Wordlikeness and Novel Word Learning

James Bartolotti (j-bartolotti@u.northwestern.edu)
Northwestern University, Department of Communication Sciences and Disorders
2240 Campus Drive, Evanston, IL, USA

Viorica Marian (v-marian@northwestern.edu)
Northwestern University, Department of Communication Sciences and Disorders
2240 Campus Drive, Evanston, IL, USA

Abstract

Many adults struggle with second language acquisition, but learn new words in their native language relatively easily. Most second language words do not follow native language patterns, but those that do may be easier to learn because they make use of existing language knowledge. Twenty English monolinguals learned to recognize and produce 48 novel written words in five repeated testing blocks. Half of the words were wordlike (e.g., 'nish') in form (high neighborhood density, high orthotactic probability), while half were not (e.g., 'gofp'). Participants were more accurate at recognizing and producing wordlike compared to unwordlike items. In addition, participants were faster to respond correctly in wordlike trials. English vocabulary size predicted wordlike learning, while phonological memory predicted learning for both wordlike and unwordlike items. Results suggest that existing language knowledge affects acquisition of novel written vocabulary, with consequences for second language instruction.

Keywords: Second language acquisition; neighborhood density; orthotactic probability

Introduction

There is large variability in second language acquisition (SLA) success, particularly in adults (Birdsong, 2009). Even in cases of successful acquisition, proficiency rarely approaches native-like in all areas (Baker & Trofimovich, 2005; DeKeyser, 2005; MacWhinney, 2005). This difficulty inherent in SLA is in contrast to the relative ease with which people expand their native-language vocabularies. Vocabulary learning is a lifelong process: new words continually enter the language, including recent additions to the Merriam-Webster English dictionary *staycation*, and *truthiness*; and few hobbies or professional pursuits are free of jargon words to learn. While word learning clearly does not cease with adulthood, SLA difficulties suggest that not all words are created equal, and some words are easier to learn than others.

Languages are characterized by enormous potential variation, but are relatively limited in the number of acceptable combinations. As an illustrative example, consider the case of four-letter words in English. Given English's 26 letters, there are 456,976 possible four-letter words (26^4). Yet only about 2,200—less than half a percent—are estimated to exist in speakers' vocabularies (unpublished calculations based on CLEARPOND, Marian, Bartolotti, Chabal, & Shook, 2012). Importantly, those 2,200 words do not reflect a random sampling of all possible combinations, but are highly regular. Many words are 'neighbors' of each other, differing by only one letter, and certain sequences of letters appear frequently (e.g., CE, LY) while others do not appear at all (e.g., BK, FD). People are sensitive to these regularities, and rate nonsense words that follow a language's rules as more wordlike than sequences with low typicality (Bailey & Hahn, 2001).

Two metrics, neighborhood size and ortho-/phono-tactic probability, can characterize patterns within a language and determine how closely novel words adhere to those patterns. Orthotactic and phonotactic probability measure the likelihood of a given sequence of letters or sounds based on how often they appear in other words. Neighborhood size is a measure of how many other words are similar to a given sequence. Ortho/phonotactic probabilities and neighborhood size provide distinct, complementary metrics of word typicality, the former based on sublexical phonological/graphemic effects, the latter based on lexical effects (Vitevitch & Luce, 1999).

These regularities, however, are often not shared across languages. Written English words tend to have 5-7 times more English neighbors than neighbors in related languages like Dutch, French, German, or Spanish (Marian et al., 2012). These dissimilarities in structure may contribute to second language vocabulary learning difficulties. However, second language words are not equally difficult, for example, cognatesoverlap in form and meaning across languages and are easier to learn than noncognates (De Groot & Keijzer, 2000). Similarly, an English learner of German may find it easier to learn words that resemble English, like 'sind' (meaning *are*) and 'hinter' (meaning *behind*), compared to atypical words such as 'jetzt' (meaning *now*). By identifying the degree to which prior lexical knowledge influences second language vocabulary learning, we can determine how much to prioritize shared-structure words during early language instruction. In beginning second language (L2) learners, increased proficiency in the L2 has a snowball effect, where sequences that are wordlike in the L2 become easier to learn (Majerus, Poncelet, van der Linden, & Weekes, 2008; Stamer & Vitevitch, 2012), making early vocabulary acquisition an important goal.

