CrossMark

Applied Psycholinguistics **38** (2017), 427–456
doi:10.1017/S0142716416000242

Orthographic knowledge and lexical form influence vocabulary learning

JAMES BARTOLOTTI and VIORICA MARIAN
Northwestern University

Received: August 18, 2015 Accepted for publication: April 29, 2016

ADDRESS FOR CORRESPONDENCE
James Bartolotti, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: j-bartolotti@u.northwestern.edu

ABSTRACT
Many adults struggle with second language acquisition but learn new native-language words relatively easily. We investigated the role of sublexical native-language patterns on novel word acquisition. Twenty English monolinguals learned 48 novel written words in five repeated testing blocks. Half were orthographically wordlike (e.g., *nish*, high neighborhood density and high segment/bigram frequency), while half were not (e.g., *gofp*, low neighborhood density and low segment/bigram frequency). Participants were faster and more accurate at recognizing and producing wordlike items, indicating a native-language similarity benefit. Individual differences in memory and vocabulary size influenced learning, and error analyses indicated that participants extracted probabilistic information from the novel vocabulary. Results suggest that language learners benefit from both native-language overlap and regularities within the novel language.

There is considerable variability in adult second language acquisition (SLA; Birdsong, 2009). The difficulties that adult learners encounter are varied, and include challenges in the phonological (Baker & Trofimovich, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, & Díaz, 2006), lexical (Bialystok, 2008), morphological (Franceschina, 2005), and syntactic (DeKeyser, 2005; Tokowicz & MacWhinney, 2005) domains. Even highly proficient second language (L2) speakers often display differences compared to native speakers in the form of accented speech (phonological), vocabulary size (lexical), or preposition use (syntactic). The persistence of these differences has led to debate on the source of adults' SLA difficulties and whether a critical or sensitive period exists for adult SLA (Birdsong, 2014; Friederici, Steinhauer, & Pfeifer, 2002; Johnson & Newport, 1989; Singleton, 2005). Theories of SLA must also expain why adults continue to learn native language vocabulary throughout life (Ramscar, Hendrix, Shaoul, Milin, & Baayen, 2014). In order to distinguish between first- and second-language learning as an adult, learning theories need to account for how the target language interacts with the learner's cognitive skills and specific language background.

Monolingual adults are more adept at learning native language content throughout life. People can shift their phonemic categories after sustained exposure to new dialects (Bigham, 2010; Evans & Iverson, 2007; Munro, Derwing, & Flege, 1999)

© Cambridge University Press 2016 0142-7164/16

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

and continue to learn new words with age (Ramscar et al., 2014; Verhaeghen, 2003). Some of these new words will be neologisms, including such recent additions to the Merriam–Webster English dictionary as *staycation* or *truthiness*, while others constitute low-frequency words or jargon found in many hobbies and professional pursuits, such as *morpheme* or *aphasia*. Word learning aptitude is associated with several cognitive predictors, including short-term memory (Majerus, Poncelet, van der Linden, & Weekes, 2008; Martin & Ellis, 2012) and attention/inhibition (Gass, Behney, & Uzum, 2013; Yoshida, Tran, Benitez, & Kuwabara, 2011). In addition, the characteristics of the word itself can influence success at different stages of lexical acquisition. New words with irregular letter or sound patterns are often better at triggering *initial* stages of learning (Storkel, Armbrüster, & Hogan, 2006), because their novelty makes them more salient when embedded in a context of known words. In the study by Storkel et al., participants encountered new words in sentences, and a low-wordlike advantage was observed during the acquisition phase (before the word was fully learned), because the novelty of the low-wordlike items triggered learning better. However, novel items that resemble one's native language tend to be remembered more accurately *after training* (Frisch, Large, & Pisoni, 2000; Luce & Large, 2001; Roodenrys & Hinton, 2002; Thorn & Frankish, 2005), because their forms are easier to integrate with prior language knowledge. This training advantage is evident from one of the first investigations of lexical similarity (Schwartz & Leonard, 1982), in which young children with limited vocabularies (<5 words) learned novel words consistent with their lexical inventory better after several sessions.

However, the types of linguistic patterns that promote first language (L1) growth may also interfere with L2 vocabulary learning. This is because languages are highly regular, but these regularities diverge across languages. For example, while less than half a percent of the 456,976 possible four-letter English words (26^4) are estimated to exist in speakers' vocabularies (calculations based on CLEARPOND; Marian, Bartolotti, Chabal, & Shook, 2012), the words that do exist share many similarities. These four-letter English words have an average of 10.33 neighbors (words that differ by only one letter), and while certain letters and bigrams (pairs of letters) appear frequently (e.g., CE or LY), others do not appear at all (e.g., ZW or FD). However, similarity drops when you compare across languages: written English words tend to have five to seven times fewer neighbors in related languages like Dutch, French, German, or Spanish than they do English neighbors (Marian et al., 2012), highlighting the language-specific nature of lexical patterns.

Because of the significant differences across languages that make L2 word learning difficult, learners ought to capitalize on any similarities available. Teachers and students have long known to make use of similarities when acquiring an L2 (Ringbom, 2007). Cognates, for example, are an effective learning tool. Cognates are words that overlap in form and meaning across languages, and they are easier to learn than noncognates (De Groot & Keijzer, 2000; Lotto & De Groot, 1998). However, because the inventory of cognates is fairly limited, even between two related languages (Schepens, Dijkstra, Grootjen, & van Heuven, 2013), they cannot make up the entirety of a learner's initial L2 vocabulary. In addition, learners' conservative assumptions about cognates' meanings may make cognates harder to use in context (Rogers, Webb, & Nakata, 2014), limiting their utility to the

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

language learner. Fortunately, there exists another large class of words that capitalizes on the learner's prior language knowledge, which are words whose spellings or sounds happen to adhere to native language patterns. For an English learner of German, words like *sind* or *hinter* (meaning *are* and *behind*) may be easier to learn than atypical words such as *jetz* (meaning *now*), because of the formers' closer resemblance to existing English words. Studies using artificially designed auditory words have shown that phonological wordlikeness does improve novel word repetition and learning (Frisch et al., 2000; Luce & Large, 2001; Roodenrys & Hinton, 2002; Storkel et al., Hogan, 2006; Thorn & Frankish, 2005), and one reason for this advantage may be that highly wordlike sequences, as they begin to degrade in memory, can be more accurately reconstructed.

Although most L2 exposure in immersion and instructional contexts is phonological at early stages, there has been an increase in text-based instruction, especially via web-based learning resources (Golonka, Bowles, Frank, Richardson, & Freynik, 2012; Liu et al., 2013). In addition, when considering the effect of cross-language similarity, orthographic overlap is often higher than phonological overlap in languages that share a script due to the conservation of symbols across languages (Marian et al., 2012). These factors combined make the study of orthographic effects on L2 learning an important area of study.

The *orthographic wordlikeness* of a word in the L2 is its similarity to existing vocabulary and letter patterns in the L1. Here we define wordlikeness based on a combination of neighborhood size, positional segment frequency, and positional bigram frequency. Neighborhood size and segment/bigram frequencies provide two distinct, complementary metrics of typicality at lexical and sublexical levels of processing, respectively (Vitevitch & Luce, 1999). Neighborhood size was defined as the total number of words in the L1 that differed from the target in the substitution of a single letter (Marian et al., 2012). Positional segment frequency was calculated for each letter in a target word as the total log frequency of all words that contain that letter in the same position, relative to the total log frequency of all words containing any letter at that position (Vitevitch & Luce, 2004); frequency estimates are thus token rather than type based, which better accounts for total exposure to a letter. The total positional segment frequency for a word was the sum across all letters in that word. Positional bigram frequency was calculated in a similar manner, based on the frequency of occurrence of each two-letter sequence. We use the abbreviated terms *segment frequency* and *bigram frequency* to refer to total positional segment frequency and total positional bigram frequency.

Wordlikeness is an especially useful aid early in learning, when little is known about the structure of the novel language itself. With increased experience, though, L2 learners are able to utilize L2-specific letter and sound patterns (Stamer & Vitevitch, 2012) or morphological rules (Brooks, Kempe, & Donachie, 2011) to improve acquisition of L2-wordlike vocabulary. Participants in Stamer and Vitevitch's (2012) study had several years of college experience with an L2, but learners may detect basic L2 patterns even earlier. Results from statistical learning of nonsense syllables or tone sequences suggest that the mechanism to extract regularities in the input operates rapidly on novel stimuli (Mirman, Magnuson, Graf Estes, & Dixon, 2008; Newport & Aslin, 2004; Saffran, Johnson, Aslin,

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

& Newport, 1999), and suggests that even early stages of L2 learning may be influenced by L2-specific patterns.

The current study investigates how novel words' similarity to one's native language and to other novel words interacts to affect word learning in a L2. The present study assessed learning of carefully designed words in an artificial language that emulated an SLA context. Novel written words were paired with familiar picture referents (e.g., airplane or cat), and we tracked participants' gains in performance over time on two measures of learning (word recognition and word production) within a single training session. Words were designed to have either high or low English wordlikeness based on their neighborhood size, segment frequency, and bigram frequency. We predicted an accuracy and response time advantage for wordlike items, which are easier to store and retrieve from memory due to their similarity to known lexical patterns in the native language. In addition, if learners intuitively identify letter and bigram patterns within the L2, we should observe characteristic errors in production caused by overgeneralization of L2 patterns.

