are rated as more familiar are recognized more quickly (Connine, Mullennix, Shernoff, & Yelen, 1990; Gernsbacher, 1984; Stadthagen-Gonzalez & Davis, 2006). A variety of studies have demonstrated that word familiarity ratings—unlike written, corpus-based frequency indices—more accurately reflect spoken word frequency and therefore demonstrate a bias toward natural exposure (Stadthagen-Gonzalez & Davis, 2006). Thus, word familiarity ratings may correlate more strongly to the age of acquisition of lexical items (G. Brown & Watson, 1987). The familiarity indices used in this study were not strongly correlated with word frequency nor with contextual diversity and were the second-strongest predictors in both our verb and noun analysis. Thus, beyond word frequency and contextual diversity, human judgments of familiarity are important indicators of whether a word will be produced or not produced.

We were surprised that our indices of contextual distinctiveness were not selected as classifiers in our DFAs. Our LSA indices yielded no significant differences between produced and unproduced nouns and verbs, whereas our USF indices demonstrated significant differences between produced and unproduced nouns but not verbs. For the noun analysis, the USF indices showed medium effect sizes with the dependent variables, but they were not selected as independent variables in the DFA, which indicates that they provided no facilitative effects independent of word frequency. It is possible, then, that contextual distinctiveness, although important for explaining L1 lexical processing, is not an important property of L2 lexical production. One potential explanation for this is that beginning-level L2 learners may not attend to local contexts as much as global contexts when acquiring lexical items.

From a pedagogical perspective, this study has implications for L2 instruction and materials design. The findings indicate that, when learning nouns, L2 instructors can likely rely on the natural distributions of words as found in representative corpora. That is to say, nouns may not need to be taught in widely dispersed contexts, although such dispersion would likely not prove detrimental. However, verb learning is likely facilitated by introducing L2 learners to specific verbs in a variety of contexts to provide the learners with the extralinguistic information necessary for acquisition of these more abstract and less specific words. Thus, L2 practitioners would do well to complement verb learning by introducing students to verbs in a variety of different contexts. Similar suggestions hold true for material designers. Because L2 learners appear to benefit from a wide range of contexts when learning verbs, textbooks should also present new verbs in a variety of contexts to introduce learners to the range of possible uses of verbs.

CONCLUSION

Our primary research goals in this study were to investigate if word frequency was the strongest predictor of L2 word production and if lexical differences existed between nouns and verbs that would explain word production. We have demonstrated that, although word frequency is the strongest predictor of noun production, it is not the strongest predictor of verb production. We thus call into question the primacy of frequency effects in lexical production and acquisition. We have also demonstrated that differences in the lexical properties (i.e., concreteness, imageability, meaningfulness, specificity, and ambiguousness) can help explain why nouns are best classified using a word frequency index and why verbs are best explained using a contextual diversity index.

Even though the statistical results from this study are quite robust—reporting classification accuracies that average about 80%—we see this study as an initial investigation into comparing the power of word frequency, contextual diversity, and contextual distinctiveness effects in explaining L2 lexical acquisition. Future studies would benefit from examining lexical acquisition as a receptive skill by analyzing word-naming or lexical decision tasks. The indices used in this study could also be operationalized differently. Specifically, as larger corpora become available, additional word frequency and contextual diversity indices could be calculated. This is especially true for large corpora that contain a greater variety of American English words, which may differ slightly from the English found in the BNC. Future studies should also consider examining contextual diversity not only as calculated in the current study but also as a result of unique semantic contexts (as compared to context defined as an individual text; Johns & Jones, 2008) or as a measure of collocational productivity (i.e., the number of collocates a word demonstrates in a selected corpus that is over a specific threshold; Wolter, 2009). Such additions may combine to provide us with a better understanding of L2 word learning in relation to the effects of word frequency, contextual diversity, and contextual distinctiveness.