Previous research has shown that auditory nonword learning is affected by neighborhood size and phonotactic probability. Nonwords with dense neighborhoods and high phonotactic probability are easier to repeat after holding them in working memory (Frisch, Large, & Pisoni, 2000; Luce & Large, 2001; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005). In addition, nonwords with high neighborhood density that are encountered in a sentence context are learned better than low density nonwords (Storkel, Armbrüster, & Hogan, 2006). One reason for this advantage may be that highly wordlike sequences, as they begin to degrade in memory, can be reconstructed based on similar neighbors at the lexical level and sequence typicality at the sub-lexical level. This process would allow novel words to be available in memory for a longer period of time, sufficient for transfer to long-term memory storage.

While effects on auditory word repetition and learning are well studied, the effect of wordlikeness on novel orthographic word recognition and production is comparatively understudied. Due to conservation of orthographic inventories between related languages, written language may be especially important when considering how wordlikeness affects L2 learning. Orthographic consistency does affect how well learners spell newly learned words that they hear(Burt & Blackwell, 2008), suggesting that orthographic knowledge plays an important role in vocabulary learning.

In the present study, we compared learning of novel words having either high or low English wordlikeness. Novel words were paired with familiar picture referents to simulate a second language learning context, and by using five repeated testing blocks with feedback, we were able to gain fine-grained detail on participants' learning over time. Based on previous wordlikeness effects on auditory nonword learning, (Roodenrys & Hinton, 2002; Storkel et al., 2006; Thorn & Frankish, 2005), we predicted an accuracy and response time advantage for wordlike items. The repeated nature of the tests let us also determine how wordlikeness affects learning rate, and whether there are consistent differences between sets of words over time. In addition, we evaluated how individuals' linguistic and cognitive backgrounds interacted with wordlikeness to affect learning outcomes, by correlating learning performance with measures of English vocabulary size, phonological short term memory, and nonverbal IQ. Because vocabulary learning involves acquisition and storage of novel letter/sound sequences, we expected high phonological short term memory ability to predict word learning success.(Majerus et al., 2008; Martin & Ellis, 2012). Nonverbal IQ has been shown to predict learning of linguistic patterns such as grammatical rules, but may not be as good a predictor of vocabulary acquisition (Kempe, Brooks, & Kharkhurin, 2010). Finally, we expected vocabulary size to predict learning only for wordlike items, because they can directly benefit from existing language knowledge.

Methods

Participants

Twenty English monolinguals participated for cash or course credit. Informed consent was obtained in accordance with Northwestern University's IRB. After the experiment, participants completed tests of phonological memory (*Comprehensive Test of Phonological Processing*, Phonological Memory Composite Score of digit span and nonword repetition subtests; Wagner, Torgesen, & Rashotte, 1999), English receptive vocabulary (*Peabody Picture Vocabulary Test III*; Dunn, 1997), and nonverbal IQ (*Weschler Abbreviated Scale of Intelligence*, block design and matrix reasoning subtests; PsychCorp, 1999).

Materials

Forty-eight orthographic CVCC words were created in a novel language named Colbertian. Half of the words were designed to have high English wordlikeness (e.g., 'nish' or 'baft') based on orthographic neighborhood size and orthotactic probability (sum of grams and sum of bigrams), while the other half of the words were unwordlike (e.g., 'gofp' or 'kowm') (Table 1, calculations from CLEARPOND, Marian et al., 2012). Although there was no auditory component to the learning task, all Colbertian words were assessed for phonological wordlikeness. Six English monolinguals (not participants in the current study) pronounced each Colbertian word, and their responses were phonologically transcribed. The Wordlike and Unwordlike lists also differed ps $< .05$ on English phonological neighborhood size, sum of phone probability, and sum of biphone probability (CLEARPOND, Marian et al., 2012). Each novel word was paired with a color line drawing (Rossion & Pourtois, 2004). Pictures were chosen to be highly recognizable (naming reliability: M = 99.1%, SD = 2.0%, Bates et al., 2003), and did not overlap orthographically or phonologically with the Colbertian words. Pictures for wordlike and unwordlike items did not differ on lexical frequency, orthographic or phonological neighborhood size, or gram, bigram, phoneme, or biphone probabilities (CLEARPOND, Marian et al., 2012).