A secondary goal of the study was to determine how the learner's individual differences in English vocabulary size, short-term memory capacity, and general intelligence affected learning of wordlike and unwordlike items. Because vocabulary learning involves acquisition and storage of novel letter/sound sequences, we expected high memory capacity to predict overall learning success (Majerus et al., 2008; Martin & Ellis, 2012). While nonverbal IQ has been shown to predict learning of linguistic patterns such as grammatical rules (Kempe, Brooks, & Kharkhurin, 2010), its role in vocabulary learning is less clear. Finally, we expected native language vocabulary size to predict learning for wordlike items that are consistent with patterns in the native language.

METHODS

Participants

Twenty English monolinguals (12 females) at a university in the United States participated for monetary compensation or course credit. Informed consent was obtained in accordance with the university's Institutional Review Board. The mean age was 20.3 years ($SD = 1.42$). After the experiment, participants completed a battery of cognitive tests to assess the effects of phonological short-term memory, vocabulary size, and nonverbal intelligence on word learning performance. Assessments and standard scores were as follows: phonological short-term memory, $M = 112.6$, $SD = 12.3$, Comprehensive Test of Phonological Processing, phonological memory composite score of digit span and nonword repetition subtests (Wagner, Torgesen, & Rashotte, 1999); English receptive vocabulary, $M = 119.3$, $SD = 9.9$, Peabody Picture Vocabulary Test—Third Edition (Dunn, 1997); nonverbal IQ, $M = 113.2$, $SD = 6.6$, Wechsler Abbreviated Scale of Intelligence, block design and matrix reasoning subtests (PsychCorp, 1999).

Materials

Forty-eight orthographic consonant–vowel–consonant–consonant words were created in the novel language Colbertian (named after comedy show wordsmith

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Stephen Colbert to engage participants in the learning task). Half of the words were designed to have high English wordlikeness (e.g., *nish* or *baft*). These were formed by substituting one letter of an English word, and had high orthographic neighborhood sizes, segment frequency, and bigram frequency. The other half of the words were unwordlike (e.g., *gofp* or *kowm*), and were formed by substituting multiple letters from English words; these words also had low segment frequency and bigram frequency. Letter and bigram distributions in the Colbertian language perceptibly differed from English; the most frequent letters at onset were G and V, while the most frequent bigrams at offset included FT, TZ, and WM. See Table 1 for orthotactic calculations from CLEARPOND (Marian et al., 2012), and Appendix A for a list of all stimuli.

Although there was no auditory component to the learning task, readers often generate pronunciations for novel written words (Johnston, McKague, & Pratt, 2004), and thus all Colbertian words were also assessed for English phonological wordlikeness. Six English monolinguals (not participants in the current study) pronounced each Colbertian word, and their responses were phonologically transcribed. Responses were 70% consistent across speakers, and more variable in the unwordlike (58.7%, $SD = 28.4$) than the wordlike (80.7%, $SD = 19.6$) conditions, $t\,(46) = 2.89$, $p < .01$. To account for variability in responses, phonological neighborhood size, total positional phoneme frequency, and total biphone frequencies were calculated for each speaker's production (using the same procedure as the orthographic calculations); these individual speaker values were averaged to yield mean scores for each word (using the CLEARPOND database; Marian et al., 2012). The orthographically wordlike and unwordlike lists differed (all $ps < .05$) on all phonological wordlikeness metrics, ensuring that any effects of orthography on learning were not obscured by differences in phonological characteristics. In addition, a phonologically matched subset of 12 wordlike and 12 unwordlike items was created in which the two lists differed in orthographic neighborhood size, total segment frequency, and total bigram frequency (all $ps < .05$), but did not differ in phonological neighborhood size, phoneme frequency, or biphone frequency (all $ps > .1$; Table 1).

Each novel word was paired with a color line drawing from the revised Snodgrass and Vanderwart picture set (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 their paired Colbertian words (all picture-word pairings available in Appendix A). Pictures for wordlike and unwordlike items did not differ on lexical frequency, orthographic or phonological neighborhood size, or mean segment, bigram, phoneme, or biphone frequencies (CLEARPOND; Marian et al., 2012).

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 2 s. Following the exposure block, participants performed five blocks of word

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table 1. *Stimuli characteristics*

Measure	Wordlike ($N = 24$)	Unwordlike ($N = 24$)	t (46)	Orthographic Wordlike ($N = 12$)	Orthographic Unwordlike ($N = 12$)	t (22)
Orthographic neighborhood	5.54 (1.72)	0 (0)	15.79***	5.42 (1.68)	0 (0)	11.19***
Segment frequency	0.289 (0.042)	0.223 (0.048)	4.96***	0.286 (0.043)	0.220 (0.050)	3.46**
Bigram frequency	0.026 (0.011)	0.010 (0.009)	5.16***	0.024 (0.010)	0.011 (0.011)	3.06**
Phonological neighborhood	16.27 (11.05)	9.45 (5.74)	2.68*	11.38 (7.75)	10.67 (6.40)	0.24
Phoneme frequency	0.191 (0.055)	0.155 (0.058)	2.19*	0.176 (0.057)	0.176 (0.054)	0.003
Biphone frequency	0.012 (0.010)	0.006 (0.005)	2.49*	0.010 (0.009)	0.007 (0.005)	0.84

Note: Orthographic wordlike and orthographic unwordlike subsets were designed to be matched on phonological wordlikeness metrics. Segment, bigram, phoneme, and biphone metrics are total positional frequencies. Values are means (standard deviations).
*$p < .05$. **$p < .01$. ***$p < .001$.

recognition with feedback, and five blocks of word production with feedback, alternating between the two tasks. After each block, participants were given a short self-paced break (totaling on average 69.7 s, $SD = 14.6$, across the experiment). The entire learning and testing procedure lasted on average 55.1 min ($SD =$ 12.5 min).

Word learning: Recognition. Each testing block included 48 recognition trials. In each trial, a randomly selected target picture and three distractors were displayed in the four corners of the screen, and the written target word appeared in the center of the screen. The distractors were other training items selected at random, with the constraint that each picture appeared once as a target and three times as a distractor within a single testing block. The participant was instructed to click on the correct picture, and accuracy and response time (RT) were recorded. After making a response, the three distractors disappeared, and the target picture and written word remained onscreen for 1000 ms. Because the word–picture pair remained visible during feedback, participants could use it to relearn the correct association.

Word learning: Production. In each testing block, the recognition task was followed by 48 production trials. In each trial, a randomly selected 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. Errors in production were collected and analyzed separately. 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. By providing the correct answer as feedback, participants were able to improve their memory for the correct picture–word association. After completing all 48 trials, a new testing block of recognition and production began. After the fifth series of recognition and production blocks, the experiment concluded.

Data analysis

Accuracy and RT. RT analyses were performed on correct responses only in order to control for accuracy differences between blocks and conditions. Excessively long outlier RTs were identified within each combination of block and condition using a threshold of mean plus 2 *SD* (4.2% of all trials). Outliers were replaced with the threshold ($M + 2\ SD$) for that combination in order to minimize noise from long trials. All analyses were repeated not omitting outliers and the same pattern of results was observed; for brevity, only results with adjusted outliers are reported.

Change across blocks in accuracy and RT in the recognition and production tasks was analyzed using growth curve analysis (Mirman, Dixon, & Magnuson, 2008; Mirman, Magnuson, Graf Estes, & Dixon, 2008). Growth curve analysis is a form of multilevel regression that simultaneously estimates the effects of individuals and of experimental manipulations on timecourse data. Accuracy and RT data were first fit with second-order orthogonal polynomials to capture the curvilinear shape of learning gains over time. Each of the polynomial terms in

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

these base models was then estimated in a Level 2 model that assessed the effects of wordlikeness and of participants on the parameter estimate from the base model. In these models, changes in the intercept term correspond to changes in the average height of the curve across the window of analysis (i.e., across all five testing blocks). The linear term reflects the overall slope of the curve, while the quadratic reflects its curvature. The base model included all time terms and random effects of participant and participant by condition on all time terms. Additional models were built that added a fixed effect of condition (wordlike and unwordlike items) to each time variable in turn. Significant improvements in model fit (chi-square test on the –2 log likelihood change in model fit) indicated an effect of condition on independent properties of the curve (i.e., height, slope, and curvature). To assess the effect on learning of individual differences in phonological memory, English vocabulary size, and nonverbal IQ, these individual difference predictors were added to the full model as fixed effects after recentering each predictor to zero. The change in model fit was assessed for each factor on the different time terms.

