LANGUAGE LEARNING
A Journal of Research in Language Studies

Language Learning ISSN 0023-8333

# EMPIRICAL STUDY

# Bilinguals' Existing Languages Benefit Vocabulary Learning in a Third Language

James Bartolotti and Viorica Marian

Northwestern University

Learning a new language involves substantial vocabulary acquisition. Learners can accelerate this process by relying on words with native-language overlap, such as cognates. For bilingual third language learners, it is necessary to determine how their two existing languages interact during novel language learning. A scaffolding account predicts transfer from either language for individual words, whereas an accumulation account predicts cumulative transfer from both languages. To compare these accounts, 20 English-German bilingual adults were taught an artificial language containing 48 novel written words that varied orthogonally in English and German wordlikeness (neighborhood size and orthotactic probability). Wordlikeness in each language improved word production accuracy, and similarity to one language provided the same benefit as dual-language overlap. In addition, bilinguals' memory for novel words was affected by the statistical distributions of letters in the novel language. Results indicate that bilinguals utilize both languages during third language acquisition, supporting a scaffolding learning model.

**Keywords** bilingualism; vocabulary learning; third language acquisition; neighborhood size; orthotactic probability

---

This research was supported in part by grants NICHD R01 HD059858 and T32 NS 47987–8. We would like to thank the members of the Northwestern University Bilingualism and Psycholinguistics Research Group for comments on this work.

This article has been awarded an Open Materials badge. All materials are publicly accessible in the IRIS digital repository at http://www.iris-database.org. Learn more about the Open Practices badges from the Center for Open Science: https://osf.io/tvyxz/wiki.

Correspondence concerning this article should be addressed to James Bartolotti, Department of Communication Sciences and Disorders, Northwestern University, 2240 Campus Drive, Evanston, IL 60208. E-mail: j-bartolotti@u.northwestern.edu



DOI: 10.1111/lang.12200

## Introduction

Knowledge of multiple languages is a desirable skill. Eighty percent of Americans believe that children should learn a second language (L2) fluently before graduating high school (Rivers, Robinson, Harwood, & Brecht, 2013). Similarly, 84% of Europeans believe that everyone in the European Union should be able to speak an L2 fluently (European Commission Special Barometer, 2006). In actuality, multilingual rates in both the United States and the European Union are far lower than desired levels. Estimates from U.S. Census data in 2007 yield bilingualism rates of slightly more than 20%, and only 26% of respondents to a 2001 Gallup poll described themselves as able to hold a conversation in an L2 (Gallup Organization, 2001). Even in the European Union, where primary and secondary school instruction in two foreign languages is widespread (Barcelona European Council, 2002), only 56% of adults report fluency in an L2 (European Commission Special Barometer, 2006).

The primary factor working against L2 acquisition in adults is the knowledge base that learners must acquire. Even setting aside the challenges involved in acquiring a new language's orthographic/phonological system and grammatical rules, the raw number of vocabulary items that must be learned can seem insurmountable. It is estimated that understanding written texts fluently requires coverage of 98% of words in the text, at which point only 1 word in 50 is unknown to the reader (Hu & Nation, 2000; Nation, 2006; Schmitt, Jiang, & Grabe, 2011). Achieving 98% coverage for learners of English requires knowing 8,000–9,000 base words as well as their inflected forms—a total of approximately 34,600 individual words (Nation, 2006). As a result, it can take years to build up the necessary vocabulary in a new language. Even high school and university students who have studied L2 English for several years are estimated to know only between 1,000–4,000 base words, leaving significant vocabulary gaps at lower word frequencies (Laufer, 2000). Four thousand base words corresponds to roughly 95% text coverage (one unknown word in 20), producing significant gaps in reading comprehension (Schmitt et al., 2011). Because of the difficulty L2 vocabulary acquisition presents, there is high variability in adult language learning (Birdsong, 2009). Even in cases of successful acquisition, proficiency rarely approaches nativelike levels in all areas (Baker & Trofimovich, 2005; DeKeyser, 2005; Johnson & Newport, 1989; MacWhinney, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, & Díaz, 2006). There is a clear need to identify the factors that contribute to language learning difficulties in children and adults, as well as to understand how to optimize instruction for each learner to maximize his or her success.