Table 1: Stimuli characteristics.

Measure	Wordlike	Unwordlike	t(46)
Orthographic N	5.54 (1.72)	0 (0)	15.79***
Sum of Grams	.289 (.042)	.223 (.048)	4.96***
Sum of Bigrams	.026 (.011)	.010 (.009)	5.16***

Note: N = Neighborhood size; *** = $p < .001$

Procedure

Participants began training with a single exposure block of 48 randomized trials to familiarize them with the novel language. In each exposure trial, a picture was presented in the center of the computer screen, and the written target word in Colbertian appeared below the picture. Trials advanced automatically after two seconds. Following the exposure block, participants performed five blocks of word recognition and word production tasks with feedback. The entire learning component lasted on average 51.7 minutes (SD = 9.7 minutes).

Word learning: Recognition In 48 recognition trials, a random target picture and three randomly selected foil pictures were displayed in the four corners of the screen, and the written target word appeared in the center of the screen. The participant was instructed to click on the correct picture; accuracy and response time were recorded. After making a response, the three foils disappeared, and the target picture and written word remained onscreen for 1000 ms. This feedback period provided an additional learning opportunity. After completing all 48 trials, the participant was shown his or her accuracy for the block, then began the production task.

Word learning: Production In 48 production trials, a random target picture was presented in the center of the screen; the participant was instructed to type the name of the picture in Colbertian and the participant's response and RT were recorded. After making a response, the picture and the participant's answer remained on the screen, and the correct name of the target was printed below the participant's response for 1000 ms. This feedback prompted additional learning. After completing all 48 trials, the participant was shown his or her accuracy for the block, then a new testing block of recognition and production began. After the fifth block the experiment concluded.

Data Analysis

Recognition and production accuracy and RT were analyzed with 2 (Condition: wordlike, unwordlike) × 5 (Block) repeated measure ANOVAs. Follow up comparisons were performed using paired or two-sample t-tests. Outlier RTs within each combination of Block and Condition were identified (RT > M + 2SD) and replaced with M + 2SD (4.2% of trials). Analyses were repeated on untrimmed RT data and yielded the same pattern of results. RT analyses were performed on correct responses only in order to control for accuracy differences between blocks and conditions. In production blocks where a subject's or item's mean accuracy was zero (Block one: 8 subjects or 14 items; Block two: 1 subject or 1 item), RT was multiply imputed using the Amelia II software package in R (Honaker, King, & Blackwell, 2011).

Results

Recognition Accuracy

There was a significant main effect of Block on accuracy, $F_1(4,76) = 113.8$, $p < .001$, $F_2(4,184) = 96.60$, $p < .001$, and a significant interaction of Block and Condition by subjects $F_1(4,76) = 4.007$, $p < .01$, but not by items, $F_2(4,184) = 1.90$, ns. Additionally, there was no main effect of Condition, $F_1(1,19) = 2.52$, ns, $F_2(1,46) = 1.21$, ns.

Follow up comparisons (Figure 1) revealed that accuracy improved between each block (*ps* < .05), except blocks four and five, likely due to ceiling effects on accuracy (Block One: M = 64.4%, SD = 13.1%; Block Two: M = 80.7%, SD = 11.7%; Block Three: M = 92.6%, SD = 8.0%; Block Four: M = 96.7%, SD = 5.9%; Block Five: M = 97.7%, SD = 4.8%). The interaction revealed significant subject effects of Condition on accuracy (*ps* < .05 by subjects) in blocks two (Wordlike: M = 83.8%, SD = 11.1%; Unwordlike: M = 77.7%, SD = 14.3%) and three (Wordlike: M = 94.4%, SD = 6.2%; Unwordlike: M = 90.8%, SD = 10.7%).

Learning rate (i.e., the average difference in accuracy between subsequent blocks) was not different between wordlike (M = 9.2%, SD = 3.3%) and unwordlike (M = 7.5%, SD = 3.3%) trials, ns.