Error analysis. Production errors were analyzed to determine the influence of other languages on word recall. The distribution of incorrect single-letter substitutions can be used to isolate the effects of interference from Colbertian and English letter knowledge. Errors were analyzed at each letter position within the target word structure separately. Because all Colbertian letters were of the form consonant–vowel–consonant–consonant, and only five letters were valid in the vowel position, we restricted analyses to the three consonant positions (i.e. letter positions 1, 3, and 4). At each position, the relative frequencies with which each letter in the alphabet was used incorrectly across all items and participants was calculated in each block. As predictor variables, we calculated positional segment frequencies in the complete Colbertian vocabulary and for four-letter English words as in Vitevitch and Luce (2004). Colbertian positional segment frequencies for each letter were calculated as the number of times that letter appeared in a Colbertian word at that position, divided by the total number of Colbertian words (because each word appeared the same number of times in the experiment, Colbertian words have equivalent token frequency, obviating the need for a log frequency transformation). English positional segment frequencies for each letter were calculated as the sum of SUBTLEX log frequencies for all four-letter words containing that letter in that position, divided by the sum of SUBTLEX log frequencies for all four-letter English words. Despite some notable differences across the two languages (e.g., the letters G and V are more common onsets in Colbertian than English), segment frequencies were distributed similarly, and no differences between the two languages were observed at any letter position (all $ps > .1$).

At each of the three letter positions, three linear models were constructed with letters' error frequencies as the dependent value. The base model included a single predictor, time (i.e., testing block). The Colbertian model included both time and Colbertian positional segment frequencies as predictors, and the English model included both time and English positional segment frequency as predictors. To compare how well English and Colbertian segment frequencies predicted actual errors, three model comparisons were performed at each letter position (letters that were never produced or that had zero frequency in that position in English or

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Colbertian were excluded to enable model comparisons). The Colbertian model and the English model were each compared to the base model using likelihood ratio tests. Then the Colbertian model was compared to the English model using Vuong's nonnested model comparisons (Merkle, You, & Preacher, 2015; Vuong, 1989).

RESULTS

Novel word recognition

Recognition accuracy improved over time and reached an asymptote at ceiling performance, improving from 64.4% (SD = 13.1) in the first block to 97.7% (SD = 4.8) in the fifth block. The base model was fit with a second-order polynomial: linear term, change in log likelihood (ΔLL) = 8.48, χ^2 (1) = 16.96, $p <$ .0001; quadratic term, ΔLL = 15.17, χ^2 (1) = 30.35, $p <$ .0001. RTs became faster over time, from 2921 ms (SD = 446) in the first block to 2102 ms (SD = 539) in the fifth block. The base RT model was fit with a second-order polynomial: linear term, ΔLL = 14.50, χ^2 (1) = 27.00, $p <$ .0001; quadratic term, ΔLL = 0.53, χ^2 (1) = 1.06, $p >$ .1.

Wordlikeness affected the shape of the learning curve for accuracy. Wordlike items rose in accuracy more steeply initially between blocks, and reached an asymptote at mastery performance earlier than unwordlike items, but there was no difference in the amount learned across the experiment. In model comparisons, there was a marginal effect of wordlikeness on the intercept, ΔLL = 1.58, χ^2 (1, 17) = 3.16, p = .075, no effect on the linear term, ΔLL = 0.01, χ^2 (1) = 0.02, $p >$ .1, and a significant effect on the quadratic term, ΔLL = 2.33, χ^2 (1) = 4.67, $p <$ .05. Observed data and the fitted models for wordlike and unwordlike conditions are shown in Figure 1.

Wordlikeness also increased overall speed of recognition, as pictures corresponding to wordlike items were identified faster than those corresponding to unwordlike items: intercept, ΔLL = 6.41, χ^2 (1) = 12.81, $p <$ .001. The rate at which RT improved between blocks, and the shape of the RT curve over time, however, were not affected by items' wordlikeness: linear, ΔLL = 0.63, χ^2 (1) = 1.28, $p >$ .1; quadratic, ΔLL = 0.98, χ^2 (1) = 1.96, $p >$ .1. Observed data and the fitted models are shown in Figure 2.

Novel word production

Production accuracy improved from 10.2% (SD = 8.9) in the first block to 73.2% (SD = 22.3) in the fifth block. The base model was fit with a second-order polynomial, and though overall accuracy increased between blocks: linear term, ΔLL = 23.37, χ^2 (1) = 46.74, $p <$ .0001, the increase was largest in earlier blocks: quadratic term, ΔLL = 3.89, χ^2 (1) = 7.78, $p <$ .01. RTs on correct trials became faster over time, improving from 3566 ms (SD = 1335) in the first block to 2907 ms (SD = 789) in the fifth block. RT was also fit with a second-order polynomial, and RT decreased between blocks: linear term, ΔLL = 2.64, χ^2 (1) = 5.27,

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

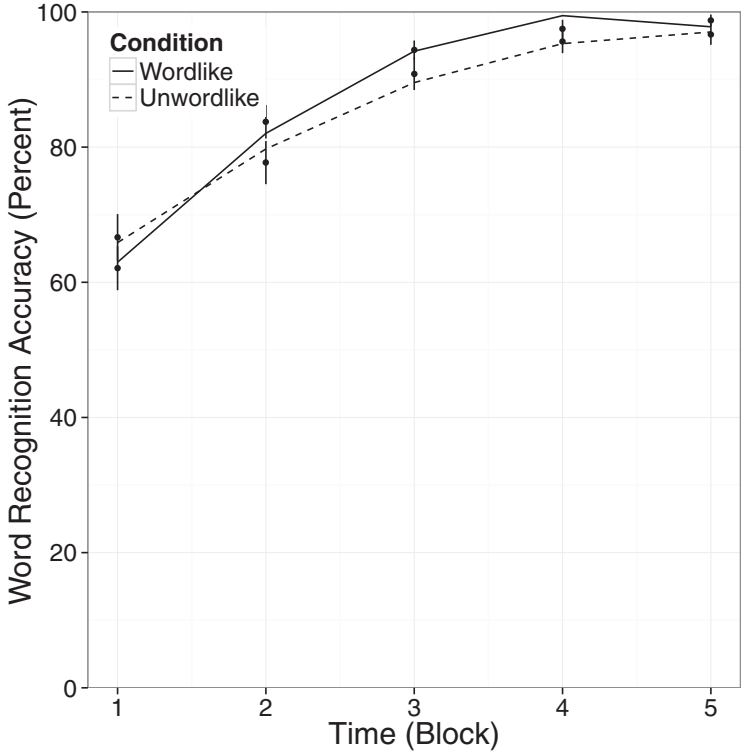


Figure 1. Novel word recognition accuracy. Dots mark observed data, and lines are best fit quadratic growth curve models. Wordlike items (solid) had significantly steeper curvature (effect of wordlikeness on the quadratic term); English vocabulary size positively affected learning rate (linear term).

$p < .05$, with the largest decrease in earlier blocks: quadratic term, $\Delta LL = 2.48$, $\chi^2\ (1) = 4.97$, $p < .05$.

Wordlike items were produced more accurately throughout the experiment than unwordlike items, but the rate of improvement for the two types of words was comparable. Adding wordlikeness to the intercept significantly improved model fit, $\Delta LL = 17.88$, $\chi^2\ (1) = 35.77$, $p < .0001$, but it had no effect on the linear, $\Delta LL = 0.96$, $\chi^2\ (1) = 1.91$, $p > .1$, or quadratic terms, $\Delta LL = 1.07$, $\chi^2\ (1) = 2.16$, $p > .1$. Observed data and the fitted models are shown in Figure 3. Wordlikeness improved accuracy by 16.9%, a difference of roughly four words. The lack of an effect on either the linear or quadratic term indicates that accuracy for both types of words improved at a similar rate over time, and the relative difference between the two conditions remained the same.

Similar to the recognition task, RTs were faster overall for wordlike items compared to unwordlike items, but both item types decreased at similar rates. Adding wordlikeness to the intercept significantly improved model fit,

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

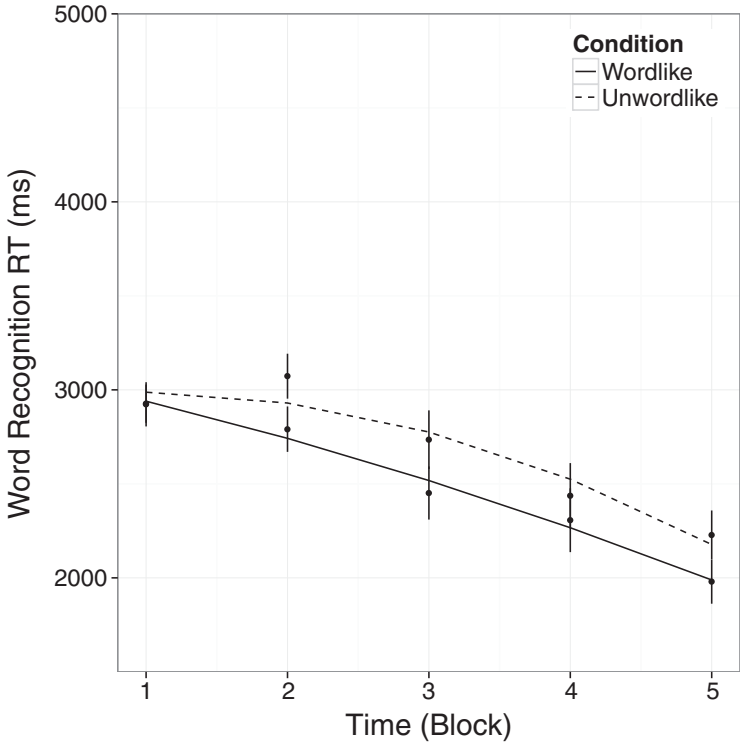


Figure 2. Novel word recognition response time for correct trials. Dots mark observed data, and lines are best fit quadratic growth curve models. Wordlike items (solid) were recognized faster (effect of wordlikeness on intercept), and higher memory capacity was associated with faster response times.

