Studies in Second Language Acquisition **41** (2019), 721–744
doi:10.1017/S0272263118000268

Research Article

ABSOLUTE FREQUENCY EFFECTS IN SECOND LANGUAGE LEXICAL ACQUISITION

Scott A. Crossley*

Georgia State University

Stephen Skalicky

Georgia State University

Kristopher Kyle

University of Hawai'i Manoa

Katia Monteiro

Georgia State University

Abstract

A number of longitudinal studies of L2 production have reported frequency effects wherein learners' produce more frequent words as a function of time. The current study investigated the spoken output of English L2 learners over a four-month period of time using both native and non-native English speaker frequency norms for both word types and word tokens. The study also controlled for individual differences such as first language distance, English proficiency, gender, and age. Results demonstrated that lower level L2 learners produced more infrequent tokens at the beginning of the study and that high intermediate learners, when compared to advanced learners, produced more infrequent tokens at the beginning of the study and more frequent tokens toward the end of the study. Main effects were also reported for proficiency level, age, and language distance. These results provide further evidence that L2 production may not follow expected frequency trends (i.e., that more infrequent tokens are produced as a function of time).

We very much appreciate the feedback from the anonymous reviewers and from those that attended our presentation at the 2018 American Association for Applied Linguistics conference. Specifically, we thank Nick Ellis for his sage advice to read the transcripts.

*Correspondence concerning this article should be addressed to Scott A. Crossley, Department of Applied Linguistics/ESL, Georgia State University, 25 Park Place, Suite 1500, Atlanta, GA 30303. E-mail: sacrossley@gmail.com

Copyright © Cambridge University Press 2019

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

INTRODUCTION

Research on lexical development in second language (L2) learners can provide important information about L2 acquisition (e.g., Crossley, Salsbury, McNamara, & Jarvis, 2011a, 2011b; Zhang & Lu, 2013; Zheng, 2016). Traditionally, lexical development has been investigated by examining the proportion of relatively advanced or infrequent words in a learner's output or their knowledge of frequent and infrequent words as found in standardized tests (Read, 2000). Frequency is an important component of language acquisition because the more times learners are exposed to a linguistic item, the stronger the linguistic representation of that item becomes and the more likely the item will be incorporated into a language system allowing for quicker activation and processing of items (Diessel, 2017; Ellis, 2002a). However, while frequency is an important determinant of language learning, it is only one of many including contingency, recency, and context (Ellis, 2006b; Gries, 2018). A number of cross-sectional studies examining word production report that beginning L2 users produce more frequent words in both writing (Bell, 2003; Crossley et al., 2011a; Laufer & Nation, 1995) and speaking (Crossley et al., 2011b; Kyle & Crossley, 2015) than advanced learners supporting the assertion that more frequent words are learned earlier, likely as a results of frequency effects. However, longitudinal studies of word production for small populations of L2 learners have been mixed, with at least three studies (Crossley, Salsbury, & McNamara, 2010; Crossley, Salsbury, Titak, & McNamara, 2014; Kim, Crossley, & Kyle, 2018) demonstrating that the frequency of words produced increases as a function of time. These findings contrast with findings from cross-sectional studies as well as the expectation that more frequent words will be produced in the early stages of acquisition.

In light of these differences, the current study investigates developmental patterns in L2 lexical production in terms of word frequency in a large longitudinal corpus of 50 L2 speakers of English in an English as a second language (ESL) setting. In contrast to previous studies, the current study uses finer-grained indices of word frequency as well as controls for a number of individual differences variables. Spoken data was collected over a 4-month period along with individual difference variables including proficiency, age, gender, and first language distance from English. Word frequency was computed using native speaker and nonnative speaker frequency norms calculated at the type (unique words) and token (all words) level. Linear mixed effects (LME) models were conducted to examine developmental patterns in the L2 speakers in terms of these frequency measures while controlling for individual differences.

FACTORS INFLUENCING LEXICAL DEVELOPMENT

A number of different variables can influence lexical development during acquisition of an additional language, including influence from an L1 (Jarvis, 2000; MacWhinney, 1997; Odlin, 2003), distance between an L1 and the target language (Miller & Chiswick, 2005; Odlin, 2003; Ringbom & Jarvis, 2011), relative age and age of learning onset (Birdsong & Molis, 2001; Cenoz, 2011; DeKeyser, 2000; Dörnyei & Skehan, 2003; Hyltenstam & Abrahamsson, 2003; Jarvis, 2000; Johnson & Newport, 1989; McDonald, 2000), and gender (Baker & MacIntyre, 2000; Pavlenko & Piller, 2008; Roquet, Llopis, & Pérez-Vidal, 2016; Sunderland, 2000).

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

Specifically, a large body of research has demonstrated that language learners rely on L1 lexical items when developing knowledge of lexical items in an additional language (Ellis, 2002a; Ellis & Larsen-Freeman, 2006; Jiang, 2000), and this L1 influence is partially determined by the relative similarity or difference in lexical items between the two languages (Miller & Chiswick, 2005; Odlin, 2003; Ringbom & Jarvis, 2011). These studies indicate that participants with different L1 backgrounds may develop lexical items in the target language at different rates. Lexical development is also partially influenced by age, both in terms of the relative age of a language learner and the age of learning onset (i.e., when the learner began learning the target language). Studies investigating whether critical periods of age exist for attaining nativelike proficiency generally report that for language features such as pronunciation, initiating language learning at a younger age results in a higher chance of attaining nativelike competence (e.g., Dekeyser, 2000). However, in terms of lexical development, age of onset appears to exert less of an influence in attaining nativelike proficiency, but instead is predictive of performance during semantic priming and idiom recognition tasks (see Dekeyser, 2012, for a review). These performances represent receptive vocabulary knowledge using automaticity of lexical integration as compared to nativelike production. Additionally, the relative age of a language learner may determine how strongly that learner is influenced by L1 lexical influence during lexical acquisition, with older learners experiencing more L1 influence when compared to younger learners (Cenoz, 2001). Thus, age is an important feature of language learning that should be controlled for during investigations of language acquisition (Jarvis, 2000).

Gender is another variable that may exert a potential influence on lexical development. In general, studies investigating the role of gender on language acquisition suggest that females tend to outperform males on a variety of language proficiency tests (Sunderland, 2000). However, whether this effect reflects an inherent, biological difference in language learning or whether it is a function of cultural norms and stereotypes that serve to gender the types of skills men and women are interested in pursuing is still in question. For instance, Baker and MacIntyre (2000) compared the motivations and attitudes for language learning among woman and men enrolled in either an immersion or nonimmersion language learning program and found that gender only played a role among students enrolled in the nonimmersion program, with females, in this case, demonstrating significantly higher levels of positive attitude and motivation when compared to the males. The findings suggest that gender may interact with other social and contextual features during language acquisition.

MEASURING LEXICAL PROFICIENCY

Lexical proficiency is generally considered to be a composite of a learners' metalinguistic knowledge of lexical items, metacognitive strategies related to using those items, and use of lexical items in context (Chapelle, 1994; Kim et al., 2018; Read, 2000). One effective method for examining lexical proficiency is through the analysis of lexical output in L2 learners' speech and writing (e.g., Crossley et al., 2011a, 2011b; Daller, Milton, & Treffers-Daller, 2007; Meara, 1996; Nation, 1990; Read, 2000; Salsbury, Crossley, & McNamara, 2011). Research in this area operates under the assumption that linguistic features found in a learner's production are reflective of a learner's lexical

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

proficiency, and therefore different linguistic features within learner's production may reflect different stages of proficiency and vocabulary acquisition (Crossley et al., 2011a, 2011b; Daller & Xue, 2007; Jarvis, 2013; Lu, 2012; Saito, Webb, Trofimovich, & Isaacs, 2016).