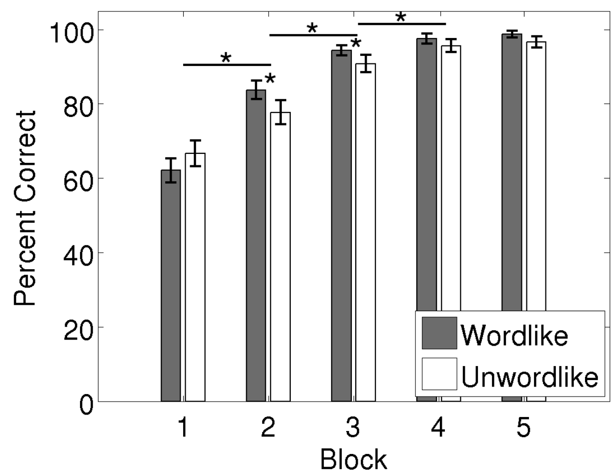

Figure 1: Novel word recognition accuracy. Asterisks indicate significance at $p < .05$, error bars indicate standard error.

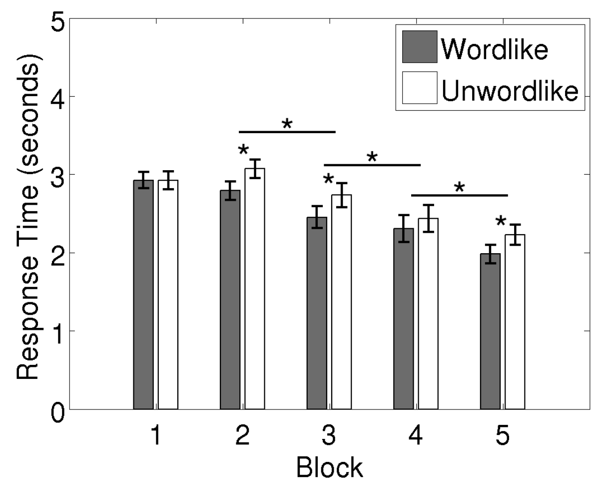

Figure 2: Novel word recognition RT on correct trials. Asterisks indicate significance at $p < .05$, error bars indicate standard error.

Recognition Response Time

On RTs for correct trials, there were main effects of Block, $F_1(4,76) = 16.29$, $p < .001$, $F_2(4,184) = 66.68$, $p < .001$, and Condition $F_1(1,19) = 16.96$, $p < .001$, $F_2(1,46) = 3.97$, $p = .052$, and an interaction, $F_1(4,76) = 3.18$, $p < .05$, $F_2(4,184) = 2.58$, $p < .05$.

Follow up comparisons (Figure 2) revealed that RTs improved ($ps < .05$) between each block, except from block one to two (Block One: M = 2.92 s, SD = 0.45 s; Block Two: M = 2.93 s, SD = 0.48 s; Block Three: M = 2.59 s, SD = 0.65 s; Block Four: M = 2.37 s, SD = 0.75 s; Block Five: M = 2.10 s, SD = 0.54 s). Overall, RTs were faster in wordlike (M = 2.49 s, SD = 0.48 s) than unwordlike (M = 2.68 s, SD = 0.48 s) trials, $p < .05$. The interaction revealed that the wordlike RT advantage was only in blocks two, three, and five ($ps < .05$).

Production Accuracy

Only trials where participants produced the exact target were coded as correct, as a conservative measure of accuracy. There were significant main effects of Block, $F_1(4,76) = 136.4$, $p < .001$, $F_2(4,184) = 420.58$, $p < .001$, and Condition on accuracy $F_1(1,19) = 80.22$, $p < .001$, $F_2(1,46) = 16.96$, $p < .001$, and a marginal interaction $F_1(4,76) = 2.33$, $p = .064$, $F_2(4,184) = 3.17$, $p < .05$.

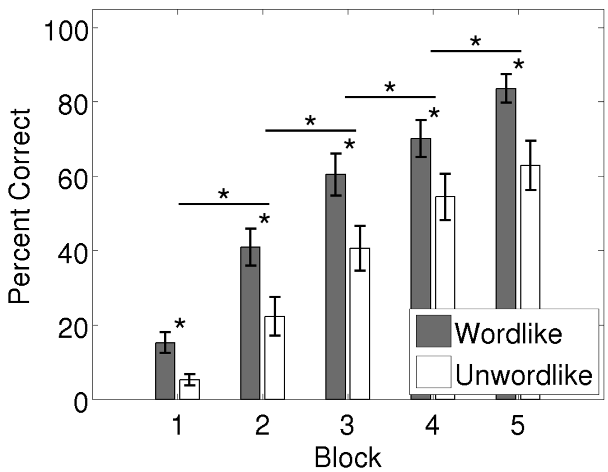

Figure 3: Novel word production accuracy. Asterisks indicate significance at $p < .05$, error bars indicate standard error.