Received 21 March 2012
Accepted 26 July 2013
Final Version Received 26 September 2012

NOTES

1. Contextual diversity is sometimes referred to as dispersion (Gries, 2008).
2. The TASA corpus is the largest semantic space available for LSA. The corpus comprises educational texts used at the 3rd, 6th, 9th, and 12th grade levels along with college-level texts.
3. There is no real agreement among researchers about the meaningful interpretation of cosine values. We selected .26 on the basis of research in text comprehension and word recall that indicated that this cosine value denoted strong semantic associations such as those found in frequent words (Hulme, Stuart, Brown, & Morin, 2003; Lemaire, Denhiere, Bellissens, & Jhean-Larose, 2006).
4. Multicollinearity between indices indicates that the indices are effectively measuring the same patterns in the data.
5. A confusion matrix displays the number of correct and incorrect predictions made by a model.
6. Cohen's kappa coefficient is a statistical measure between 0 and 1 of interrater agreement for categorical items that incorporates an estimate of chance agreement.
7. A lower hypernymy score relates to less specific words, whereas a higher score relates to more specific words.

REFERENCES

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*, 703–719.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, *2*, 396–408.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283–316.

Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *The Quarterly Journal of Experimental Psychology: Section A*, *40*, 5–40.
Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *Modern Language Journal*, *93*, 418–429.
The British National Corpus (Version 2). (2001). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Retrieved from http://www.natcorp.ox.ac.uk/
Brown, G., & Watson, F. (1987). First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory & Cognition*, *15*, 208–216.
Brown, R. (1958). How shall a thing be called? *Psychological Review*, *65*, 14–21.
Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990.
Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 1084–1096.
Crossley, S. A., & Salsbury, T. (2010). Using lexical indices to predict produced and not produced words in second language learners. *The Mental Lexicon*, *5*, 115–147.
Crossley, S. A., Salsbury, T., & McNamara, D. (2009). Measuring L2 lexical growth using hypernymic relationships. *Language Learning*, *59*, 307–334.
Crossley, S. A., Salsbury, T., & McNamara, D. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning*, *60*, 573–605.
Crossley, S. A., Salsbury, T., & McNamara, D. S. (2012). Predicting the proficiency level of language learners using lexical indices. *Language Testing*, *29*, 240–260.
Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011a). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing*, *28*, 561–580.
Crossley, S. A., Salsbury, T., McNamara, D. S., & Jarvis, S. (2011b). What is lexical proficiency? Some answers from computational models of speech data. *TESOL Quarterly*, *45*, 182–193.
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, *24*, 143–188.
Ellis, N. C. (2006). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review*, *19*, 100–121.
Ellis, N., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function—Introduction to the special issue. *Modern Language Journal*, *93*, 329–335.
Ellis, N. C., & Ferreira-Junior, F. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal*, *93*, 370–385.
Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.
Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (Eds.), *New approaches to language mechanisms* (pp. 257–287). Amsterdam: North Holland.
Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635.
Frederiksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 361–379.
Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281.
Gilhooly, K., & Logie, R. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods*, *12*, 395–427.
Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*, 1–16.
Glenberg, A. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*, 95–112.
Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics*, *13*, 403–437.