**Prior Language Knowledge and Novel Word Learning**

The unique challenge that adults face while learning a novel language is that, unlike children learning their mother tongue, adults already know one or more languages, and this experience has thoroughly shaped how they process language. In particular, first language (L1) experience tends to sharpen the mind to features and regularities of future L1 input, such as word forms (Ellis, 2006; Schmitt, 2008). A consequence of this linguistic sharpening is that adults become particularly attuned to learning new vocabulary in their native language with increased experience; as age increases, vocabulary size becomes a better predictor of word learning ability than other factors like working memory skill (Long & Shaw, 2000). Throughout life, native English speakers learn low-frequency words like *allay* or *obviate*, as well as domain-specific vocabulary: A novice knitter quickly learns words like *purl*, *skein*, or *selvage*, and professional psycholinguists learn words like *morpheme*, *electroencephalogram*, or *aphasia*. Acquisition of these words comes relatively seamlessly, even to people who may have struggled to learn a foreign language, by consequence of their similarity to existing words and lexical patterns. Knowing a word's form is only one required component of word knowledge—learners must also understand the word's meaning, morphological inflections, and grammatical usage, but mastery of form provides a base to link a novel word's appearance in different contexts.

While linguistic sharpening can facilitate acquisition of familiar words, it can also interfere with learning of words that do not match native language patterns. This has important implications for novel language learning, because lexical similarity drops across languages. Letters and sounds can be combined in a myriad of ways, but each language allows different 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 according to calculations based on CLEARPOND (Marian, Bartolotti, Chabal, & Shook, 2012) and $\text{SUBTLEX}_{\text{us}}$ (Brysbaert & New, 2009). These four-letter words have an average of 10.33 neighbors (i.e., words that are the same in all but one letter), and while certain sequences of letters appear quite frequently (e.g., CE, LY) others do not appear at all (e.g., ZW, FD). This high regularity within a language is lost when compared between languages—written English words tend to have 5–7 times fewer neighbors in related languages like Dutch, French, German, or Spanish than they do English neighbors (Marian et al., 2012). Thus, when attempting to learn a new language, native language knowledge is very likely to interfere with most novel language vocabulary as a result of the languages' divergent structures.

However, there will always be pockets of consistency between any two languages that the learner can utilize. Learners and instructors have long known the benefit of cognates, which overlap in both form and meaning across languages. Cognates are relatively easy to acquire (De Groot & Keijzer, 2000; Lotto & De Groot, 1998) and can give the learner a head start in vocabulary acquisition. As a result, cognates are an invaluable teaching aid, but they are limited in that they constitute a small set of vocabulary. Fortunately, other kinds of overlap can also provide a benefit, with the potential to greatly expand the pool of easily learnable words. Native Spanish adults learning L2 English responded positively to use of the native language in L2 instruction, particularly as a way to recognize overlap between the two languages in vocabulary and grammatical structures (Brooks-Lewis, 2009). In addition, learners in laboratory contexts acquire novel words that do adhere to native language lexical patterns better than words that do not (Bartolotti & Marian, 2014, in press; Storkel, Armbrüster, & Hogan, 2006; Thorn & Frankish, 2005). This native language overlap can be characterized in several different ways; two useful metrics are *neighborhood size* and *orthotactic probability*.

Neighborhood size refers to the number of known words that differ from the target in only one letter. A broad definition of neighborhood size includes all words that differ from a target in the substitution, addition, or deletion of a single letter; for example, the English word *plant* has neighbors *plan*, *plank*, and *planet* (Marian et al., 2012). The concept of neighborhood size can also be applied to nonwords: *Baft* has English neighbors *raft* and *bat*. People judge nonwords with denser neighborhoods to be more subjectively wordlike (Bailey & Hahn, 2001), and they are demonstrably easier to learn (Roodenrys & Hinton, 2002; Storkel et al., 2006; Thorn & Frankish, 2005) than nonwords with sparse neighborhoods.