Follow up comparisons (Figure 3) revealed that accuracy improved ($ps < .001$) on each block (Block One: M = 10.2%, SD = 8.9%; Block Two: M = 31.2%, SD = 22.1%; Block Three: M = 50.5%, SD = 25.4%; Block Four: M = 62.3%, SD = 24.0%; Block 5: M = 73.2%, SD = 22.3%). Overall accuracy was higher for wordlike (M = 54.0%, SD = 18.3%) compared to unwordlike(M = 37.1%, SD = 21.6%) trials, $p < .001$. The Condition effect was present in each block ($ps < .05$); the interaction was driven by a smaller Condition effect (i.e., wordlike minus unwordlike) in Block One compared to Blocks two, three, and five (Block One: M = 10.0%, SD = 9.4%; Block Two: M = 18.5%, SD = 10.1%; Block Three: M = 19.8%, SD = 13.2%; Block Four: M = 15.6%, SD = 16.3%; Block Five: M = 20.6%, SD = 19.0%), likely due to low overall accuracy in Block One.

Learning rate (i.e., the average improvement in accuracy between blocks) was marginally higher for wordlike (M = 17.1%, SD = 3.5%) compared to unwordlike (M = 14.4%, SD = 6.7%) trials, $t_1(19) = 1.88$, $p = .08$, $t_2(46) = 3.96$, $p < .001$.

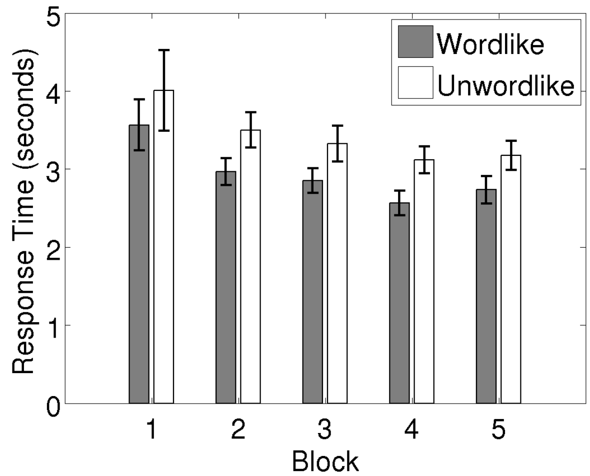

Figure 4: Novel word production RT on correct trials, error bars indicate standard error. Overall RT was faster on Wordlike trials compared to Unwordlike trials.

Production Response Time

On RTs for correct trials, there were significant main effects of Block $F_1(4,76) = 5.77$, $p < .001$, $F_2(4,184) = 6.62$, $p < .001$, and Condition, $F_1(1,19) = 20.20$, $p < .001$, $F_2(1,46) = 7.40$, $p < .01$, but no interaction $F_1(4,76) = 0.11$, ns, $F_2(4,184) = 0.34$, ns.

Follow up comparisons (Figure 4) revealed modest improvements in RT over time, with block four faster than one or two, and block five faster than one ($ps < .05$) (Block One: M = 3.69 s, SD = 1.53 s; Block Two: M = 3.20 s, SD = 0.81 s; Block Three: M = 3.09 s, SD = 0.77 s; Block Four: M = 2.84 s, SD = 0.61 s; Block Five: M = 2.95 s, SD = 0.78 s). RTs were faster overall ($p < .01$) for wordlike (M = 2.92 s, SD = 0.70) than unwordlike (M = 3.37 s, SD = 0.80) trials.

Individual Differences

To examine how cognitive factors contribute to early word learning, stepwise multiple regressions including phonological memory (PM), English vocabulary size (EV), and nonverbal IQ (IQ) were used to predict participants' wordlike and unwordlike scores, collapsed across blocks.