In general, if a learner's output contains more sophisticated lexical items, it follows that their lexical proficiency is higher.[1] While a large number of different lexical features represent the construct of lexical sophistication (Kyle & Crossley, 2015), one of the most common measures used is word frequency. Word frequency represents the relative frequency of use for any particular word in a language and is captured through a representative, large corpus for the language in question. In general, sophisticated words are those that are less frequent (Laufer & Nation, 1995), and a wealth of L2 studies have demonstrated that the use of less frequent words is positively correlated with higher L2 language proficiency (Crossley et al., 2011a, 2011b; Daller et al., 2007; Kyle & Crossley, 2015, 2016; Meara, 1996; Milton, 2009).

FREQUENCY EFFECTS

Within a usage-based framework of language learning, frequency effects are regarded as a key influence on development along with other factors such as contingency, context, and recency. Frequency effects refer to learning of language knowledge based on cumulative experience with different linguistic items (Ellis, 2002a, 2002b; Ellis & Collins, 2009; Ellis, O'Donnell, & Römer, 2013, 2015; Ellis, Römer, & O'Donnell, 2016; Gries, 2018; Tomasello, 2000). From this perspective, cumulative experiences with language form memories that interact with cognitive learning mechanisms and ultimately drive language acquisition in both first and second languages. In terms of usage-based learning in a second language, repeated language input and interactions drive associative learning of constructions, which are conventionalized form-meaning mappings that range from single words to longer constructions with many possible variations, such as verb argument constructions (Ellis et al., 2016; Ellis & Wulff, 2015). Frequency effects are thus related to the relative frequency of a construction, in that more frequent constructions at all levels (e.g., lexical items, phrases, and verb argument constructions) will be more likely to occur as input and allow for interaction, which in turn drives the overall cumulative frequency of exposure for a particular learner (Ellis et al., 2016; Ellis & Wulff, 2015).

Frequency effects can therefore explain prototypical learning patterns of constructions, with more frequent constructions being learned before less frequent constructions (Ellis & Ferreira-Junior, 2009a, 2009b; Ellis, O'Donnell, & Römer, 2014; Wulff, Ellis, Römer, Bardovi-Harlig, & Leblanc, 2009). At the same time, other characteristics of constructions can influence typical developmental patterns, specifically related to L1 transfer. For example, corpus and psycholinguistic investigations of verb argument construction production among German and Spanish learners of English demonstrate some consistency with native English speaker production, but also distinct construction patterns associated with features of a learners' L1 (Römer, Roberson, O'Donnell, & Ellis, 2014; Römer, Skalicky, & Ellis, in press). Thus, while a powerful influence in driving overall acquisition of L2 constructions, frequency effects are still modulated by speakers' individual experiences and L1s.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

In terms of lexical acquisition, empirical studies generally tend to support a strong role for frequency effects. For instance, frequency in L2 output has also been used successfully in a number of studies to discriminate among different L2 proficiency levels (Bell, 2003; Crossley, Salsbury, & McNamara, 2012; Laufer & Nation, 1995; Meara & Bell, 2001; Palfreyman & Karaki, in press). Studies investigating both learners' lexical proficiency (i.e., the breadth and depth of lexical knowledge) as well as learners' speaking proficiency (i.e., fluency and accuracy of L2 speech production) all demonstrate that more advanced language samples contain more infrequent words (Crossley et al., 2011a; Crossley & McNamara, 2013; Kyle & Crossley, 2015). Lastly, studies examining human ratings of L2 proficiency consistently demonstrate that texts with lower frequency words are rated as higher in L2 proficiency (Crossley & McNamara, 2012; Crossley et al., 2011a; Dabbagh & Enayat, in press; Kim et al., 2018; Kyle & Crossley, 2015).

In addition, studies measuring the relative frequency of nouns and verbs in target language input demonstrated that frequency of input is a strong predictor of whether the L2 learners' production will contain specific words from the target language input. Specifically, nouns and verbs with higher frequency in the target language input were more likely to be found in subsequent L2 learner output (Crossley & Salsbury, 2010; Crossley, Subtirelu, & Salsbury, 2013). More recent investigations have provided evidence that frequency measures based on formulaic sequences are more predictive of proficiency when compared to single-word measures. However, single-word measures were still reported to be important considerations (Bestgen, 2017).

Frequency effects are also seen during the acquisition of more difficult constructions, such as phrasal verbs in English. For instance, Garnier and Schmitt (2016) studied L2 learner knowledge of phrasal verbs in English and found that among several different features of phrasal verbs that are thought to aid in their acquisition (e.g., semantic opacity), only corpus frequency of the phrasal verbs and time spent reading (i.e., additional input in the target language) were significant predictors of phrasal verb knowledge. These studies as a whole cohere to suggest that frequency measures are strongly associated with perceptions of proficiency as well as learner knowledge.

However, the majority of the studies reported in the preceding text were cross-sectional and not longitudinal. When L2 longitudinal studies of word production are considered, a frequency effect is often reported such that learners first produce less frequent words followed by more frequent words as a function of time. For instance, using the Salsbury longitudinal spoken corpus (Salsbury, 2000), Crossley et al. (2010) reported that word frequency increased as a function of time spent learning a second language even as scores on a standardized test of English increased (although these scores were not co-varied with the frequency measures). A follow-up study (Crossley et al., 2014) using the Salsbury longitudinal spoken corpus found that L2 learners produce words with lower frequency values, fewer words in the first frequency band, and more words beyond the most frequent 2,000 words at the beginning of a 1-year longitudinal study as compared to the end of the study. Over time, the data showed significant linear trends toward the production of words with higher frequency values, more words in the first frequency band, and fewer words beyond the most frequent 2,000 words by the end of the observation period. However, when only word types were examined (i.e., the frequency of unique words as compared to all words), trends indicated that L2 learners produce more infrequent word types as a function of time. Other research

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

has reported differences in terms of content and function word frequency. For instance, using the Salsbury longitudinal spoken corpus, Kim et al. (2018) found that L2 learners produced more frequent content words as a function of time, but less frequent function words.

These frequency effects, however, seem specific to studies that focus on L2 production. Other longitudinal studies that examine lexical development using standardized vocabulary tests find that L2 learners follow expected patterns (i.e., develop knowledge of infrequent words over time). For instance, Zhang and Lu (2013) examined lexical development in Chinese university students using the Vocabulary Size Test (VST; Nation, 1983, 1990). The study was longitudinal and assessed lexical knowledge at three different points in time over 22 months. Zhang and Lu reported that learners in their study demonstrated increased knowledge of infrequent words (described as words beyond the first 3,000 most common words in English). Similar findings were reported by Dóczi and Kormos (2016), who used the VST to examine changes in lexical knowledge as a function of time for Hungarian L1 EFL students and university-level ESL students in Britain from a variety of L1 backgrounds. Over the course of eight months, learners in both groups demonstrated significant improvements in vocabulary test score that were most strongly explained by word frequency features.

Thus, while production data shows one frequency effect explaining lexical development, receptive vocabulary studies show a different frequency effect. A number of potential reasons may explain why word frequency in L2 output can increase as a function of time spent studying English. One explanation may be due to the relative polysemy of frequent words. For instance, Crossley et al. (2010) examined the polysemy values in L2 production over time and found an initial acquisition of words followed by subsequent acquisition of different senses for those same words. Therefore, production of more frequent words may be a result of learners acquiring less frequent senses for those words, which may not be assessed in receptive tests of vocabulary knowledge. In addition, some highly frequent words, especially function words, can be difficult to acquire depending upon cross-linguistic influences and the salience of the form and function of the word, delaying acquisition until a learner gains a threshold in target language proficiency. Function words are generally not assessed in receptive vocabulary tests. Other contextual variables such as topic and prompt affects may also influence the relative frequency of words in L2 output. For instance, Yoon (2017) compared a large corpus of argumentative essays written by Chinese EFL learners. The essays were based on two separate prompts, and Yoon's analysis demonstrated that prompt topics with more complex language resulted in more complex lexical production, including word frequency.