$\Delta LL = 9.87$, χ^2 (1) = 19.74, $p < .0001$, but it had no effect on the linear, $\Delta LL = 0.01$, χ^2 (1) = 0.01, $p > .1$, or quadratic terms, $\Delta LL = 0.03$, χ^2 (1) = 0.06, $p > .1$. Observed data and the fitted models are shown in Figure 4. Correct wordlike items were overall 480 ms faster than correct unwordlike items in the model, indicating that wordlike items were easier to produce.

Relationship between recognition and production

To determine whether word recognition and production performance were affected by a common set of lexical representations in the novel language, correlations between accuracy and RT in the two tasks were conducted. For accuracy, recognition and production were strongly correlated in each testing block, Block 1: r (18) = .65, $p < .01$; Block 2: r (18) = .79, $p < .001$; Block 3: r (18) = .77, $p < .001$; Block 4: r (18) = .76, $p < .001$; and Block 5: r (18) = .71, $p < .001$, suggesting that participants accessed common representations in the two tasks. RTs were not

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

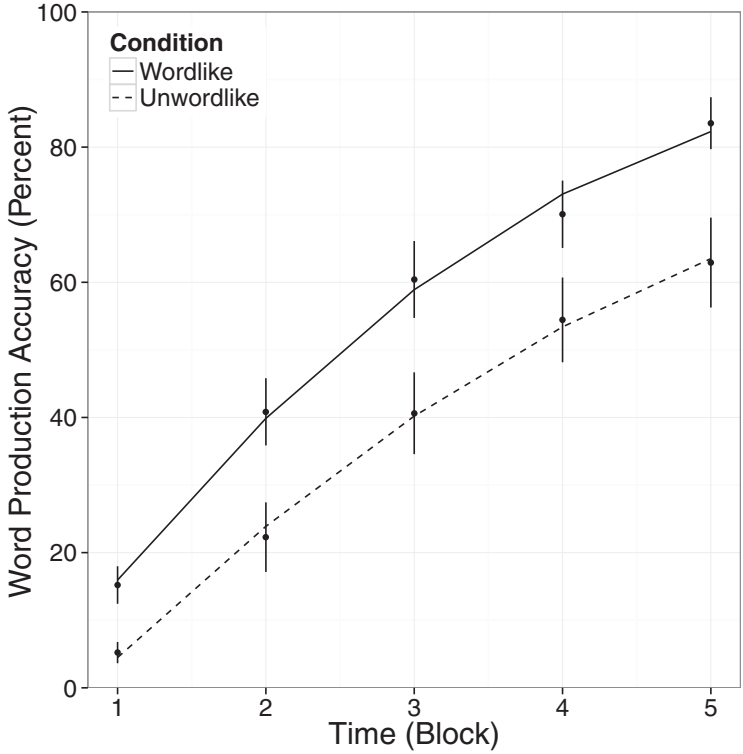


Figure 3. Novel word production accuracy. Dots mark observed data, and lines are best fit quadratic growth curve models. Wordlike items (solid) were produced more accurately (effect of wordlikeness on intercept); wordlike accuracy was positively associated with English vocabulary size. Larger memory capacity was associated with faster learning rate for both item types (effect on linear term).

correlated in the first four testing blocks, Block 1: $r\,(18) = .31$, $p > .1$; Block 2: $r\,(18) = .37$, $p > .1$; Block 3: $r\,(18) = .26$, $p > .1$; and Block 4: $r\,(18) = .09$, $p > .1$, but in the final testing block, there was a moderate correlation between the two tasks, $r\,(18) = .47$, $p < .05$. The lack of correlation until participants became most fluent in the novel language reflects the large difference in response type between the two tasks (point and click for recognition versus typing a word in production).

Phonological control

To control for the effect of phonological wordlikeness on performance, analyses were run on a subset of wordlike and unwordlike trials closely matched for phonological similarity. For word production accuracy, adding orthographic wordlikeness to the intercept significantly improved model fit, $\Delta LL = 7.16$, $\chi^2\,(1) = 14.31$, $p < .0001$, but it had no effect on the linear, $\Delta LL = 0.64$, $\chi^2\,(1) = 1.27$,

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

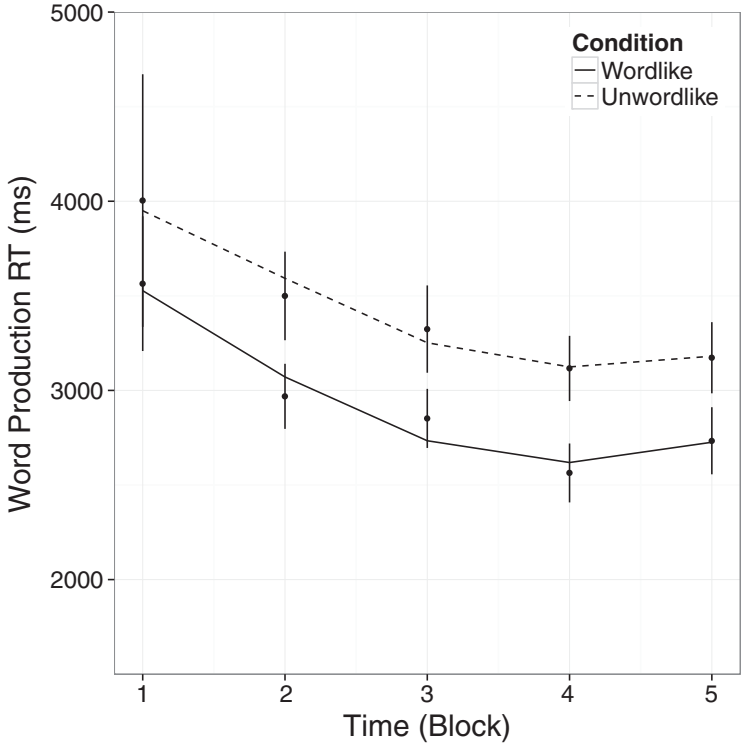


Figure 4. Novel word production response time for correct trials. Dots mark observed data, and lines are best fit quadratic growth curve models. Wordlike items (solid) were recognized faster (effect of wordlikeness on intercept), and higher memory capacity was associated with faster response times.

$p > .1$, or quadratic, $\Delta LL = 0.22$, χ^2 (1) $= 0.45$, $p > .1$, terms. The overall curve height was increased by wordlikeness (estimate $= 9.6$, $SE = 2.1$, $p < .001$), though the size of this increase is lower than that observed in the full analysis. For word production RTs, adding orthographic wordlikeness to the intercept significantly improved model fit, $\Delta LL = 2.02$, χ^2 (1) $= 4.04$, $p < .05$, but it did not affect the linear, $\Delta LL = .01$, χ^2 (1) $= 0.01$, $p > .1$, or quadratic, $\Delta LL = 0.11$, χ^2 (1) $= 0.22$, $p > .1$, terms. Wordlikeness reduced the overall curve height (estimate $=$ –224.8, $SE = 112.6$, $p < .05$), to a lesser degree than in the full analysis.

For word recognition, orthographic wordlikeness did not improve model fit for recognition accuracy, intercept model: $\Delta LL = 0.36$, χ^2 (1) $= 0.73$, $p > .1$; linear model: $\Delta LL = 0.44$, χ^2 (1) $= 0.89$, $p > .1$; quadratic model $\Delta LL = 0.19$, χ^2 (1) $= 0.37$, $p > .1$, or recognition RT: intercept model: $\Delta LL = 0.21$, χ^2 (1) $= 0.41$, $p > .1$; linear model: $\Delta LL = 1.23$, χ^2 (1) $= 2.47$, $p > .1$; quadratic model $\Delta LL = 0.30$, χ^2 (1) $= 0.61$, $p > .1$. The effect of wordlikeness on performance in the recognition task was small in the overall analysis, and the subset analysis shows that orthographic similarity alone did not yield a detectable effect.

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Bigram legality

Because the unwordlike item set varied in whether they contained any bigrams that were illegal in English (e.g., the bigram FP in *gofp*), we conducted a secondary analysis comparing wordlike ($N = 24$), legal-unwordlike ($N = 9$), and illegal-unwordlike ($N = 15$) items. In all analyses, the legal-unwordlike condition was used as the reference level. For accuracy on the production task, adding condition to the intercept significantly improved model fit, $\Delta LL = 22.07$, χ^2 (2) = 44.14, $p < .001$. Illegal items were learned less overall than legal-unwordlike items (intercept estimate = –7.0, $SD = 2.3$, $p < .01$), and wordlike items were learned better overall (intercept estimate = 12.6, $SD = 2.3$, $p < .001$). For RT on the production task, adding condition to the intercept significantly improved model fit, $\Delta LL = 11.94$, χ^2 (2) = 23.88, $p < .001$. Illegal items were produced slower overall than legal-unwordlike items (estimate = 541, $SD = 140$, $p < .001$), but there was no difference between legal and wordlike items (estimate = –160, $SD = 122$, $p > .1$). This pattern of results suggests that items containing novel bigrams were especially difficult to recall and produce. Low similarity while not violating English orthotactic rules (i.e., the legal-unwordlike condition) still led to a large drop in accuracy compared to high English similarity, but did not affect speed.