Hulme, C., Stuart, G., Brown, G., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*, 500–518.
Johns, B. T., & Jones, M. N. (2008). Predicting word-naming and lexical decision times from a semantic space model. In V. Sloutsky, B. Love, & K. McRae (Eds.), *Proceedings of the 30th Cognitive Science Society Meeting* (pp. 279–284). Washington, DC: Cognitive Science Society.
Kirsner, K. (1994). Implicit processes in second language learning. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 283–312). San Diego, CA: Academic Press.
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.
Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.
Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.
Lemaire, B., Denhiere, G., Bellissens, C., & Jhean-Larose, S. (2006). A computational model for stimulating text comprehension. *Behavior Research Methods*, *38*, 628–637.
MacWhinney, B. (1997). Second language acquisition and the completion model. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Mahwah, NJ: Erlbaum.
McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech*, *44*, 295–322.
Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. (1993). *Five papers on WordNet* (Report No. 43). Cognitive Science Laboratory, Princeton University. Retrieved from http://wordnetcode.princeton.edu/5papers.pdf
Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393–424.
Murphy, G. L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). The University of South Florida word association, rhyme, and word fragment norms. Retrieved from http://www.usf.edu/FreeAssociation/
Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, *76*, 1–25.
Salsbury, T. (2000). *The acquisitional grammaticalization of unreal conditionals and modality in L2 English: A longitudinal perspective* (Unpublished doctoral dissertation). Indiana University, Bloomington.
Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*, 343–360.
Schmitt, N., & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *Studies in Second Language Acquisition*, *19*, 17–36.
Scott, M. (2004). *WordSmith tools*. Oxford: Oxford University Press.
Stadthagen-Gonzalez, H., & Davis, C. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*, 598–605.
Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 760–766.
Toglia, M. P., & Battig, W. F. (1978). *Handbook of semantic word norms*. Mahwah, NJ: Erlbaum.
Verkoeijen, P. P. J. L., Rikers, R. M. J. P., & Schmidt, H. G. (2004). Detrimental influence of contextual change on spacing effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*, 796–800.
Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods*, *20*, 6–10.
Wolter, B. (2009). Meaning-last vocabulary acquisition and collocational productivity. In T. Fitzpatrick & A. Barfield (Eds.), *Lexical processing in second language learners: Papers and perspectives in honour of Paul Meara* (pp. 111–127). Bristol, UK: Multilingual Matters.
Wulff, S., Ellis, N. C., Römer, U., Bardovi-Harlig, K., & LeBlanc, C. J. (2009). The acquisition of tense-aspect: Converging evidence from corpora and telicity ratings. *Modern Language Journal*, *93*, 354–369.
Zipf, G. K. (1935). *The psycho-biology of language*. New York: Houghton-Mifflin.