Another potential explanation may come from complex adaptive system approaches, which fall under a usage-based framework. Complex adaptive system approaches describe learning as an emergent process wherein simple language input and social interactions lead to the development of the dynamic networks that underlie language acquisition. The language patterns that emerge during acquisition may not be linear, but rather polynomial with peaks and valleys that shift as learners encounters more input (Ellis & Wulff, 2015). Often, looking at individual parts of the system (i.e., only looking at lexical frequency effects or even the lexicon) will not afford an understanding of the entire system. Zheng (2016) provides a good example of a complex adaptive system

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

explanation of L2 lexical acquisition. In this study, Zheng analyzed the writing of 15 Chinese L1 English L2 university students over the period of 1 year. The writing was analyzed using measures of lexical diversity, frequency, and formulaic sequences, all of which comprise aspects of lexical sophistication. Their results revealed different developmental trajectories for these three features, with polynomial learning present in frequency development. Moreover, their results suggested that frequency served as a stabilizing mechanism, in that once learners' frequency stabilized, so did the other measures of lexical sophistication.

Thus, there may be limitations to a frequency account of language learning that posits that lower-level learners would produce more frequent words initially and move toward the production of more infrequent words over time. While these trends are generally supported in cross-sectional studies, they are not strongly supported in longitudinal studies that examine learner production. These studies suggest a more nonlinear and dynamic pattern of lexical proficiency development.

CURRENT STUDY

In the current study, we are interested in examining L2 lexical development in terms of word frequency. We use a longitudinal corpus of 50 L2 speakers' data between L2-L2 dyads and L1-L2 dyads. The data were collected over a 4-month period. In addition to the spoken data, we collected a number of individual difference variables that may influence lexical production including age, gender, proficiency level, and first language. To assess word frequency production, we used native speaker and nonnative speaker frequency norms. These norms were computed both for token and type frequency to examine the potential for all words (token) and unique words (types) to explain growth patterns. The following research questions guide this study:

1. Are there growth patterns in word frequency production for L2 learners?
2. Are there relationships between individual difference factors and word frequency production in L2 learners?
3. Are there interactions between word frequency production and proficiency level?

METHODS

DATA

In the current study, we examine the naturalistic oral output produced during conversations among English L1 and L2 speakers at a large university located in the southeastern United States. The speakers included 50 nonmatriculated L2 English learners enrolled in an intensive English program (IEP). These 50 IEP students conversed with 50 matriculated L1 and L2 speaking students (25 L1 and 25 L2) enrolled in undergraduate and graduate-level TESOL courses. Conversation sessions occurred once every month over a period of 4 months, lasting for an average of 30 minutes each session, with a maximum of four sessions per learner. Speakers interacted with the same partner each session. Speakers were given no specific topics to discuss. Due to time and cost, only a portion of each conversation session was transcribed. The sections of the

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

conversations that were transcribed were random so that the focus was not on the beginning or the end of conversations. These transcripts were cleaned to remove all filler words and proper names. Proper names with associated social concepts (i.e., Korea, Thanksgiving, English) were retained. Transcriptions for each conversation were split into two separate transcripts: one for the IEP student and one for the matriculated student. Not every student was able to participate in all conversation sessions.

Our analysis focuses only on the transcripts for the 50 IEP students. Thirty-five of these participants were male, and ages ranged between 18 and 55 years old ($M = 26.1$, $SD = 8.26$). L1 backgrounds included Korean, Mandarin Chinese, Japanese, French, Mexican Spanish, Malay, Portuguese, Sinhala, Thai, Arabic, and Uzbek (see Appendix A for a detailed list of participants' L1s and corresponding language distance measures). Table 1 displays descriptive information about the total corpus for the IEP students, including number of transcripts collected, information on missing transcripts, and the average transcribed word counts per month. Table 2 presents the 10 most frequent topics for the entire corpus and by meeting.

INDIVIDUAL DIFFERENCE MEASURES

In addition to transcripts of their oral production, we also collected the following individual difference measures from each of the IEP participants: English proficiency level, age, sex, and language distance (i.e., how different a participant's L1's linguistic features are from English). We included age, gender, and first language as additional variables because of multiple studies that have demonstrated the influences these variables can have on language learning and use (see preceding text). We computed language distance by using measures reported in Miller and Chiswick (2005). Each L1 is assigned a score from 1 to 3, with a score of 1 (e.g., Korean) being further from English than a score of 3 (e.g., Afrikaans).

English proficiency was measured using an ACT Compass ESL test administered at the beginning of the semester in which the conversations took place. The test included sections related to listening, reading, and grammar, and students were placed into proficiency levels based on recommendations from the test makers. Thus, students who scored 126 or below on the ACT Compass ESL reading/grammar tests were classified as high beginning level. Students that scored 127 to 142 were classified as low intermediate level and those that scored 143 to 162 were classified as high intermediate. Students who scored 163 or higher were classified as advanced. In total, the corpus included 8 high

TABLE 1. Descriptive statistics for corpus

Month	Number of transcripts	Participants without transcripts	Number of words (SD)
One	43	7	480.837 (254.299)
Two	47	3	430.192 (205.821)
Three	42	8	572.595 (336.760)
Four	35	15	387.657 (268.101)

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

TABLE 2. Incidence of top ten topics in transcripts

All transcripts		First meeting		Second meeting		Third meeting		Fourth meeting	
Topic	Incidence (normed)	Topic	Incidence (normed)	Topic	Incidence (normed)	Topic	Incidence (normed)	Topic	Incidence (normed)
Food	39 (6.588)	Food	10 (6.494)	Food	13 (7.692)	Food	11 (6.962)	English	7 (6.306)
Family	29 (4.899)	Sports	10 (6.494)	Traveling	11 (6.509)	Family	7 (4.43)	Friends	7 (6.306)
English	23 (3.885)	Family	8 (5.195)	Languages	10 (5.917)	English	5 (3.165)	Family	5 (4.505)
Friends	21 (3.547)	Friends	6 (3.896)	Family	9 (5.325)	Shopping	5 (3.165)	Food	5 (4.505)
Traveling	20 (3.378)	English	5 (3.247)	Cities	6 (3.55)	Sports	5 (3.165)	Grammar	4 (3.604)
Sports	19 (3.209)	School	5 (3.247)	English	6 (3.55)	ESL	4 (2.532)	Animal	3 (2.703)
Languages	14 (2.365)	Study	5 (3.247)	Class	5 (2.959)	Major	4 (2.532)	ESL	3 (2.703)
Study	14 (2.365)	Movie	4 (2.597)	Friends	5 (2.959)	Religion	4 (2.532)	Exam	3 (2.703)
Cities	13 (2.196)	Traveling	4 (2.597)	Cultures	4 (2.367)	Study	4 (2.532)	School	3 (2.703)
Class	12 (2.027)	Animals	3 (1.948)	Grammar	4 (2.367)	Cities	3 (1.899)	Shopping	3 (2.703)

beginner learners, 14 low intermediate learners, 9 high intermediate learners, and 19 advanced learners.

FREQUENCY FEATURES

In this study, we focus on raw frequency features for individual words (i.e., low-complexity features), which fit the criteria for a usage-based language study (Gries, 2018). We calculated frequency features based on NS and NNS corpora. The NS frequency features were calculated from the spoken section of the Corpus of Contemporary American English (COCA; Davies, 2008). The NNS frequency features were calculated using the TOEFL11 corpus (Blanchard, Tetreault, Higgins, Cahill, & Chodorow, 2013). For both corpora, we calculated frequency features related to word tokens (i.e., each individual lemmas) and word types (i.e., each unique lemmas). The features were automatically calculated using the Tool for the Automatic Analysis of Lexical Sophistication 2.2 (TAALES; Kyle & Crossley, 2015; Kyle, Crossley, & Berger, 2018). The features are discussed in more depth in the following text.