For recognition accuracy, there were marginal effects of condition on the intercept, $\Delta LL = 2.6$, χ^2 (2) = 5.19, $p < .1$, and quadratic terms, $\Delta LL = 2.5$, χ^2 (2) = 5.01, $p < .1$, and a significant effect of condition on the linear term, $\Delta LL = 5.5$, χ^2 (2) = 10.90, $p < .01$. The illegal items were learned less overall (estimate = –2.9, $SD = 1.3$, $p < .05$) and had a shallower slope (estimate = 8.8, $SD = 2.6$, $p < .001$) than the legal-unwordlike condition. In contrast, the wordlike items were learned faster (estimate = 8.4, $SD = 2.6$, $p < .01$) and reached ceiling performance earlier (quadratic term: estimate = –6.2, $SD = 2.8$, $p < .05$). This pattern of results was caused by higher initial accuracy for the legal- unwordlike condition compared to wordlike items, followed by shallower learning gains (thus the lack of an intercept effect). For recognition RTs, there were significant effects of adding condition to the intercept, $\Delta LL = 10.74$, χ^2 (2) = 21.48, $p < .001$, and linear terms, $\Delta LL = 5.34$, χ^2 (2) = 10.67, $p < .01$. Illegal items were slower overall than the legal-unwordlike condition (estimate = 249, $SD = 540$, $p < .001$). The legal-unwordlike items had a shallower slope than both the illegal items (estimate = –325, $SD = 108$, $p < .01$) and the wordlike items (estimate = –300, $SD = 108$, $p < .01$). This pattern of results was caused by the legal-unwordlike items being responded to faster in the initial testing block, followed by smaller speed increases between blocks compared to the other two conditions.

Error analysis

Error production by block and letter position was analyzed with a 5 (block) × 4 (letter position) within-subjects analysis of variance; mean errors for each combination are provided in Table 2. Errors decreased over time: main effect of block, F (4, 76) = 105.6, $p < .001$, with each block containing fewer errors than the

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table 2. *Percentage errors by letter-position and block*

Block	First Letter (Consonant1)	Second Letter (Vowel)	Third Letter (Consonant2)	Fourth Letter (Consonant3)
1	63.3 (20.0)	52.7 (16.3)	68.0 (17.2)	67.5 (18.4)
2	42.2 (19.0)	34.7 (18.8)	48.3 (21.2)	52.8 (22.0)
3	26.7 (18.9)	21.1 (16.0)	31.9 (20.0)	36.6 (21.1)
4	16.6 (14.5)	13.3 (13.9)	22.0 (19.4)	25.3 (20.0)
5	12.1 (13.4)	10.2 (11.1)	17.0 (16.1)	20.4 (18.6)

Note: Values are participant means (standard deviations).

preceding block (all ps $< .05$, Holm correction). In addition, errors were unevenly distributed across letter positions in a word: main effect of letter-position, F (3, 57) $= 27.69$, $p < .001$. All letter positions were significantly different from each other (all ps $< .05$, Holm correction), with error percentage increasing from the vowel, to the first consonant, the second consonant, and the final consonant in order. The low vowel error rate reflects the smaller inventory of only five letters at that position, compared to 14–15 possible consonants at the other positions. In addition, an interaction of block and position, F (12, 228) $= 2.98$, $p < .001$, captured a change in consonant error from uniformly high in block one (all comparisons $p > .1$) to more nuanced with increased training, including a sharp decrease in first-consonant errors.

To examine participants' ability to extract Colbertian regularities, the effects of Colbertian and English letter frequencies on the distribution of incorrect single-letter substitutions was examined. At each of the three consonant positions within a word, three linear models were created with the frequency of errors for individual letters as the dependent variable, a base linear model containing only time (testing block) as a predictor, and two language models that included either Colbertian segment frequencies or English segment frequencies as an additional predictor.

In the first consonant position, both Colbertian segment frequency and English segment frequency significantly improved fit compared to the base model (Colbertian model: likelihood ratio [LR] $= 28.95$, $p < .001$; English model: $LR = 11.48$, $p < .001$). However, the Colbertian model was a significantly better fit to the data than the English model (Vuong's nonnested model comparison $z = 1.65$, $p < .05$). In the second consonant position, only Colbertian segment frequency improved model fit (Colbertian model: $LR = 42.78$, $p < .001$; English model: $LR = 17.93$, $p > .1$). Accordingly, the Colbertian model provided significantly better fit than the English model ($z = 2.24$, $p < .05$). In the final consonant position, both Colbertian and English segment frequencies improved model fit (Colbertian model: $LR = 34.33$, $p < .01$; English model: $LR = 22.06$, $p < .05$); however, in this position there was no difference in model fit between the two languages ($z = 1.23$, $p > .1$).

Whereas the frequency with which letters appeared in the Colbertian vocabulary predicted the types of errors participants made at all consonant positions in a word,

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

English frequency was only related to errors at word onset and offset. In addition, Colbertian was a better predictor of errors than English for two letter positions, and as good a predictor as English for the remaining position. These results indicate that even when participants were unable to completely retrieve a target item, they filled in gaps in their recollection based on their knowledge of sublexical letter patterns, especially those patterns present in Colbertian. This specific sensitivity to letter frequencies in Colbertian suggests that as participants attempted to learn whole words, they were also able to identify regularities in the input and used this language-general knowledge to supplement their memory for individual words.

Individual differences

The effects of English vocabulary size, phonological memory capacity, and nonverbal intelligence on accuracy and RT was explored in both the recognition and production tasks. For accuracy on the recognition task, adding English vocabulary size to the model had a significant effect on the linear term, $\Delta LL = 3.29$, χ^2 (1, 12) = 6.58, $p < .05$, suggesting that a large native language vocabulary can improve the rate of learning novel words' meanings. This effect on the linear term was present for both wordlike, $\Delta LL = 2.75$, χ^2 (1, 12) = 5.50, $p < .05$, and unwordlike, $\Delta LL = 2.18$, χ^2 (1, 12) = 4.35, $p < .05$, items analyzed separately, indicating that the effect of vocabulary size was not dependent on novel words' orthographic structure.

Adding phonological memory capacity to the model for RT in the recognition task significantly improved estimation of the intercept, $\Delta LL = 4.30$, χ^2 (1, 11) = 8.60, $p < .05$, indicating that individuals with larger memory capacity identified correct word–picture associations faster than individuals with lower memory capacity throughout training. The absence of an effect of memory on the linear or quadratic terms indicates that the improvements in RT that were observed over training did not depend on memory capacity. In addition, the effect of memory on the intercept was observed for both wordlike, $\Delta LL = 2.42$, χ^2 (1, 11) = 4.83, $p < .05$, and unwordlike, $\Delta LL = 4.56$, χ^2 (1, 11) = 9.13, $p < .01$, items analyzed separately; larger memory capacity improved RTs for both types of words. Nonverbal IQ had no effect on accuracy or RT in the recognition task.

For accuracy on the production task, adding phonological memory to the model had a significant effect on the linear term, $\Delta LL = 4.95$, χ^2 (1, 12) = 10.07, $p < .01$, indicating that higher memory capacity improved word learning rate. This effect of memory was found for both wordlike, $\Delta LL = 6.05$, χ^2 (1, 12) = 12.09, $p < .001$, and unwordlike, $\Delta LL = 1.90$, χ^2 (1, 12) = 3.81, $p = .05$, items analyzed separately, suggesting that the effect of memory was not affected by the words' similarity to existing lexical forms. English vocabulary size, however, had a significant effect on the intercept term, $\Delta LL = 2.96$, χ^2 (1, 11) = 5.91, $p < .05$, suggesting that a large English vocabulary increased the number of words a participant was able to learn. However, this effect was observed only for wordlike items, $\Delta LL = 2.40$, χ^2 (1, 11) = 4.79, $p < .05$, and did not hold for the unwordlike items analyzed separately, $\Delta LL = 1.59$, χ^2 (1, 11) = 3.18, $p > .1$, indicating that

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

larger English vocabulary size conferred a specific advantage to learning the novel words that resembled English.

Reaction times in the production task were affected by phonological memory in a similar manner to the recognition task; adding phonological memory to the intercept term significantly improved model fit, $\Delta LL = 20.9$, χ^2 (1, 11) = 40.23, $p < .0001$, an effect observed for both wordlike, 17.148, χ^2 (1, 11) = 34.30, $p < .0001$, and unwordlike, $\Delta LL = 9.34$, χ^2 (1, 11) = 18.70, $p < .0001$, items. These results suggest that learners with larger memory capacity more quickly retrieved novel words' forms, but training improved RTs for all learners at a similar rate. Nonverbal IQ had no effect on accuracy or RT in the production task.