For recognition task accuracy, the final model for wordlike items included PM and EV, $F(2, 16) = 11.429$, $p < .001$, $R^2 = .588$, Adj.$R^2 = .537$. PM was a better individual predictor ($r = .680$) than EV ($r = .458$). The final model for unwordlike items included only PM, $F(1, 17) = 14.65$, $p < .01$, $R^2 = .463$, Adj.$R^2 = .431$. No factors predicted recognition task RT.

For production task accuracy, the final model for wordlike items included PM and EV, $F(2, 16) = 9.84$, $p < .01$, $R^2 = .551$, Adj.$R^2 = .495$. PM was a better individual predictor (r

= .656 than EV (r = .455). The final model for unwordlike items included only PM, $F(1, 17) = 8.23$, $p < .05$, $R^2 = .326$, Adj.$R^2 = .286$. For production task RT, the final model for wordlike items included only PM, $F(1, 17) = 6.65$, $p < .05$, $R^2 = .281$, Adj.R2 = .239. No factors predicted unwordlike RT. IQ was not correlated with any measures of performance.

Discussion

As expected, participants improved in performance over time, becoming faster and more accurate on both the novel word recognition and production tasks. Previous work has shown that phonological neighborhood size and phonotactic probability affect learning of novel auditory words (Frisch et al., 2000; Luce & Large, 2001; Thorn & Frankish, 2005; Roodenrys & Hinton, 2002; Storkel et al., 2006). Here, we show that orthographic wordlikeness affects written word recognition and production over time, and interacts with vocabulary size and working memory to impact learning success.

In the production task, novel words that resembled English were much more accurately produced, with accuracy 10-20% higher than unwordlike items at all time points. Even unwordlike items that were correctly produced were nearly half a second slower than wordlike items. Participants were, however, able to eventually learn a large number of unwordlike items (62.9% by the fifth block), and there was only weak evidence for a difference in learning rate between conditions. While wordlikeness is not the sole predictor of learning, it was a consistent factor in performance over time.

In the recognition task, only modest effects of wordlikeness were observed. Accuracy approached ceiling in blocks four and five, but in blocks two and three, wordlike items were correctly recalled significantly better than unwordlike items, and faster in blocks two, three, and five. Because the novel word was provided in recognition trials and the participants' task was only to select the matching picture from four choices, memory demands for the novel words' forms were low. The fact that wordlikeness effects still emerged suggests that constructing the link between a word and its referent may have been easier for the wordlike items. In post-experiment debriefings, participants tended to report that they attempted to learn the novel words by creating visual associations (e.g., "The cat *purred* on the *fence*" to remember 'purd' = fence), similar to the successful keyword method (Shapiro & Waters, 2005). It is possible that it was easier for participants to generate useful, robust associations for the wordlike items, accelerating learning of these words' meanings. Future studies could specifically address this by asking participants to provide the mental associations they used to learn each word.

Both vocabulary size and phonological memory were related to individual learning ability, but in different ways. As expected, higher English vocabulary size was associated with higher accuracy for wordlike items only, but did not predict performance on unwordlike items. This lends support to the idea that as memories for novel words began to degrade, they could be reconstructed based on existing language knowledge, such as lexical similarity (i.e., neighbors) and sublexical sequence typicality (e.g., gram and bigram probabilities). More direct measures of participants' lexical and sublexical knowledge (e.g., probing which of the novel words' neighbors participants actually know, or asking participants to rank order selected bigrams by frequency) may help to determine how different aspects of participants' existing language knowledge affect word learning success.

Phonological memory in particular, and working memory in general, have been previously associated with novel word learning skill (Baddeley, Gathercole, & Papagno, 1998; Gathercole, 2006; Majerus et al., 2008; Martin & Ellis, 2012), but how this increase interacts with wordlikeness has not been explored. In the current study, we found that phonological memory was associated to a similar degree with increased learning for both wordlike and unwordlike items. This suggests that short-term storage and rehearsal of the novel words was not affected by existing language knowledge, but reflects a general language learning ability.

Nonverbal IQ was a poor predictor for all measures, failing to reach significance. IQ has previously been shown to affect other domains of second language learning such as listening comprehension (Andringa & Olsthoorn, 2012) or grammar learning (Kempe et al., 2010), but may not be a large factor on memory-based tasks such as vocabulary learning.