NS FREQUENCY NORMS

We selected COCA spoken norms to examine changes in word frequency over time. COCA comprises texts collected between 1990 and 2015 that represent five registers (academic, fiction, magazine, news, and spoken). The total size of the corpus is 520 million words equally divided among the registers. The spoken corpus thus comprises 104 million words. The spoken subcorpus was derived from transcripts of unscripted conversation from around 150 different radio and television shows. We selected both type and token indices for all words. These indices were logarithmically transformed to control for Zipfian distributions. Frequency scores are calculated in TAALES by dividing the sum of the frequency scores for the tokens in a text by the number of tokens in that text that received a frequency score (i.e., words in the text that are not in the selected frequency lists are not included in the frequency counts). Initial results indicated that 97% of the words in the L2 corpus were found in the COCA spoken section.

NNS FREQUENCY NORMS

A large spoken NNS corpus from which to derive frequency norms is not available. As a proxy for NNS spoken frequency, we used the TOEFL11, which comprises 11,000 essays taken from the TOEFL iBT test. This corpus is composed of argumentative essays and contains 3,509,001 words. Test takers, all NNS of English, were required to respond to a brief topic in the time span of 30 minutes. To guarantee a variety of language backgrounds and topics, a total of 11 different language backgrounds (i.e., Arabic, Chinese, French, German, Hindi, Italian, Japanese, Korean, Spanish, Telugu, and Turkish) and eight distinct essay prompts were used as criteria to develop the corpus (Blanchard et al., 2013). To create the NNS norms, we excluded tokens whose frequency equaled one (i.e., words that appeared only once in the corpus and its subcorpora). This measure prevented the inclusion of neologisms and rare misspellings. We developed type

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

and token frequency indices using the entire corpus. Like the NS spoken corpus, the indices were logarithmically transformed to control for Zipfian distributions. Initial results indicated that ~90% of the words in the L2 corpus were found in the TOEFL11 corpus.

STATISTICAL ANALYSIS

We used LME models in R (R Core Team, 2016) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) to examine any changes in frequency over time (i.e., as participants completed conversation sessions over the 4-month period). We opted to use LME models because they offer statistical advantages over traditional repeated measures analyses of variance (RM ANOVAs). Specifically, LMEs account for both pooled and individual variance among participants as opposed to one pooled group mean by including subjects as random effects (i.e., assigning a unique intercept for each participant), resulting in more accurate estimates based on individual participant variation. LMEs also better account for the impact of experimental variables of interest by allowing a researcher to assign random slopes per variable per participant. This process allows for variation attributed to hypothesis-related variables to be better separated from variance attributable to error based on other, unrelated features of the participants.

When building our models, we constructed a unique model for each frequency measure of interest, for a total of four models. The purpose for these four models was because each model included a unique dependent variable (i.e., the different measures of frequency explained previously). The purpose of each model was to test whether any of the independent variables (e.g., age, language distance, session order) significantly predicted changes in word frequency (i.e., the dependent variable), allowing us to determine whether any of the independent variables affected the production of the participants in the study. Accordingly, in each model, we entered the frequency measure as the dependent variable, with age, gender (female or male), language distance (1–3), English proficiency (advanced, high intermediate, low intermediate, or high beginning), session order (1–4), and pair language match (i.e., L1/L2 or L2/L2) as fixed effects (i.e., predictor variables). The baselines for our categorical variables were as follows: proficiency: advanced, gender: female, pair language match: L2/L2. We also included subjects as random effects, with a random slope of session order fit to participants. In each model, we also fit an interaction between session order and English proficiency to measure whether English proficiency influenced word frequency production over time. To properly interpret potential main effects for session order, we first tested models without the interaction between session order and English proficiency and only included significant interactions in our final models.

We also used several other packages to aid in our construction and interpretation of our models. We used *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2016) to derive *p* values from the models, *multcomp* (Hothorn, Bretz, & Westfall, 2008) to obtain full pairwise comparisons between the four proficiency levels (advanced, high intermediate, low intermediate, and high beginner), the effects (Fox, 2003) package to retrieve model information for our figures, and the *MuMIn* package (Nakagawa & Schielzeth, 2013) to obtain two measures of variance explained: a marginal R^2 measuring the variance explained by the fixed effects only, and a conditional R^2 measuring the variance explained by the fixed and random effects combined.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

RESULTS

COCA SPOKEN TOKEN INDEX

An LME model using COCA spoken token frequency (all words, logarithmic correction) as the dependent variable without interactions (intercept = 3.316) reported a significant, negative main effect of age (coefficient = -.001, t = -2.599, p = .013), proficiency level (high beginner compared to advanced, coefficient = -.113, t = -2.027, p = .049), and a significant, negative effect for language distance (coefficient = .087, t = 2.138, p = .038) such that older learners used less frequent words, high beginner learners produced more infrequent words when compared to advanced learners, and decreasing levels of language distance predicted participants' spoken frequency values. No other main effects were reported. An LME model with an interaction fit between proficiency and session order reported a significant main effect for age, and a significant interaction between English proficiency (high intermediate learners compared to advanced learners) and session order. In the case of the interaction, the high intermediate participants' COCA spoken token frequency was significantly lower during earlier sessions when compared to advanced participants, but then increased significantly as session order increased. This interaction is visualized in Figure 1. This model reported a marginal R^2 of .211 and a conditional R^2 of .651. Table 3 displays the estimates, standard errors, t values, and p values for the fixed effects entered into this model.

COCA SPOKEN TYPE INDEX

An LME model with COCA spoken type frequency (all words, logarithmic correction) as the dependent variable reported no significant main effects or interactions. This model reported a marginal R^2 of .062 and a conditional R^2 of .358. Table 4 displays the

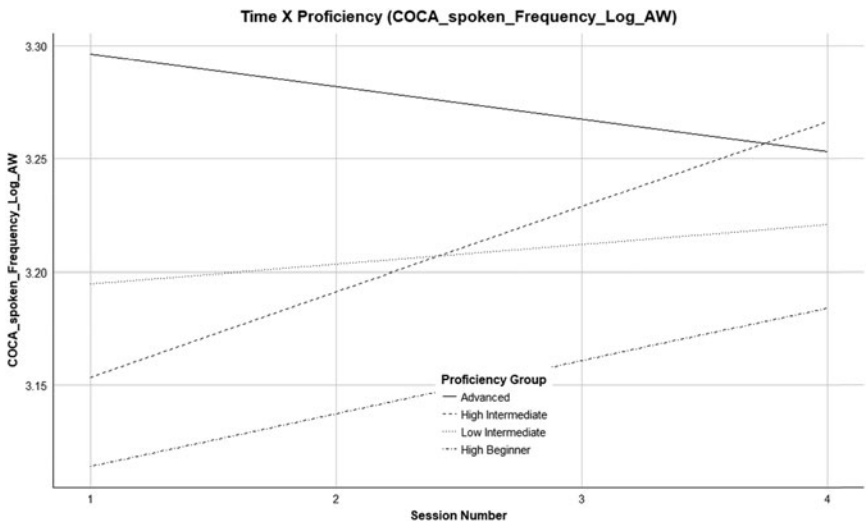


FIGURE 1. Interaction plot demonstrating the effect of proficiency and session order on COCA spoken token frequency (all words, logarithmic correction). High intermediate participants' COCA spoken token frequency (all words, logarithmic correction) was significantly lower during earlier sessions when compared to the advanced participants, and then significantly increased as session order increased.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

TABLE 3. LME model predicting COCA spoken token frequency (all words, logarithmic correction)

Fixed effects	Estimate	Std. Error	*t*	*p*
(Intercept)	3.386	0.094	36.061	<.001
Sex: Male	-0.033	0.042	-0.804	0.425
Match: L1-L2	0	0.045	0.002	0.998
Age	-0.006	0.003	-2.472	0.018
Proficiency: High Intermediate	-0.194	0.085	-2.282	0.027
Proficiency: Low Intermediate	-0.124	0.076	-1.625	0.11
Proficiency: High Beginning	-0.22	0.094	-2.343	0.023
Session Order	-0.014	0.014	-1.052	0.295
Language Distance	0.081	0.041	1.957	0.057
Proficiency X Session Order Interaction				
High Intermediate	0.052	0.022	2.399	0.019
Low Intermediate	0.023	0.019	1.197	0.234
High Beginning	0.038	0.026	1.456	0.148

Note: Baselines for categorical variables are as follows: Sex = Female, Match = L2-L2, Proficiency = Advanced

estimates, standard errors, *t* values, and *p* values for the fixed effects entered into this model.