DISCUSSION

The current study investigated how orthographic typicality in the L1 and L2 affects the rate and amount of learning of written L2 words, and explored the relative contributions of memory, vocabulary, and nonverbal intelligence to vocabulary learning. As expected, accuracy and RT improved with training, both for recognizing novel words' matching pictures and for producing the novel words when cued with the meaning. Of critical importance, we found that high similarity to existing English words (based on both high neighborhood density and high segment/phoneme and bigram/biphone frequencies) improved learning of both novel words' forms and their meanings. After controlling for phonological similarity, orthographic wordlikeness significantly improved word production, but was not sufficient to affect word recognition alone. Learners also successfully extracted and utilized regularities in the novel language, leading to a characteristic pattern of overgeneralization errors. In addition, higher phonological memory capacity was associated with better form and meaning learning regardless of English similarity. High vocabulary size also provided a general benefit to meaning acquisition (picture–word correspondences), but for word form acquisition, it selectively improved items with high similarity to English. Results of the current study support the view that word similarity can assist L2 vocabulary learning by scaffolding on overlapping structures in the native language.

After 1 hr of study involving repeated retrieval attempts with feedback, learners were able to produce about 35 of the 48 words (73%) correctly, given only a picture as a cue to the novel word's meaning. These results highlight the effectiveness of repeated testing as an instructional tool (Roediger & Karpicke, 2006) given the rapid gains in performance. The persistence of these single-session learning benefits, however, depends on successful transition from their representation in initial episodic memory traces to stable lexical representations, as described in the complementary systems account of word learning (Davis & Gaskell, 2009; Dumay & Gaskell, 2007).

Within this period of initial learning, we have shown that form similarity in the absence of meaning overlap can accelerate learning. Words that overlap in both form and meaning across languages (i.e., cognates) are valuable teaching tools and are easier to learn than noncognates (De Groot & Keijzer, 2000; Lotto &

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

De Groot, 1998; Rogers et al., 2014). In the current study, participants mastered the picture–word associations for wordlike items faster than unwordlike items, and successfully produced more wordlike items over the course of training. In addition, RTs for both word recognition and production were faster overall for wordlike than for unwordlike items. These results corroborate existing studies of vocabulary acquisition using nonwords in the auditory domain showing that phonological neighborhood size and phonotactic probability influence learning outcomes (Frisch et al., 2000; Luce & Large, 2001; Roodenrys & Hinton, 2002; Storkel et al., 2006; Thorn & Frankish, 2005).

Orthographic similarity, as manipulated in the current study, can affect phonological processing through lexical and sublexical structures (Ziegler, Muneax, & Grainger, 2003), making orthography an important component for understanding vocabulary learning. Because of the close link between orthography and phonology, we conducted a secondary analysis on a subset of items matched for phonological wordlikeness (phonological neighborhood size, phoneme frequency, and biphone frequency) in order to isolate the effect of orthographic similarity alone. In the production task, orthographic wordlikeness increased accuracy and decreased RT (although to a lesser extent than the full analysis), but no effect of orthographic wordlikeness alone was observed for word recognition. These results suggest the contribution of both a general wordlike effect on recognition and production as well as independent effects of orthography and phonology on word production. The additive effects of orthographic and phonological wordlikeness may occur through a dual coding advantage due to participants generating phonological forms of the novel written words. The contribution of phonological knowledge is particularly likely given the observation that participants' phonological memory performance affected learning performance for both word recognition and production. To precisely account for orthographic and phonological effects during learning would require assessing acquisition of novel words fully crossed for orthographic and phonological wordlikeness.

The effects of wordlikeness on different components of novel word learning were identified in the recognition and production tasks. The recognition task probed participants' ability to forge a link between a novel word and an existing semantic concept, whereas the production task assessed memory for the novel written form itself. To emulate an SLA context, the novel written words were paired with pictures of familiar objects (e.g., airplane or cat). In the recognition task, the participants' goal was to identify the matching picture from four alternatives when presented with one of the novel words, which served to minimize memory demands for word forms and isolate knowledge of the word–meaning link itself. Accuracy for wordlike items was slightly lower than for unwordlike items in the first testing block, but wordlike accuracy increased at a faster rate. This pattern was only observed for the novel items that had low similarity, yet still contained entirely legal English bigrams; words that contained illegal bigrams (e.g., the FP in *gofp*) were learned worse overall. This pattern conforms to models of word learning, whereby unwordlike items trigger learning better, because their novelty is more salient than wordlike items (Gupta & MacWhinney, 1997; Storkel et al., 2006). Wordlike items are advantaged with additional training, because they are easier to associate with existing vocabulary. In postexperiment debriefings, participants commonly

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

reported that they learned the novel words by creating visual associations (e.g., "The chained dog moaned and howled" to remember that the novel word *mowl* means *chain*), similar to the successful keyword learning method (Shapiro & Waters, 2005). It is possible that participants found it easier to generate useful, robust associations for wordlike items, accelerating learning of these words' meanings. Providing learners with sample keyword associations could be more beneficial for unwordlike items, for which strong associations are more difficult to self-generate due to their further distance from existing vocabulary. Whether or not this effect of wordlikeness generalizes to word classes beyond the highly imageable nouns used in the current study is yet to be determined. Keyword method learning techniques are generally less successful when used with low imageability concepts, like *liberty* (Shapiro & Waters, 2005), and it is possible that wordlikeness affects learning differently for words relating to abstract concepts. While keyword learning is an effective tool for the initial stages of vocabulary learning, full lexical consolidation requires abstraction of a word's form and meaning from specific episodic memories (Davis & Gaskell, 2009), at which point the role of the keyword in recall will be diminished.

In contrast to the recognition task's assessment of semantic learning, the production task directly assessed memory for the novel written word forms. Although production and recognition accuracy were highly correlated, production accuracy lagged substantially behind recognition accuracy. Even when participants had acquired word–meaning links, they continued to experience difficulty retrieving the exact form. Production was much more sensitive to the effect of wordlikeness than recognition, with an effect equivalent to roughly a four-word difference between conditions. This large effect of wordlikeness supports the idea that as newly learned words begin to decay in short-term memory, they can be reconstructed based on a combination of lexical similarity to known words (i.e., neighbors) or sublexical sequence typicality (i.e., segment and bigram frequencies).

While similarity to English had a notable effect on accuracy, there were also indications that letter frequencies within the novel language itself influenced word production. There is a strong bias for spontaneous errors during speech production to adhere to learned phonotactic rules, and the types of errors one produces is susceptible to short-term training (Dell, Reed, Adams, & Meyer, 2000). In the current study, the novel language's orthotactic structure diverged in several ways from English, with certain letters occurring more or less frequently at different positions within a word. In correct responses, it is not possible to separate correct retrievals of an individual word versus successful gap-filling utilizing letter patterns. However, overgeneralization of these patterns within the novel language leads to a characteristic pattern of errors. Positional segment frequency in the novel language was a better predictor than English positional segment frequency for the types of single-letter errors that participants made in their responses, indicating that learners had already begun to extract statistical regularities in the vocabulary of the novel language, and used this general language knowledge to supplement their memory for individual words.

Successful use of a L2 not only requires a large vocabulary but also depends on fluency. Fluency is especially important for L2 communication in time-sensitive

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

contexts, and slow responses impede L2 utility in interpersonal and vocational settings. Participants in the current study correctly produced wordlike items 10%–15% faster than unwordlike items throughout training, equivalent to roughly half a second. Response fluency is a skill that can be effectively trained in instructional settings (Snellings & van Gelderen, 2002; van Gelderen, Oostdam, & van Schooten, 2011), and these results suggest that unwordlike items may be more effective targets for fluency training during L2 instruction, as RTs continue to lag behind wordlike items. However, overall RTs were relatively slow compared to the rate needed for fully fluent speech. The nature of the task, involving focused, effortful recall with no time penalty, may not directly translate to advantages in natural speech production.

While the characteristics of the novel words (i.e., neighborhood size and segment/bigram frequencies) had clear and dramatic effects on learning, there were also effects of the learners' cognitive and linguistic backgrounds. Larger phonological memory capacity improved recognition and production RT for both types of words and the rate of learning for the novel word forms. Larger English vocabulary size improved recognition learning rate and interacted with wordlikeness on word production. Nonverbal intelligence, however, was not associated with any individual differences in learning accuracy or RT.

Phonological memory capacity was assessed using the digit span and nonword repetition subtests of the Comprehensive Test of Phonological Processing, which score accuracy for oral repetition of auditorily presented number or nonword sequences. Higher scores indicate increased short-term memory for an ordered list, and participants with high memory capacity may have been able to maintain the novel words' picture associations and written forms in working memory long enough to encode a robust, easily retrieved memory.

In terms of accuracy, English vocabulary size was associated with learning rate for picture–word associations in the recognition task. As discussed earlier, a commonly reported learning strategy in this task was to build an arbitrary mental association linking the novel word and the picture. Learners with larger vocabularies may have been able to generate more distinctive associations, because they had more words of lower frequency available to form links. Alternatively, because a large vocabulary is evidence of robust word learning skills in the L1, unmeasured predictors of learning skills such as motivation, attention, or inhibition could have affected both English vocabulary size and novel word recognition performance. In addition, individual differences in vocabulary size may change the relative frequencies of different letter sequences. Individuals' vocabularies vary widely at lower frequencies (Keuleers, Stevens, Mandera, & Brysbaert, 2015), and the effects of these unique differences on language learning is a promising area of future research.