In conclusion, our results show that acquisition of novel written words is affected by how closely they resemble words the learner already knows. Wordlikeness had a strong effect on novel word production accuracy, and also influenced translation recognition and response times. Vocabulary learning is an especially important part of acquiring a second language and these results suggest that early stages of language instruction could benefit from focusing on words that resemble a learner's native language.

Acknowledgments

This research was funded in part by grant NICHD R01 HD059858-01A to the second author. The authors would like to thank Sana Ali for help testing participants, and the members of the Northwestern University *Bilingualism and Psycholinguistics Research Group* for comments on this work.

References

Andringa, S., & Olsthoorn, N. (2012). Determinants of success in native and nonnative listening comprehension: An individual differences approach. *Language Learning*, *62*(S2), 49–78. doi:10.1111/j.1467-9922.2012.00706.x

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. doi:10.1037//0033-295X.105.1.158

Bailey, T., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*(4), 568–591. doi:10.1006/jmla.2000.2756

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*(1), 1–27. doi:10.1177/00238309050480010101

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. doi:10.3758/BF03196494

Birdsong, D. (2009). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *New handbook of second language acquisition*. Amsterdam: Elsevier.

Burt, J. S., & Blackwell, P. (2008). Soundspelling consistency in adults' orthographic learning. *Journal of Research in Reading*, *31*(1), 77–96. doi:10.1111/j.1467-9817.2007.00362.x

De Groot, A. M. B., & Keijzer, R. (2000). What Is Hard to Learn Is Easy to Forget: The Roles of Word Concreteness, Cognate Status, and Word Frequency in Foreign-Language Vocabulary Learning and Forgetting. *Language Learning*, *50*(1), 1–56. doi:10.1111/0023-8333.00110

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1–25. doi:10.1111/j.0023-8333.2005.00294.x

Dunn, L. M. (1997). Examiner's Manual for the PPVT-III: Peabody Picture Vocabulary Test-Third Edition. Circle Pines, MN: American Guidance Service.

Frisch, S. A., Large, N. R., & Pisoni, D. B. (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, *42*(4), 481–496. doi:10.1006/jmla.1999.2692

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513–543. doi:10.1017/S0142716406060383

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47. Retrieved from http://www.jstatsoft.org/v45/i07/

Kempe, V., Brooks, P., & Kharkhurin, A. (2010). Cognitive predictors of generalization of Russian grammatical gender categories. *Language Learning*, *60*(1), 127–153. doi:10.1111/j.1467-9922.2009.00553.x

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes*, *16*(5-6), 565–581. doi:10.1080/01690960143000137

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: psycholinguistic approaches* (pp. 49–67). New York: Oxford University Press.

Majerus, S., Poncelet, M., van der Linden, M., & Weekes, B. S. (2008). Lexical learning in bilingual adults: The relative importance of short-term memory for serial order and phonological knowledge. *Cognition*, *107*(2), 395–419. doi:10.1016/j.cognition.2007.10.003

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PloS One*, 7(8), e43230. doi:10.1371/journal.pone.0043230

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, *34*(3), 379–413. doi:10.1017/S0272263112000125

PsychCorp. (1999). Wechsler abbreviated scale of intelligence (WASI). San Antonio, TX: Harcourt Assessment.

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 29–33. doi:10.1037//0278-7393.28.1.29

Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception*, *33*(2), 217–236. doi:10.1068/p5117

Shapiro, A. M., & Waters, D. L. (2005). An investigation of the cognitive processes underlying the keyword method of foreign vocabulary learning. *Language Teaching Research*, *9*(2), 129–146. doi:10.1191/1362168805lr151oa

Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*(3), 490–502. doi:10.1017/S1366728911000216

Storkel, H., Armbrüster, J., & Hogan, T. P. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research*, *49*, 1175–1192. doi:10.1044/1092-4388(2006/085)

Thorn, A. S. C., & Frankish, C. R. (2005). Long-term knowledge effects on serial recall of nonwords are not exclusively lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 729–735. doi:10.1037/0278-7393.31.4.729

Vitevitch, M. S., & Luce, P. a. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language*, *40*(3), 374–408. doi:10.1006/jmla.1998.2618

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). The comprehensive test of phonological processing. Austin, TX: Pro-Ed.