NNS TOKEN INDICES

An LME model with TOEFL11 frequency as the dependent variable without interactions reported a significant, positive main effect for proficiency and language distance, such that high beginner learners used less frequent words than advanced learners and decreasing levels of language distance increased participants' spoken frequency values (see Table 5). No other main effects were reported and post-hoc comparison among proficiency level reported no other significant differences. The model reported a marginal R^2 of .178 and a conditional R^2 of .581. An LME model with an interaction fit between proficiency and session order reported no significant interactions Table 5 displays the estimates, standard errors, *t* values, and *p* values for the fixed effects entered into this model.

NNS TYPE INDICES

An LME model with TOEFL11 type frequency (all words, logarithmic correction) as the dependent variable reported no significant main effects or interactions. This model reported a marginal R^2 of .032 and a conditional R^2 of .414. Table 6 displays the estimates, standard errors, *t* values, and *p* values for the fixed effects entered into this model.

POST-HOC ANALYSES

We ran a series of post-hoc analyses to examine potential frequency effects for lexical acquisition resulting from differences in content and function words. Our

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

TABLE 4. LME model predicting COCA spoken type frequency (all words, logarithmic correction)

Fixed effects	Estimate	Std. Error	*t*	*p*
(Intercept)	2.790	0.081	34.518	< .001
Sex: Male	-0.015	0.038	-0.398	0.693
Match: L1-L2	0.019	0.040	0.465	0.644
Age	-0.003	0.002	-1.522	0.136
Proficiency: High Intermediate	0.047	0.050	0.941	0.352
Proficiency: Low Intermediate	0.018	0.047	0.382	0.704
Proficiency: High Beginning	-0.026	0.052	-0.508	0.614
Session Order	0.010	0.011	0.931	0.357
Language Distance	-0.021	0.037	-0.570	0.572

Note: Baselines for categorical variables are as follows: Sex = Female, Match = L2-L2, Proficiency = Advanced

previous analyses examined all words produced by L2 learners and found a frequency effect wherein lower-level L2 learners produced more infrequent words at the beginning of the study for NS frequency norms calculated at the token level. However, at least one previous study (Kim et al., 2018) has shown that L2 learners produce more frequent content words as a function of time, but less frequent function words. For the post-hoc analyses, we conducted the same LME analyses as in the first initial analysis, but we focused on spoken frequencies taken from COCA for both types and tokens.

COCA SPOKEN TOKEN INDICES (CONTENT AND FUNCTION WORDS)

For COCA content words calculated on word tokens, a significant effect was only reported for age (coefficient = -.009, t = -3.238, p = .002) indicating that older L2

TABLE 5. LME model predicting TOEFL11 frequency (all words, logarithmic correction)

Fixed effects	Estimate	Std. Error	t	p
(Intercept)	3.052	0.106	28.886	< .001
Sex: Male	-0.004	0.050	-0.078	0.938
Match: L1-L2	-0.008	0.053	-0.159	0.874
Age	-0.002	0.003	-0.632	0.531
Proficiency: High Intermediate	-0.013	0.066	-0.198	0.844
Proficiency: Low Intermediate	-0.117	0.061	-1.908	0.063
Proficiency: High Beginning	-0.150	0.068	-2.202	0.033
Session Order	0.009	0.011	0.835	0.413
Language Distance	0.129	0.049	2.657	0.011

Note: Baselines for categorical variables are as follows: Sex = Female, Match = L2-L2, Proficiency = Advanced

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

TABLE 6. LME model predicting TOEFL11 frequency (all types, logarithmic correction)

Fixed effects	Estimate	Std. Error	t	p
(Intercept)	3913.471	438.067	8.933	< .001
Sex: Male	22.187	205.262	0.108	0.914
Match: L1-L2	55.108	216.515	0.255	0.800
Age	-1.477	12.373	-0.119	0.905
Proficiency: High Intermediate	203.182	270.638	0.751	0.456
Proficiency: Low Intermediate	126.176	251.225	0.502	0.617
Proficiency: High Beginning	80.126	280.835	0.285	0.776
Session Order	39.719	61.063	0.650	0.518
Language Distance	-251.081	197.008	-1.274	0.208

Note: Baselines for categorical variables are as follows: Sex = Female, Match = L2-L2, Proficiency = Advanced

speakers produced more infrequent content words. There were no other main effects or significant interactions. For COCA function words calculated on word tokens, a significant main effect was only reported for language distance (coefficient = .075, t = 2.381, p = .021) indicating that L2 participants whose first language was closer to English produced more frequent function words. There were no other main effects or significant interactions.

COCA SPOKEN TYPE INDICES (CONTENT AND FUNCTION WORDS)

For COCA content words calculated on word types, no significant main effects were reported and there were no significant interactions. Similarly, for COCA function words calculated on word types, no main effects or significant interactions were reported.

DISCUSSION

Linguistic frequency is an important element of language acquisition with a number of L2 studies indicating that more frequent items are acquired earlier (Ellis, 2002a, 2002b; Ellis et al., 2013, 2015, 2016; Tomasello, 2000). These frequency effects are argued to be a principle component of L2 acquisition (Ellis, 2002a; Ellis & Collins, 2009), although not the only component, and are a key tenet of usage-based theories of language acquisition (Ellis, 2002a). A number of cross-sectional studies investigating L2 lexical production have supported the role of frequency in L2 word learning such that beginning learners produce more frequent words (Bell, 2003; Crossley et al., 2011a; Crossley & McNamara, 2012, 2013; Dabbagh & Enayat, in press; Kim et al., 2018; Kyle & Crossley, 2015; Laufer & Nation, 1995; Meara & Bell, 2001; Palfreyman & Karaki, in press). However, longitudinal studies of L2 lexical production have shown contrasting results with a number of studies indicating that L2 learners begin to produce more frequent words as a function of time (as compared to producing less frequent words as a function

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

of time; Crossley et al., 2010, 2014; Kim et al., 2018). These studies relied on a single longitudinal corpus (the Salsbury longitudinal spoken corpus; Salsbury, 2000) and the results may be specific to that data. Additionally, the majority of the studies focused only on frequency effects in L2 word learning and did not take into consideration individual learner factors such as L1, age, and gender, all of which may affect lexical acquisition.

Thus, the purpose of this study was to examine L2 lexical developmental patterns using frequency measures extracted from a large longitudinal corpus of 50 L2 speakers of English. In contrast to previous studies, the current study used finer-grained indices of word frequency including native speaker and nonnative speaker frequency norms calculated at the type (unique words) and token (all words) level. In addition, the study controlled for a number of individual differences variables including proficiency, age, gender, and first language distance from English.