APPENDIX A

LEXICAL VALUES OF L2-PRODUCED NOUNS

	CDV measures				CDT measures		WF measures				
Word	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Baby	3.71	0.40	0.43	0.29	89	55	4.41	0.11	0.11	0.13	597
Book	3.51	0.79	0.82	0.65	149	46	3.96	0.43	0.43	0.39	643
Boy	3.79	0.54	0.56	0.46	54	44	4.43	0.22	0.21	0.27	606
Brother	3.58	0.50	0.53	0.37	22	46	4.16	0.12	0.13	0.11	598
Business	3.71	0.75	0.79	0.56	37	48	4.21	0.37	0.38	0.23	563
Car	3.71	0.63	0.64	0.57	259	101	4.39	0.34	0.32	0.52	634
Child	3.52	0.74	0.77	0.56	41	31	3.95	0.65	0.67	0.48	608
City	3.52	0.64	0.71	0.31	41	86	3.94	0.26	0.27	0.15	616
Day	3.89	0.93	0.93	0.88	49	39	4.61	0.98	0.97	1.14	595
Father	3.74	0.54	0.58	0.36	30	38	4.45	0.23	0.24	0.15	591
Food	3.56	0.61	0.66	0.40	324	67	3.90	0.22	0.23	0.13	579
Friend	3.79	0.70	0.73	0.53	114	36	4.33	0.27	0.28	0.19	603
Girl	3.79	0.51	0.52	0.45	127	32	4.45	0.25	0.25	0.29	645
Home	3.87	0.85	0.87	0.76	97	13	4.60	0.56	0.56	0.61	626
Horse	3.20	0.41	0.43	0.26	79	149	3.68	0.14	0.14	0.12	560
House	3.78	0.85	0.87	0.74	185	52	4.42	0.70	0.70	0.74	600
Husband	3.53	0.44	0.47	0.28	14	72	4.00	0.11	0.11	0.07	557
Job	3.78	0.74	0.74	0.73	74	77	4.32	0.31	0.29	0.51	578
Life	3.88	0.87	0.90	0.69	107	26	4.61	0.71	0.75	0.32	598
Man	3.91	0.84	0.85	0.75	171	26	4.97	0.99	1.03	0.62	623
Money	3.75	0.78	0.79	0.76	302	55	4.51	0.37	0.33	0.69	631
Morning	3.78	0.63	0.62	0.71	28	71	4.35	0.22	0.19	0.52	605
Mother	3.74	0.57	0.60	0.42	48	52	4.39	0.28	0.29	0.20	632
Movie	3.32	0.15	0.16	0.07	103	102	3.80	0.03	0.03	0.01	523
Music	3.43	0.44	0.47	0.27	132	152	3.89	0.16	0.17	0.08	599
Name	3.85	0.85	0.86	0.75	37	23	4.52	0.38	0.38	0.41	573
Night	3.87	0.69	0.69	0.66	55	87	4.52	0.38	0.36	0.58	636
Office	3.60	0.73	0.76	0.53	33	86	4.02	0.27	0.27	0.20	566
People	3.90	0.92	0.92	0.92	154	344	4.75	1.15	1.06	2.00	628
Place	3.86	0.93	0.94	0.86	62	6	4.49	0.72	0.74	0.52	612
Problem	3.77	0.86	0.88	0.78	71	41	4.23	0.58	0.59	0.56	596
Question	3.66	0.84	0.86	0.71	54	46	4.01	0.45	0.45	0.42	588
Room	3.80	0.72	0.74	0.63	79	116	4.35	0.42	0.43	0.29	627
School	3.66	0.73	0.75	0.64	183	64	4.23	0.56	0.55	0.63	582
Sister	3.48	0.42	0.44	0.31	22	55	3.96	0.10	0.10	0.08	588
Story	3.63	0.64	0.69	0.42	46	110	4.05	0.19	0.20	0.14	578

Continued

Appendix A. Continued

Word	CDV measures				CDT measures		WF measures				
	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Street	3.53	0.65	0.69	0.46	40	95	3.88	0.24	0.25	0.18	602
Student	3.06	0.48	0.53	0.24	26	100	3.34	0.19	0.20	0.08	632
Teacher	3.12	0.42	0.44	0.31	64	83	3.45	0.21	0.22	0.14	599
Thing	3.91	0.87	0.85	0.96	27	23	4.74	0.79	0.64	2.22	587
Time	3.92	0.98	0.98	0.97	119	28	5.00	2.01	1.99	2.17	604
Week	3.70	0.84	0.84	0.83	7	43	4.09	0.43	0.39	0.88	577
Woman	3.77	0.68	0.71	0.53	71	46	4.35	0.57	0.60	0.31	623
Work	3.88	0.95	0.96	0.94	196	27	4.61	1.51	1.50	1.63	603
Year	3.70	0.96	0.96	0.93	11	37	4.15	1.62	1.63	1.56	601

Note. CDV = contextual diversity; CDT = contextual distinctiveness; WF = word frequency; SBT = SUBTLEXus; BNC T = British National Corpus: total; BNC W = British National Corpus: written section only; BNC S = British National Corpus: spoken section only; USF = University of South Florida word association norms; LSA = latent semantic analysis near neighbor associations; FY = familiarity.