Memory for word forms on the production task was improved by both phonological memory capacity and English vocabulary size, although the two had distinct effects. Larger phonological memory capacity resulted in a faster learning rate for both sets of words, but did not affect the overall number of words learned. Each production test was separated by an intervening word recognition task, and thus learned word forms had to be stored for several minutes before

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

being tested again (average time between the end of one production test and the start of the next was 4.22 min, $SD = 0.73$). Higher phonological memory capacity predictably resulted in larger gains from block to block, by increasing the amount of new information learned each time. Although there was no auditory component to the task used in the current study, people typically automatically generate phonological forms for written words (Johnston et al., 2004). These self-generated phonological forms provide a mechanism for the observed influence of phonological memory capacity on written word learning. When English phonological similarity was controlled for, the effect of wordlikeness on learning was reduced, suggesting that phonological forms played an important role during learning.

In contrast, larger vocabulary size improved the total number of words learned instead of the rate, but only for the wordlike item set. This effect of vocabulary on wordlike items reinforces the idea that partially remembered items may be reconstructed based on existing language knowledge, and that this strategy is more effective for items with higher similarity to the learner's native language. Note that while vocabulary size improved wordlike learning, it did not negatively impact unwordlike learning. That is, the benefit for wordlike items did not come at the expense of memory for unwordlike items. When novel words differ from existing language patterns, they can still be learned through other available mechanisms.

The final individual difference assessed in the current study, nonverbal intelligence, failed to predict changes in accuracy or RT in either task. IQ has previously been shown to affect aspects of SLA that require decision making or pattern recognition, such as listening comprehension (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012) or grammar learning (Kempe et al., 2010; Kempe & Brooks, 2011). In contrast, the current study's vocabulary learning task relies heavily on memory and prior vocabulary knowledge, but does not tap into the skills that contribute to nonverbal IQ.

In conclusion, our results show that acquisition of novel written words is affected by their lexical characteristics, as well as by the cognitive and linguistic profile of the learner. Wordlikeness, which involved a manipulation of orthographic neighborhood size, segment frequency, and bigram frequency, had a pronounced effect on both the formation of word–meaning links and the acquisition of novel word forms. In addition, regularities within the novel language were accessible to learners and influenced their responses, as evident in overgeneralization patterns leading to response errors. Both phonological memory capacity and native-language vocabulary size predicted learning, but while vocabulary effects were more extensive for items with wordlike forms, memory capacity benefited learning for words of all types. Vocabulary learning is an especially important part of SLA, because it provides a salient benchmark to the learner of their progress, helping to increase confidence and motivation. The vocabulary hurdle is highest at the onset of learning, and improving performance at this stage is vital to afford the learner a framework on which to build further easy and difficult words alike. These results suggest that learners may accelerate vocabulary acquisition at early stages by capitalizing on words that resemble their native language, allowing them to rapidly build their L2 framework.

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

APPENDIX A

Table A.1. *Stimuli list*

		Orthographic Metrics			Phonological Metrics		
Nonword	Meaning	*N*	Segment Frequency	Bigram Frequency	*N*	Phoneme Frequency	Biphone Frequency
			Wordlike Condition				
baft	whistle	6	0.318	0.016	8	0.249	0.012
bamn	horse	4	0.319	0.021	34	0.158	0.005
birt	vacuum	4	0.352	0.022	11	0.106	0.002
comf	desk	5	0.307	0.053	35.5	0.155	0.005
copt	mushroom	9	0.373	0.044	25.2	0.141	0.007
damb	envelope	8	0.27	0.018	15	0.256	0.036
darg	bicycle	4	0.339	0.036	24	0.281	0.02
duch	glasses	5	0.208	0.012	27	0.139	0.004
farp	key	8	0.333	0.038	13.5	0.297	0.028
gach	bat	4	0.254	0.019	15	0.142	0.009
gonk	bottle	6	0.3	0.033	5.7	0.223	0.033
lamk	butterfly	5	0.258	0.016	4	0.176	0.004
mowl	chain	5	0.271	0.017	36	0.206	0.016
nish	iron	4	0.21	0.024	3.3	0.186	0.008
nist	window	5	0.273	0.038	8.7	0.238	0.007
patz	bell	3	0.293	0.029	31	0.192	0.01
polp	tent	8	0.325	0.024	9	0.105	0.002
purd	fence	4	0.291	0.022	1	0.226	0.022
sumb	fork	6	0.259	0.023	10.7	0.197	0.011
sund	pear	9	0.325	0.034	28.5	0.143	0.002
vask	plug	6	0.261	0.016	6.3	0.232	0.008
veck	wheel	6	0.242	0.027	19.2	0.219	0.014
vond	hanger	5	0.292	0.038	8	0.217	0.015
zaft	belt	4	0.255	0.003	11	0.102	0.003
			Unwordlike Condition				
cisv	tree	0	0.28	0.019	19.7	0.091	0.001
duvc	leaf	0	0.202	0.003	1.5	0.187	0.014
ferz	glove	0	0.295	0.02	12	0.117	0.002
gefn	flower	0	0.259	0.008	5	0.273	0.019
gehz	lemon	0	0.193	0.005	20	0.057	0.001
giwm	cup	0	0.177	0.002	3.8	0.224	0.007
gofp	airplane	0	0.24	0.006	15	0.123	0.004
kenf	goat	0	0.258	0.018	6.2	0.164	0.007
kowm	ruler	0	0.202	0.006	13.5	0.111	0.003
kuft	pig	0	0.191	0.004	3	0.224	0.012
kutv	onion	0	0.165	0.008	14.8	0.113	0.005
lebh	chimney	0	0.224	0.008	12.7	0.186	0.008
lerl	corn	0	0.336	0.024	19.2	0.261	0.015
nahf	umbrella	0	0.192	0.004	3.5	0.164	0.004

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table A.1 (*cont.*)

		Orthographic Metrics			Phonological Metrics		
Nonword	Meaning	*N*	Segment Frequency	Bigram Frequency	*N*	Phoneme Frequency	Biphone Frequency
nomg	thumb	0	0.243	0.019	7.7	0.14	0.006
piwv	knife	0	0.201	0.006	6.3	0.122	0.003
retz	button	0	0.27	0.039	14.3	0.161	0.003
vafh	basket	0	0.204	0.005	11.7	0.234	0.009
vorz	spoon	0	0.267	0.021	5.2	0.13	0.006
wehv	dress	0	0.2	0.006	11	0.149	0.004
wusb	chair	0	0.191	0.008	8.5	0.114	0.004
zagt	lock	0	0.265	0.005	4.7	0.155	0.007
zimd	kite	0	0.187	0.005	2.8	0.17	0.006
zuwg	fan	0	0.121	0.001	4.8	0.058	0.001

Note: N, Neighborhood density.

Table A.2. *Errors at the first letter position (first consonant)*

Letter	Error Frequency	Colbertian Frequency	English Frequency
z	3.83	8.33	0.37
g	3.48	12.5	4.19
c	2.33	6.25	6.53
v	2.29	10.42	1.19
d	2.23	8.33	6.16
w	2	4.17	5.53
b	1.81	6.25	7.39
n	1.67	8.33	2.62
p	1.67	8.33	5.99
f	1.65	4.17	5.76
l	1.54	6.25	6.45
r	1.44	2.08	4.8
s	1.33	4.17	9.52
k	1.23	8.33	2.29
m	0.85	2.08	5.77
t	0.83	0	6.99
h	0.69	0	6.11
a	0.33	0	2.89
e	0.33	0	2.24
o	0.19	0	1.96
q	0.15	0	0.18
i	0.1	0	0.78
u	0.08	0	0.61
j	0.06	0	2.69
y	0.04	0	1
x	0	0	0

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table A.3. *Errors in the second letter position (vowel)*

Letter	Error Frequency	Colbertian Frequency	English Frequency
e	6.21	18.75	14.42
o	4.48	20.83	19.91
a	4.04	27.08	20.97
i	3.46	14.58	13.84
u	3.35	18.75	9.69
l	0.83	0	3.65
r	0.75	0	4.46
h	0.69	0	4.2
w	0.42	0	0.76
s	0.38	0	0.34
f	0.31	0	0.04
n	0.23	0	2
k	0.21	0	0.39
g	0.19	0	0.29
p	0.19	0	0.7
v	0.15	0	0.59
d	0.13	0	0.58
t	0.13	0	1.01
m	0.1	0	0.4
b	0.08	0	0.23
z	0.06	0	0.04
c	0.04	0	0.58
j	0	0	0
q	0	0	0
x	0	0	0
y	0	0	0

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table A.4. *Errors in the third letter position (second consonant)*