The findings from this study support earlier longitudinal research conducted by Crossley and colleagues (Crossley et al., 2010, 2014; Kim et al., 2018) by demonstrating that lower-level L2 learners produced more infrequent words at the beginning of the study for NS frequency norms calculated at the token level. The token results indicate that high intermediate learners, when compared to advanced learners, produced more infrequent words at the beginning of the study and more frequent words toward the end of the study. Main effects were also reported for proficiency level in the NS frequency indices such that advanced learners generally produced more frequent words than high intermediate and high beginning learners. Main effects were also reported for age such that older learners produced more infrequent words, likely because they have more longer experiences with the distributional patterns of words in English. For the NNS norms, no interactions were reported, and main effects were reported for proficiency and language distance. In terms of proficiency, high beginning learners produced more infrequent words than advanced learners. For language distance, learners whose first language was closer to English produced more frequent words.

No interactions or main effects were reported for NS or NNS type frequency indices, indicating no development at the individual word level. Word types likely tap into productive knowledge (i.e., the words a learner knows) and a lack of any linear trends (either positive or negative) suggests that the L2 learners in this study likely developed words along a nonlinear frequency path.

The interaction findings for NS token measures indicate that not all L2 speakers in this analysis reported a frequency effect in which more frequent words were produced at the beginning of the study. The effect was strongest for high intermediate levels. The interaction indicates that, in general, L2 speakers at the high intermediate level produced more infrequent words at the beginning of the study and more frequent words at the end of the study when compared to advanced learners. In conjunction with the main effects reported for proficiency, which indicated that advanced L2 learners produced more frequent words than high beginning and high intermediate L2 learners, we see a pattern in which beginning and intermediate learners begin to produce more frequent words over time and start to align with the word frequency production of advanced L2 learners. While this finding fits with previous longitudinal studies of lexical production (Crossley et al., 2010, 2014; Kim et al., 2018), it runs counter to a number of cross-sectional studies of lexical acquisition (Bell, 2003; Crossley et al., 2011a, 2011b; Kyle & Crossley, 2015; Laufer & Nation, 1995). Indeed, this study, along with previous longitudinal studies,

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

provides counterevidence to the notion that frequency effects based on absolute frequency counts are strong predictors of lexical acquisition for individual words.

Instead of producing more frequent words earlier, high beginning and high intermediate learners initially produce lower frequency words and then begin to attune their production to match the frequency of more advanced learners. It is probable, then, that early learners are not fully attentive to frequency distributions in the language, likely as a result of minimal exposure, while advanced learners are more sensitive to L2 frequency distributions, as has been postulated in previous studies (Paquot, 2017). Without enough exposure to the language, L2 learners may not have enough data from which to develop accurate structural regularities. Over time, in this case 4 months, they begin to produce more frequent words that may be the result of increased language exposure that affords them the opportunity to develop statistical expectations for the lexical items in English.

To illustrate these changes, we present frequency data from a single participant (participant 63) over the course of the 4-month study. Table 7 presents the average logged frequency (COCA spoken subcorpus) for the words produced by the participant at each time point. It also includes five low-frequency words and five high-frequency words produced by the participant at each time point as well as the rank of the word within the COCA spoken subcorpus and the number of times each word was produced (a higher rank indicates a less frequent word). The transcripts from this learner along with the frequency scores for each individual word produced by this learner are available as supplemental material for this study (see supplemental material file). Table 7 indicates that the infrequent words produced by participant 63 in the first conversation were quite infrequent and many of these words were repeated (e.g., *volleyball* and *caption*). In contrast, participant 63 produced the same top five frequent words each conversation and these words were the most frequent words in the COCA spoken subcorpus. The data in Table 7 provide some support for the notion that L2 learners produce infrequent words at the beginning of the study and later begin to attune their production to match the frequency of native speakers. As mentioned earlier, this is likely the result of L2 learners becoming attuned to the word frequency distributions of language over time.

Our post-hoc analysis showed no difference in terms of development over time when we examined the frequency of content and function tokens and types, which helps to rule out the notion that frequency effects may be modulated by larger part of speech constructs. The post-hoc analysis did report that age demonstrated a significant main effect for content words such that older L2 learners produced more infrequent words than younger learners. A main effect was also reported for language distance with respect to function words such that learners with a language closer to English produced more frequent English function words.

It is also important to highlight that this study focused on single lexical items. Most studies that report frequency effects in L2 learning have generally investigated multiword characteristics and not single words. Word and phrase combinations such as construction frequency (e.g., the frequency of verb argument constructions) or type frequency (e.g., the number of lexical types that can fit into various verb argument construction) differ from single words in that they allow for larger abstractions of analogies and language patterns based on form-meaning mappings (Goldberg, 2003; Robinson & Ellis, 2008). These features examine regularities in language that go beyond the individual word, providing additional context for learning. It is possible, then, that

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

TABLE 7. Frequency data from participant 63

Time (Average corpus frequency)	Infrequent words	Occurrence	Corpus frequency (logged)	Corpus rank	Frequent words	Occurrence	Corpus frequency (logged)	Corpus rank
1 (2.703)	volleyball	5	0.306	17,941	be	16	4.727	1
	caption	3	0.451	13,087	the	6	4.666	2
	shutter	1	0.463	12,817	to	8	4.432	3
	laptop	2	0.807	7,091	and	4	4.417	4
	hockey	3	0.915	5,944	a	5	4.383	5
2 (2.768)	seventy-five	1	0.399	10,945	be	11	4.727	1
	driving	1	0.771	6,755	the	1	4.666	2
	beside	1	0.919	5,487	to	9	4.432	3
	tag	1	1.052	4,534	and	7	4.417	4
	stadium	1	1.180	3,771	a	5	4.383	5
3 (2.894)	grammar	5	0.256	13,186	be	17	4.727	1
	dormitory	2	0.353	11,627	the	1	4.666	2
	Korean	3	1.265	3,319	to	10	4.432	3
	weak	1	1.356	2,868	and	4	4.417	4
	German	2	1.485	2,339	a	2	4.383	5
4 (2.928)	roommate	1	0.787	6,592	be	9	4.727	1
	tortured	1	0.892	5,697	the	4	4.666	2
	ski	6	0.915	5,522	to	11	4.432	3
	dose	1	1.120	4,093	and	7	4.417	4
	poison	1	1.132	4,019	a	5	4.383	5

frequency effects rely on the additional context provided within frames of words and are not activated for single linguistic items. Indeed, recent studies have found that multiword combinations may be better predictors of proficiency when compared to single-word measures, though the combination of multi- and single-word frequency measures appear to produce the most accurate predictions of L2 proficiency (Bestgen, 2017; Kyle & Crossley, 2015).

In addition, it is important to note that L2 production at early stages of learning is not a random assortment of words, but rather a collection of linguistic items that are self-selected, the result of instruction, or the result of repeated, naturalistic exposure (which may not be perfectly aligned with reference-corpus frequencies). As such, language production is still likely driven by linguistic experience and, in many cases, even for single words, this experience will follow the tenets of usage-based learning theories that go beyond frequency. Thus, while the main findings of this study may provide counterevidence for frequency effects in lexical learning, as described in a usage-based framework, they may not reflect the actual language exposure for each participant, and they do not run counter to other tenets of usage-based learning. Within a complex adaptive systems framework, for example, one would not expect to find the linear trends sought in this analysis. Instead, language learning is seen as a nonlinear process that is

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

affected by a number of intervening factors such as language levels, timescales, and social interactions (Beckner et al., 2009; Ellis & Larsen-Freeman, 2009; Ellis & Wulff, 2015). In addition, a usage-based learning framework posits that learners notice and acquire linguistic items based on the salience and perception of those items (Boyd & Goldberg, 2009; Ellis, 2006a, 2012). At least one previous study (Crossley, Kyle, & Salsbury 2016) has examined the role of salience in longitudinal L2 lexical acquisition by examining word properties such as concreteness and meaningfulness and how they interact with learning. Crossley et al. found that more salient words (i.e., more concrete and more meaningful words) demonstrated development over time such that L2 learners' vocabulary became less salient (i.e., more sophisticated) as a function of time. This finding contrasts with the current frequency results and follows expected learning trends such that learners begin to produce more sophisticated language as a function of time. Indeed, recent studies conducted on larger longitudinal datasets have supported these findings (Crossley & Skalicky, in press) indicating that other elements of usage-based theories may be driving learning for single items in the absence of frequency effects.