APPENDIX B

LEXICAL VALUES OF L2-UNPRODUCED NOUNS

Word	CDV measures				CDT measures		WF measures				
	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Age	3.40	0.77	0.81	0.55	21	44	3.61	0.35	0.36	0.19	582
Art	3.20	0.51	0.58	0.19	52	147	3.56	0.23	0.25	0.06	529
Ball	3.32	0.39	0.41	0.31	84	103	3.73	0.09	0.09	0.11	575
Bar	3.35	0.51	0.55	0.30	46	54	3.64	0.11	0.11	0.08	592
Box	3.37	0.59	0.60	0.48	55	51	3.66	0.13	0.13	0.20	599
Brain	3.30	0.36	0.39	0.20	39	99	3.59	0.06	0.07	0.03	580
Bunch	3.32	0.21	0.23	0.09	7	49	3.48	0.02	0.02	0.01	503
Case	3.69	0.87	0.89	0.76	23	124	4.16	0.66	0.70	0.31	553
Chance	3.73	0.70	0.74	0.54	16	4	4.09	0.15	0.16	0.13	563
Corner	3.26	0.51	0.54	0.37	13	59	3.43	0.10	0.10	0.10	556
Death	3.63	0.62	0.69	0.29	133	72	4.04	0.22	0.24	0.06	581
Dust	2.91	0.31	0.34	0.15	21	61	3.09	0.04	0.04	0.02	558
Engine	2.91	0.31	0.33	0.16	20	159	3.21	0.07	0.07	0.06	543
Entrance	2.79	0.30	0.34	0.12	12	85	2.90	0.04	0.04	0.02	555

Continued

Appendix B. Continued

	CDV measures				CDT measures		WF measures				
Word	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Fault	3.49	0.42	0.45	0.26	9	76	3.73	0.05	0.05	0.05	541
Figure	3.57	0.75	0.81	0.46	25	30	3.82	0.30	0.31	0.22	534
Fire	3.60	0.61	0.65	0.41	91	86	4.04	0.19	0.18	0.21	580
Fool	3.37	0.26	0.29	0.10	21	23	3.66	0.03	0.03	0.02	551
Gift	3.27	0.35	0.40	0.11	33	42	3.52	0.05	0.05	0.02	566
Glass	3.27	0.48	0.52	0.29	60	78	3.49	0.14	0.14	0.10	611
Hell	3.77	0.31	0.30	0.36	19	122	4.38	0.06	0.05	0.14	564
Joy	2.97	0.29	0.33	0.10	22	114	3.16	0.03	0.03	0.02	545
Key	3.34	0.67	0.72	0.41	12	45	3.65	0.15	0.15	0.11	603
Minute	3.78	0.71	0.70	0.74	8	23	4.28	0.28	0.25	0.49	621
Mirror	2.94	0.38	0.43	0.15	19	85	3.09	0.05	0.06	0.03	593
Moment	3.64	0.70	0.71	0.66	8	211	3.98	0.26	0.25	0.31	560
Mouth	3.48	0.42	0.45	0.27	47	132	3.73	0.12	0.12	0.06	572
Pain	3.41	0.43	0.47	0.21	158	137	3.70	0.09	0.10	0.04	569
Pants	3.25	0.15	0.16	0.07	31	160	3.48	0.01	0.01	0.01	575
Part	3.75	0.93	0.95	0.84	30	2	4.13	0.71	0.73	0.43	579
Power	3.47	0.75	0.83	0.39	53	60	3.88	0.42	0.45	0.13	556
Rock	3.28	0.42	0.47	0.17	64	93	3.64	0.12	0.13	0.04	583
Seat	3.39	0.51	0.56	0.30	20	78	3.60	0.11	0.12	0.07	597
Sense	3.58	0.73	0.77	0.50	14	60	3.83	0.28	0.29	0.16	583
Shape	3.08	0.60	0.67	0.27	28	57	3.19	0.14	0.15	0.05	575
Sign	3.55	0.74	0.79	0.50	43	20	3.83	0.19	0.20	0.13	543
Size	3.23	0.67	0.71	0.48	25	47	3.37	0.17	0.18	0.13	566
Soul	3.31	0.29	0.34	0.08	8	244	3.59	0.04	0.05	0.01	544
Square	3.03	0.51	0.55	0.31	33	32	3.21	0.09	0.09	0.10	576
Star	3.26	0.45	0.50	0.22	51	79	3.62	0.12	0.13	0.05	574
Suit	3.29	0.57	0.63	0.29	23	101	3.54	0.09	0.09	0.06	543
Sun	3.29	0.46	0.51	0.20	81	72	3.55	0.11	0.11	0.05	635
Trial	2.98	0.39	0.44	0.11	14	168	3.40	0.07	0.08	0.02	509
Trouble	3.68	0.63	0.65	0.54	44	22	4.06	0.12	0.12	0.16	590
Wire	2.91	0.31	0.33	0.21	17	95	3.15	0.04	0.04	0.04	556