Letter	Error Frequency	Colbertian Frequency	English Frequency
r	6.04	14.58	9.2
n	4.04	8.33	9.97
m	3.56	14.58	3.93
s	3.44	10.42	7.01
h	3.23	6.25	0.21
f	3.15	12.5	1.42
t	2.02	6.25	5.17
l	1.46	2.08	8.98
g	1.44	2.08	1.9
c	1.04	6.25	4.87
w	1.04	10.42	1.22
e	0.85	0	9.04
d	0.83	0	2.65
p	0.81	2.08	1.98
v	0.77	2.08	1.88
i	0.65	0	5.89
a	0.6	0	9.83
o	0.6	0	6.62
k	0.56	0	2.08
u	0.56	0	2.89
b	0.31	2.08	1.26
z	0.19	0	0.53
j	0.08	0	0.1
x	0.08	0	0.37
y	0.04	0	1.01
q	0.02	0	0

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Table A.5. *Errors in the fourth letter position (third consonant)*

Letter	Error Frequency	Colbertian Frequency	English Frequency
t	6.48	14.58	10.41
k	4.15	8.33	7.01
p	3.88	6.25	3.84
f	3.65	6.25	1.3
b	3.31	6.25	1.21
d	2.92	8.33	7.07
z	2.38	10.42	0.4
g	2.02	6.25	2.79
h	1.92	10.42	3.44
n	1.85	4.17	6.29
l	1.6	4.17	6.52
v	1.58	8.33	0.01
m	1.4	4.17	3.11
s	0.83	0	11.52
r	0.65	0	3.43
w	0.4	0	1.92
c	0.38	2.08	0.36
e	0.31	0	18.1
i	0.27	0	0.53
a	0.17	0	2.77
o	0.1	0	2.05
j	0.08	0	0
u	0.08	0	0.4
y	0.06	0	5.29
q	0.02	0	0.01
x	0.02	0	0.24

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

ACKNOWLEGMENTS

This research was supported in part by NICHD Grants R01 HD059858, NIH R01 DC008333, and T32 NS 47987–8. The authors 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., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning, 62*, 49–78. doi:10.1111/j.1467-9922.2012.00706.x

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

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

Bialystok, E. (2008). Bilingualism: The good, the bad, and the indifferent. *Bilingualism: Language and Cognition, 12*, 3–11. doi:10.1017/S1366728908003477

Bigham, D. S. (2010). Mechanisms of accommodation among emerging adults in a university setting. *Journal of English Linguistics, 38*, 193–210. doi:10.1177/0075424210373542

Birdsong, D. (2009). Age and the end state of second language acquisition. In W. Ritchie & T. Bhatia (Eds.), *New handbook of second language acquisition* (pp. 401–423). Amsterdam: Elsevier.

Birdsong, D. (2014). The critical period hypothesis for second language acquisition: Tailoring the coat of many colors. In M. Pawlak & L. Aronin (Eds.), *Essential topics in applied linguistics and multilingualism, second language learning and teaching* (pp. 43–50). Geneva: Springer International.

Brooks, P. J., Kempe, V., & Donachie, A. (2011). Second language learning benefits from similarity in word endings: Evidence from Russian. *Language Learning*, *61*, 1142–1172. doi:10.1111/j.1467-9922.2011.00665.x

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364*, 3773–3800. doi:10.1098/rstb.2009.0111

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*, 1–56. doi:10.1111/0023-8333.00110

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

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. *Journal of Experimental Psychology: Learning Memory and Cognition, 26*, 1355–1367. doi:10.1037/0278-7393.26.6.1355

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*, 35–39. doi:10.1111/j.1467-9280.2007.01845.x

Dunn, L. M. (1997). *Examiner's manual for the PPVT-III: Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Evans, B. G., & Iverson, P. (2007). Plasticity in vowel perception and production: A study of accent change in young adults. *Journal of the Acoustical Society of America, 121*, 3814–3826. doi:10.1121/1.2722209

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Franceschina, F. (2005). *Fossilized second language grammars: The acquisition of grammatical gender*. Amesterdam: Benjamins.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences, 99*, 529–534. doi:10.1073/pnas.012611199

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*, 481–496. doi:10.1006/jmla.1999.2692

Gass, S. M., Behney, J. N., & Uzum, B. (2013). Inhibitory control, working memory and L2 interaction. In K. Drozdzial-Szelest & M. Pawlak (Eds.), *Psycholinguistic and sociolinguistic perspectives on second language learning and teaching* (pp. 91–114). Berlin: Springer International.

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2012). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning, 21*, 70–105. doi:10.1080/09588221.2012.700315

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59*, 267–333. doi:10.1006/brln.1997.1819

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. doi:10.1016/0010-0285(89)90003-0

Johnston, M., McKague, M., & Pratt, C. (2004). Evidence for an automatic orthographic code in the processing of visually novel word forms. *Language and Cognitive Processes, 19*, 273–317. doi:10.1080/01690960344000189

Kempe, V., & Brooks, P. J. (2011). Individual differences in adult second language learning: A cognitive perspective. *Scottish Languages Review*, *23*, 15–22. Retrieved from https://repository.abertay.ac.uk/jspui/handle/10373/1045

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

Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology, 68*, 1665–1692. doi:10.1080/17470218.2015.1022560

Liu, M., Evans, M. K., Horwitz, E., Lee, S., McCrory, M., Park, J.-B., et al. (2013). A study of the use of social network sites for language learning by university ESL students. In M.-N. Lamy & K. Zourou (Eds.), *Social networking for language education* (pp. 137–157). London: Palgrave Macmillan.

Lotto, L., & De Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*, 31–69. doi:10.1111/1467-9922.00032

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

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*, 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*, 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*, 379–413. doi:10.1017/S0272263112000125

Merkle, E. C., You, D., & Preacher, K. J. (2015). Testing nonnested structural equation models. *Psychological Methods*. Advance online publication. doi:10.1037/met0000038

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language, 59*, 475–494. doi:10.1016/j.jml.2007.11.006

Mirman, D., Magnuson, J. S., Graf Estes, K., & Dixon, J. A. (2008). The link between statistical segmentation and word learning in adults. *Cognition, 108*, 271–280. doi:10.1016/j.cognition.2008.02.003

Munro, M. J., Derwing, T. M., & Flege, J. E. (1999). Canadians in Alabama: A perceptual study of dialect acquisition in adults. *Journal of Phonetics, 27*, 385–403. doi:10.1006/jpho.1999.0101

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance: Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*, 127–162. doi:10.1016/s0010-0285(03)00128-2

PsychCorp. (1999). *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Harcourt Assessment.

Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in Cognitive Science, 6*, 5–42. doi:10.1111/tops.12078

Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon: Multilingual Matters.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249–255. doi:10.1111/j.1467-9280.2006.01693.x

Rogers, J., Webb, S., & Nakata, T. (2014). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research, 19*, 9–27. doi:10.1177/1362168814541752

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 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*, 217–236. doi:10.1068/p5117

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27–52. doi:10.1016/S0010-0277(98)00075-4

Schepens, J., Dijkstra, T., Grootjen, F., & van Heuven, W. J. B. (2013). Cross-language distributions of high frequency and phonetically similar cognates. *PLOS ONE, 8*, e63006. doi:10.1371/journal.pone.0063006

Schwartz, R. G., & Leonard, L. B. (1982). Do children pick and choose? An examination of phonological selection and avoidance in early lexical acquisition. *Journal of Child Language, 9*, 319–336. doi:10.1017/S0305000900004748

Sebastián-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., & Díaz, B. (2006). First- and second-language phonological representations in the mental lexicon. *Journal of Cognitive Neuroscience, 18*, 1277–1291. doi:10.1162/jocn.2006.18.8.1277

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*, 129–146. doi:10.1191/1362168805lr151oa

Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics, 43*, 269–285.

Snellings, P., & van Gelderen, A. (2002). Lexical retrieval: An aspect of fluent second language production that can be enhanced. *Language Learning, 52*, 723–754. doi:10.1111/1467-9922.00202

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*, 490–502. doi:10.1017/S1366728911000216

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press

Storkel, H. L., 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*, 729–735. doi:10.1037/0278-7393.31.4.729

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition, 27*, 173–204. doi:10.1017/S0272263105050102

van Gelderen, A., Oostdam, R., & van Schooten, E. (2011). Does foreign language writing benefit from increased lexical fluency? Evidence from a classroom experiment. *Language Learning, 61*, 281–321. doi:10.1111/j.1467-9922.2010.00612.x

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology and Aging, 18*, 332–339. doi:10.1037/0882-7974.18.2.332

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

Vitevitch, M. S., & Luce, P. A. (2004). A Web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavior Research Methods, Instruments, and Computers, 36*, 481–487. doi:10.1016/j.bbi.2008.05.010

Vuong, Q. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica, 57*, 307–333. doi:10.2307/1912557

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

Yoshida, H., Tran, D. N., Benitez, V., & Kuwabara, M. (2011). Inhibition and adjective learning in bilingual and monolingual children. *Frontiers in Psychology, 2*, 210. doi:10.3389/fpsyg.2011.00210

Ziegler, J. C., Muneaux, M., & Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language, 48*, 779–793. doi:10.1016/S0749-596X(03)00006-8

https://doi.org/10.1017/S0142716416000242 Published online by Cambridge University Press