There are other potential reasons for the results reported in this study that are supported by a usage-based framework. For instance, as mentioned by Crossley et al. (2014), many frequent words, especially function words, are associated with more complex syntactic and grammatical form-function relations and thus are more difficult to produce, especially for beginning L2 learners. Additionally, more difficult, but more frequent, form-function relations may be less phonologically salient and thus be more difficult to learn. Lastly, there may be cross-linguistic influence that interact with lexical acquisition that affect learning. As noted by Ellis (Ellis, 2006a, 2006b), for L2 learners, acquisition is not a matter of development from a blank slate (*tabula rasa*), but rather development built onto a preexisting first language and any number of additional languages (*tabula replete*). For instance, Crossley et al. (2014), argued that lexical experience in the first language likely influences production in a second language such that words that may be frequent in a first language but infrequent in an L2 may be produced before frequent words in an L2. Thus, L2 learners that have L1s that are more dissimilar to English may produce words that do not match English frequency distributions as a result of cross-linguistic influence. There is some evidence for that in the current study. For each of our token frequency indices, there was a marginal or significant main effect for language distance indicating that L2 learners whose first language was closer to English produced more frequent words in English. Thus, L2 learners of English whose first language may share similar frequency distributions with English produced more frequent words, matching the expectations of the L1, while speakers whose first language did not share similar distribution patterns produced more infrequent words, counter to the expectations of the L1. Simply put, L2 learners may rely on the distributional and statistical regularities in their L1 when first learning an L2, and this may influence the frequency of the words they produce. This finding mirrors that of a recent study by Paquot (2017) in which she reported parallel frequencies between discourse-oriented lexical bundles in L2 learners' first and second languages leading her to conclude that L1 frequency has an impact on L2 frequency.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

CONCLUSION

This study examined links between time spent learning English and a variety of individual difference factors on the word frequency output of L2 learners. The aim of the study was to investigate if word frequency decreased as a function of time learning English, as would be predicted by frequency effects or if word frequency would increase as a function of time learning English, as has been reported in previous longitudinal studies. Unlike previous studies, we used a large, longitudinal corpus of 50 L2 speakers of English and finer-grained indices of word frequency computed from both NS and NNS corpora and calculated at both the type and token level. The findings from the study support previous longitudinal studies of frequency that demonstrated that L2 learners produce more frequent words as a function of time spent learning English. This counters a central tenet of usage-based theories of language learning (frequency effects in lexical acquisition) but does not challenge other premises of usage-based approaches. While we know that frequency is an important element of a usage-based framework, it is not the only usage-based factor that impacts second language acquisition (Ellis & Wulff, 2015). Thus, the results can help refine how tenets of a usage-based framework are applied across linguistic items and potential limitations within the framework. Specifically, it appears that frequency effects are not strong predictors of acquisition for individual linguistic items like words. We are not the first to note this. For instance, Ellis and Wulff (2015, p. 84) wrote that "frequency effects come in different kinds … and they will have differently weighted impacts depending on the target structure under examination." In the case of single lexical items, other factors important to usage-based learning like salience most likely play an important role. In addition, the study provides evidence that language distance also affects lexical acquisition with L2 learners whose first language is closer to English producing more frequent words that likely meet the distribution patterns of English.

While we used a much larger longitudinal corpus of L2 learners than previously available, the distribution of learners within proficiency levels was not well balanced. While this would not affect the statistical techniques used in the study, a larger sample of high beginner and high intermediate learners is desirable. Beyond a larger sample, the amount of exposure to English participants receive should be tracked in future studies. While it is almost impossible to accurately measure exposure outside of the classroom, surveys could be used to better estimate students' interactions with L1 and L2 speakers of English along with exposure to English media (both print and visual). In addition, the manner in which proficiency was assigned was less than optimal because the ACT Compass ESL test does not measure speaking or lexical proficiency and instead focuses on listening, reading, and grammar skills. Lastly, this study focused solely on individual words and not on longer segments of linguistic features at the phrasal, utterance, or discourse level. Future studies should consider language development at phrasal, syntactic, and text cohesion levels.

NOTE

[1]Not seeing more sophisticated lexical items in production does not necessarily indicate that the learner has not acquired more complex words. Rather, it may only indicate that the learner did not produce more complex words.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

REFERENCES

Baker, S. C., & MacIntyre, P. D. (2000). The role of gender and immersion in communication and second language orientations. *Language Learning*, *50*, 311–341.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using *lme4*. *Journal of Statistical Software*, *67*, 1–48.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., & Schoenemann, T. (2009). Language is a complex adaptive system. *Language Learning*, *59* (Suppl. 1), 1–26.

Bell, H. M. (2003). *Using frequency lists to assess L2 texts* (Unpublished doctoral dissertation). University of Wales Swansea. Retrieved from EThoS (https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.636075).

Bestgen, Y. (2017). Beyond single-word measures: L2 writing assessment, lexical richness and formulaic competence. *System*, *69*, 65–78.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249.

Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., & Chodorow, M. (2013). TOEFL11: A corpus of non-native English. *ETS Research Report Series*, *2*, 1–15.

Boyd, J. K., & Goldberg, A. E. (2009). Input effects within a constructionist framework. *The Modern Language Journal*, *93*, 418–429.

Cenoz, J. (2001). The effect of linguistic distance, L2 status, and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 8–20). Buffalo, NY: Multilingual Matters.

Cenoz, J. (2011). Toward a multilingual approach in the study of multilingualism in school contexts. *Modern Language Journal*, *95*, 339–478.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research*, *10*, 157–187.

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading*, *35*, 115–135.

Crossley, S. A., & McNamara, D. S. (2013). Applications of text analysis tools for spoken response grading. *Language Learning & Technology*, *17*, 171–192.

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., Kyle, K., & Salsbury, T. (2016). A usage-based investigation of L2 lexical acquisition: The role of input and output. *The Modern Language Journal*, *100*, 702–715.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (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., & Jarvis, S. (2011a). Predicting lexical proficiency in language learners 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.

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., Titak, A., & McNamara, D. S. (2014). Frequency effects and second language lexical acquisition: Word types, word tokens, and word production. *International Journal of Corpus Linguistics*, *19*, 301–332.

Crossley, S. A., & Skalicky, S. (in press). *Examining lexical development in second language learners: An approximate replication of Salsbury, Crossley & McNamara* (2011). Language Teaching.

Crossley, S. A., Subtirelu, N., & Salsbury, T. (2013). Frequency effects or context effects in second language word learning: What predicts early lexical production? *Studies in Second Language Acquisition*, *35*, 727–755.

Dabbagh, A., & Enayat, M. (in press). The role of vocabulary breadth and depth in predicting second language descriptive writing performance. *The Language Learning Journal*.

Daller, H., Milton, J., & Treffers-Daller, J. (Eds.). (2007). *Modelling and assessing vocabulary knowledge*. Cambridge, UK: Cambridge University Press.

Daller, H., & Xue, H. (2007). Lexical richness and the oral proficiency of Chinese EFL students. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 150–164). Cambridge, UK: Cambridge University Press.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

Davies, M. (2008). *The Corpus of Contemporary American English (COCA): 560 million words, 1990-present*. Available online at https://corpus.byu.edu/coca/.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

DeKeyser, R. M. (2012). Age effects in second language learning. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 442–460). New York, NY: Routledge.