Note. CDV = contextual diversity; CDT = contextual distinctiveness; WF = word frequency; SUBT = SUBTLEXus; BNC T = British National Corpus: total; BNC W = British National Corpus: written section only; BNC S = British National Corpus: spoken section only; USF = University of South Florida word association norms; LSA = latent semantic analysis near neighbor associations; FY = familiarity.

APPENDIX C

LEXICAL VALUES OF L2-PRODUCED VERBS

	CDV measures				CDT measures		WF measures				
Word	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Buy	3.63	0.73	0.74	0.69	34	69	192.43	0.28	0.25	0.62	575
Call	3.88	0.93	0.94	0.90	26	36	861.39	0.65	0.64	0.78	599
Choose	3.21	0.74	0.81	0.44	13	32	48.06	0.19	0.20	0.09	534
Clean	3.54	0.56	0.59	0.41	109	101	121.24	0.12	0.12	0.15	610
Come	3.92	0.96	0.96	0.97	16	14	3,140.98	1.56	1.37	3.34	608
Cry	3.28	0.42	0.46	0.22	46	115	65.65	0.10	0.10	0.05	566
Dance	3.39	0.39	0.42	0.23	62	158	148.04	0.09	0.10	0.07	550
Drink	3.64	0.54	0.58	0.38	101	81	247.39	0.19	0.18	0.23	628
Eat	3.68	0.54	0.56	0.43	126	70	251.88	0.17	0.16	0.31	529
Feel	3.86	0.85	0.86	0.81	21	57	627.24	0.75	0.76	0.59	588
Find	3.89	0.95	0.95	0.92	26	21	830.96	1.05	1.07	0.78	580
Get	3.92	0.91	0.90	0.97	28	41	4,583.76	2.18	1.39	9.63	604
Give	3.91	0.97	0.97	0.95	67	5	1,167.82	1.33	1.32	1.41	595
Go	3.92	0.96	0.95	0.98	70	21	3,793.04	2.44	1.89	7.66	618
Help	3.89	0.89	0.90	0.80	105	8	921.12	0.51	0.52	0.42	594
Kill	3.76	0.54	0.58	0.36	73	125	452.57	0.15	0.16	0.11	549
Kiss	3.44	0.25	0.28	0.13	23	56	121.16	0.06	0.07	0.04	592
Know	3.92	0.96	0.95	0.97	38	8	5,721.18	1.97	1.48	6.52	605
Learn	3.53	0.72	0.76	0.51	39	26	118.57	0.27	0.28	0.16	552
Like	3.92	0.95	0.94	0.98	40	18	3,998.96	1.76	1.50	4.19	610
Listen	3.84	0.53	0.53	0.53	29	77	544.78	0.13	0.12	0.23	543
Live	3.78	0.83	0.85	0.70	21	28	344.59	0.44	0.44	0.40	608
Look	3.92	0.93	0.92	0.95	66	11	1,947.27	1.35	1.24	2.42	607
Love	3.87	0.59	0.62	0.46	181	106	1,114.98	0.34	0.34	0.32	619
Make	3.91	0.98	0.98	0.97	40	17	1,387.75	2.27	2.30	1.94	618
Meet	3.78	0.82	0.85	0.65	11	9	352.27	0.28	0.29	0.20	575
Need	3.91	0.94	0.95	0.87	32	8	1,294.90	0.95	0.93	1.12	589
Play	3.73	0.83	0.87	0.68	95	116	354.53	0.47	0.47	0.51	586
Read	3.68	0.78	0.80	0.68	38	32	241.22	0.30	0.30	0.36	568
Say	3.90	0.97	0.96	0.99	24	27	1,639.78	3.21	2.86	6.65	600
See	3.92	0.97	0.97	0.98	60	6	2,556.73	1.90	1.73	3.51	625
Sleep	3.67	0.45	0.47	0.33	75	86	227.94	0.14	0.13	0.14	610
Speak	3.63	0.77	0.78	0.74	31	59	187.18	0.29	0.29	0.30	600
Stay	3.83	0.70	0.71	0.64	21	15	515.65	0.21	0.21	0.27	573
Take	3.92	0.98	0.98	0.96	63	17	1,891.04	1.81	1.77	2.10	555
Teach	3.33	0.54	0.58	0.35	19	27	72.84	0.17	0.18	0.11	552

Continued

Appendix C. Continued

	CDV measures				CDT measures		WF measures				
Word	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Tell	3.91	0.87	0.85	0.93	42	27	1,724.49	0.79	0.74	1.29	596
Test	3.32	0.68	0.73	0.42	71	91	84.08	0.26	0.27	0.13	566
Think	3.92	0.94	0.93	0.97	70	11	2,691.39	1.71	1.31	5.41	612
Travel	3.07	0.63	0.68	0.35	42	51	33.37	0.13	0.14	0.08	550
Use	3.80	0.96	0.96	0.96	19	13	343.84	1.65	1.62	1.84	567
Walk	3.69	0.64	0.65	0.60	59	36	215.86	0.30	0.30	0.35	625
Want	3.92	0.88	0.86	0.93	33	14	2,759.18	0.93	0.76	2.51	606
Watch	3.77	0.64	0.66	0.52	45	18	330.02	0.24	0.24	0.32	576
Write	3.47	0.85	0.88	0.72	54	63	126.80	0.51	0.51	0.45	560

Note. CDV = contextual diversity; CDT = contextual distinctiveness; WF = word frequency; SBT = SUBTLEXus; BNC T = British National Corpus: total; BNC W = British National Corpus: written section only; BNC S = British National Corpus: spoken section only; USF = University of South Florida word association norms; LSA = latent semantic analysis near neighbor associations; FY = familiarity.

APPENDIX D

LEXICAL VALUES OF L2-UNPRODUCED VERBS

	CDV measures				CDT measures		WF measures				
Word	SBT	BNC T	BNC W	BNC S	USF	LSA	SBT	BNC T	BNC W	BNC S	FY
Act	3.49	0.80	0.86	0.46	28	76	109.12	0.40	0.44	0.12	566
Allow	3.22	0.83	0.87	0.65	19	10	44.37	0.32	0.33	0.19	504
Beat	3.50	0.48	0.52	0.29	25	90	131.69	0.10	0.10	0.06	536
Benefit	2.75	0.65	0.70	0.39	3	30	14.35	0.19	0.19	0.12	481
Blow	3.45	0.51	0.55	0.33	25	68	97.57	0.08	0.08	0.08	528
Build	3.19	0.85	0.88	0.67	35	54	48.08	0.42	0.43	0.29	554
Count	3.42	0.56	0.60	0.36	5	37	89.96	0.09	0.09	0.08	574
Cross	3.23	0.69	0.75	0.40	21	48	55.04	0.18	0.18	0.10	525
Cut	3.69	0.79	0.82	0.62	81	108	229.76	0.27	0.27	0.30	581
Drop	3.55	0.66	0.69	0.52	30	24	130.61	0.14	0.14	0.15	577
Fill	3.23	0.66	0.70	0.46	9	21	43.94	0.13	0.13	0.12	521
Form	3.19	0.85	0.90	0.61	13	26	42.75	0.58	0.62	0.21	553
Hang	3.59	0.54	0.54	0.52	12	21	147.75	0.11	0.11	0.17	514