Diessel, H. (2017). Usage-based linguistics. In M. Aronoff (Ed.), *Oxford Research Encyclopedia of Linguistics*. New York, NY: Oxford University Press.

Dóczi, B., & Kormos, J. (2016). *Longitudinal developments in vocabulary knowledge and lexical organization*. New York, NY: Oxford University Press.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell.

Ellis, N. C. (2002a). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 143–188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies in Second Language Acquisition*, *24*, 297–339.

Ellis, N. C. (2006a). Cognitive perspectives on SLA: The associative-cognitive CREED. *AILA Review*, *19*, 100–121.

Ellis, N. C. (2006b). Language acquisition as rational contingency learning. *Applied Linguistics*, *27*, 1–24.

Ellis, N. C. (2012). What can we count in language, and what counts in language acquisition, cognition, and use? In S. T. Gries & D. S. Divjak (Eds.), *Frequency effects in language learning and processing* (pp. 7–34). Berlin: Mouton de Gruyter.

Ellis, N. C., & Collins, L. (2009). Input and second language acquisition: The roles of frequency, form, and function introduction to the special issue. *The Modern Language Journal*, *93*, 329–335.

Ellis, N. C., & Ferreira-Junior, F. (2009a). Construction learning as a function of frequency, frequency distribution, and function. *The Modern Language Journal*, *93*, 370–385.

Ellis, N. C., & Ferreira-Junior, F. (2009b). Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics*, *7*, 188–221.

Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics (introduction to the special issue). *Applied Linguistics*, *27*, 558–589.

Ellis, N. C., & Larsen-Freeman, D. (Eds.). (2009). *Language as a complex adaptive system*. Boston, MA: Wiley-Blackwell.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2013). Usage-based language: Investigating the latent structures that underpin acquisition. *Language Learning*, *63*, 25–51.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2014). Second language processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Linguistic Approaches to Bilingualism*, *4*, 405–431.

Ellis, N. C., O'Donnell, M. B., & Römer, U. (2015). Usage-based language learning. In B. MacWhinney & W. Grady (Eds), *The handbook of language emergence* (pp. 163–180). Malden, MA: Wiley-Blackwell.

Ellis, N. C., Römer, U., & O'Donnell, M. B. (Eds.). (2016). *Usage-based approaches to language acquisition and processing: Cognitive and corpus investigations of construction grammar*. Hoboken, NJ: John Wiley & Sons Limited.

Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 75–93). New York, NY: Routledge.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, *8*, 1–27.

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System*, *59*, 29–44.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*, 219–224.

Gries, S. T. (2018). Operationalizations of domain-general mechanisms cognitive linguists often rely on: A perspective from quantitative corpus linguistics. In S. Engelberg, H. Lobin, K. Steyer, & S. Wolfer (Eds.), *Wortschätze: Dynamik, Muster, Komplexität* (pp. 75–90). Berlin and Boston, MA: De Gruyter.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous inference in general parametric models. *Biometrical Journal*, *50*, 346–363.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 539–588). Malden, MA: Blackwell.

Jarvis, S. (2000). Methodological rigor in the study of transfer: Identifying L1 influence in the interlanguage lexicon. *Language Learning*, *50*, 245–309.

Jarvis, S. (2013). Defining and measuring lexical diversity. *Vocabulary Knowledge: Human Ratings and Automated Measures*, *47*, 13.

Jiang, N. (2000). Lexical representation and development in a second language. *Applied Linguistics*, *21*, 47–77.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Kim, M., Crossley, S. A., & Kyle, K. (2018). Lexical sophistication as a multidimensional phenomenon: Relations to second language lexical proficiency, development, and writing quality. *The Modern Language Journal*, *102*, 120–141.

Kuznetsova, A., Brockhoff, B., & Christensen, H. B. (2016). lmerTest: Tests in linear mixed effects models. R package version 2.0-32. Retrieved from https://CRAN.R-project.org/package=lmerTest.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, *49*, 757–786.

Kyle, K., & Crossley, S. A. (2016). The relationship between lexical sophistication and independent and source-based writing. *Journal of Second Language Writing*, *34*, 12–24.

Kyle, K., Crossley, S., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods*, *50*, 1030–1046.

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal*, *96*, 190–208.

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. De Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Mahwah, NJ: Lawrence Erlbaum.

McDonald, J. L. (2000). Grammaticality judgments in a second language: Influences of age of acquisition and native language. *Applied Psycholinguistics*, *21*, 395–423.

Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge, MA: Cambridge University Press.

Meara, P., & Bell, H. (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect*, *16*, 5–19.

Miller, P. W., & Chiswick, B. R. (2005). Linguistic distance: A quantitative measure of the distance between English and other languages. *Journal of Multilingual and Multicultural Development*, *26*, 1–11.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Toronto, ON: Multilingual Matters.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines*, *5*, 12–25.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle & Heinle Publishers.

Odlin, T. (2003). Cross-linguistic influence. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 436–486). Malden, MA: Blackwell.

Palfreyman, D., & Karaki, S. (in press). Lexical sophistication across languages: A preliminary study of undergraduate writing in Arabic (L1) and English (L2). *International Journal of Bilingual Education and Bilingualism*.

Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL writing. *Second Language Research*, *33*, 13–32.

Pavlenko, A., & Piller, I. (2008). Language education and gender. In N. H. Hornberger (Ed.), *Encyclopedia of language and education* (pp. 57–69). Springer.

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Read, J. (2000). *Assessing vocabulary*. Cambridge, MA: Cambridge University Press.

Ringbom, H., & Jarvis, S. (2011). The importance of cross-linguistic similarity in foreign language learning. In M. Long & C. Doughty (Eds.), *The handbook of language teaching* (pp. 106–118). Malden, MA: Wiley-Blackwell.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press

Robinson, P., & Ellis, N. C. (Eds.). (2008). *Handbook of cognitive linguistics and second language acquisition*. New York, NY: Routledge.

Römer, U., Roberson, A., O'Donnell, M. B., & Ellis, N. C. (2014). Linking learner corpus and experimental data in studying second language learners' knowledge of verb-argument constructions. *ICAME Journal*, *38*, 115–135.

Römer, U., Skalicky, S., & Ellis, N. (in press). Verb-argument constructions in L2 English learner production: Insights from corpora and psycholinguistic experiments. *Corpus Linguistics and Linguistic Theory*.

Roquet, H., Llopis, J., & Pérez-Vidal, C. (2016). Does gender have an impact on the potential benefits learners may achieve in two contexts compared: Formal instruction and formal instruction + content and language integrated learning? *International Journal of Bilingual Education and Bilingualism*, *19*, 370–386.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition*, *38*, 677–701.

Salsbury, T. (2000). *The acquisitional grammaticalization of unreal conditionals and modality in L2 English: A longitudinal perspective* (Unpublished doctoral dissertation). Bloomington, IN: Indiana University.

Salsbury, T., Crossley, S. A., & McNamara, D. S. (2011). Psycholinguistic word information in second language oral discourse. *Second Language Research*, *27*, 343–360.

Sunderland, J. (2000). Issues of language and gender in second and foreign language education. *Language Teaching*, *33*, 203–223.

Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*, 61–82.

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. *The Modern Language Journal*, *93*, 354–369.

Yoon, H. J. (2017). Linguistic complexity in L2 writing revisited: Issues of topic, proficiency, and construct multidimensionality. *System*, *66*, 130–141.

Zhang, X., & Lu, X. (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, *35*, 283–304.

Zheng, Y. (2016). The complex, dynamic development of L2 lexical use: A longitudinal study on Chinese learners of English. *System*, *56*, 40–53.

https://doi.org/10.1017/S0272263118000268 Published online by Cambridge University Press