Continued

Appendix D. Continued

Word	CDV measures SBT	BNC T	BNC W	BNC S	CDT measures USF	LSA	WF measures SBT	BNC T	BNC W	BNC S	FY
Hide	3.37	0.52	0.58	0.23	32	57	69.69	0.09	0.09	0.04	515
Hold	3.81	0.88	0.92	0.71	63	31	436.73	0.50	0.53	0.26	596
Hurt	3.71	0.34	0.35	0.25	137	42	246.35	0.06	0.06	0.07	579
Jump	3.30	0.43	0.46	0.31	34	65	69.82	0.07	0.07	0.08	551
Lead	3.40	0.85	0.91	0.52	17	17	83.25	0.45	0.48	0.14	526
Order	3.58	0.86	0.90	0.64	35	16	156.57	0.45	0.48	0.22	570
Own	3.84	0.93	0.95	0.86	17	11	459.20	0.80	0.84	0.51	598
Paint	3.01	0.47	0.51	0.25	40	100	36.80	0.13	0.14	0.08	551
Pass	3.50	0.80	0.84	0.62	19	29	108.12	0.26	0.27	0.17	535
Pick	3.67	0.66	0.66	0.67	27	21	198.39	0.16	0.15	0.32	524
Point	3.72	0.90	0.92	0.82	60	26	236.53	0.65	0.64	0.78	538
Pull	3.57	0.58	0.60	0.50	33	65	146.45	0.16	0.16	0.16	565
Raise	3.27	0.76	0.84	0.39	8	69	55.20	0.21	0.23	0.11	534
Rate	2.97	0.67	0.73	0.38	5	67	24.92	0.34	0.36	0.13	527
Reach	3.30	0.78	0.86	0.35	8	39	56.92	0.26	0.28	0.06	577
Recall	2.90	0.46	0.51	0.20	2	39	19.67	0.07	0.07	0.03	526
Rest	3.71	0.78	0.81	0.61	28	16	212.96	0.20	0.21	0.15	590
Save	3.61	0.68	0.72	0.49	29	14	162.31	0.13	0.13	0.16	559
Shoot	3.51	0.49	0.53	0.29	23	61	164.94	0.13	0.13	0.08	536
Sink	2.81	0.40	0.45	0.20	24	60	16.92	0.05	0.06	0.03	586
Spot	3.31	0.50	0.55	0.30	19	28	61.57	0.09	0.09	0.05	560
Steal	3.25	0.34	0.38	0.14	36	34	53.33	0.05	0.05	0.02	562
Sue	2.84	0.23	0.24	0.16	5	102	29.37	0.03	0.03	0.05	528
Throw	3.57	0.61	0.64	0.45	30	97	128.82	0.13	0.13	0.13	548
Tie	3.17	0.55	0.59	0.36	37	41	44.43	0.08	0.08	0.07	559
Trace	2.85	0.41	0.48	0.09	2	6	19.39	0.05	0.05	0.01	531
Trust	3.63	0.56	0.62	0.25	28	45	178.18	0.12	0.13	0.07	548
Turn	3.77	0.89	0.90	0.81	30	58	306.47	0.60	0.62	0.47	583
Wash	3.13	0.46	0.48	0.35	39	97	40.73	0.09	0.08	0.17	632
Waste	3.29	0.54	0.56	0.41	22	120	53.25	0.10	0.10	0.08	577
Wave	2.83	0.46	0.53	0.15	20	89	21.25	0.10	0.11	0.03	518

Note. CDV = contextual diversity; CDT = contextual distinctiveness; WF = word frequency; SBT = SUBTLEXus; BNC T = British National Corpus: total; BNC W = British National Corpus: written section only; BNC S = British National Corpus: spoken section only; USF = University of South Florida word association norms; LSA = latent semantic analysis near neighbor associations; FY = familiarity.