Orthotactic probabilities calculate how often individual letters or letter sequences are used in a language. Two complementary ways to characterize probabilities are positional segment probability (how often a letter occurs in a given position in a word) and positional bigram probability (how often two letters in sequence occur in a given position in a word) (Jusczyk, Luce, & Charles-Luce, 1994; Vitevitch & Luce, 2004). These metrics can similarly be applied to nonwords: The bigrams *su*, *um*, and *mb* in the nonword *sumb* have an average probability of .0076, four times higher than the average probability of bigrams in the nonword *kowm* (.0019). Most learning research to date has focused on phonological similarity and has shown that phonotactic probabilities contribute to nonword repetition accuracy (Gupta & Tisdale, 2009), subjective wordlikeness ratings (Bailey & Hahn, 2001), and word learning (Majerus, Poncelet, van der Linden, & Weekes, 2008; Storkel et al., 2006; Thorn & Frankish, 2005). Orthographic overlap also provides a benefit (Bartolotti & Marian, 2014, in press), with the additional advantage of having more potential overlap between common language pairs, as languages that share scripts often have more orthographic overlap than phonological overlap (Marian et al., 2012). However, whereas most languages share at least some phonological features, many language pairs do not share alphabets (e.g., English and Arabic) or writing systems (e.g., English and Mandarin), making orthography of limited use in some contexts.

**Two Models of Novel Word Learning**

Our discussion so far has shown that prior language experience shapes later language learning. Acquisition of vocabulary that resembles the native language is facilitated, whereas atypical vocabulary becomes more difficult to learn. By capitalizing on pockets of overlap between two languages where they exist, initial language learning in adults may be more effective. The patterns present in learned L2 vocabulary can facilitate further acquisition; beginning Spanish learners, for example, are better at learning words with high versus low Spanish neighborhood sizes (Stamer & Vitevitch, 2012). The mechanism by which this overlap can benefit learning at early stages constitutes the focus of the current investigation. L2 acquisition often begins with intentional vocabulary learning wherein the learner explicitly attempts to associate words with their meanings or translation equivalents (Barcroft, 2004).

In this study, we compared two possible models for how long-term knowledge affects novel vocabulary learning. The *scaffolding model* predicts that the ability to create a direct association between a newly encountered word and an existing word or concept drives memory strength. Novice learners rely heavily on L1 translations during L2 vocabulary learning (Liao, 2006; Schmitt, 1997), which anchors the relatively weak novel word to a strong existing memory. In the Revised Hierarchical Model of bilingual language processing, these word-to-word associations are strongest at the onset of L2 learning (Kroll & Stewart, 1994). Beyond translation equivalents, other lexical associations can be used to remember discrete aspects of a word. For example, the keyword learning method (Shapiro & Waters, 2005) is a pedagogical approach that emphasizes using a known word as a form intermediary between a novel word and its meaning (e.g., the word *steel* can be used to remember the phonological form of the French word *stylo*, meaning "pen"). It may be easier for learners to generate keywords for novel words that resemble the native language more closely, compared to novel words that have less nativelike forms. The scaffolding account thus emphasizes the learner's ability to directly utilize their existing linguistic framework during language acquisition.

In contrast, the *accumulation model* proposes that a novel word's consistency with lexicon-wide patterns affects the fidelity with which it is represented in short-term memory and retrieved from long-term memory. When a new word is first encountered, it is vulnerable to disruption, but rehearsal processes maintain the trace in the phonological loop until it can be stored in memory. How well information is represented and maintained in the phonological loop is affected by interactions with prior knowledge, as detailed in Baddeley's Working Memory Model (Baddeley, 1986). As applied to word learning, this long-term knowledge may be used to enhance the strength of the initial temporary storage during encoding (Gathercole & Martin, 1996; Gathercole, 2006; Gupta & MacWhinney, 1997). Novel words with more nativelike features are easier to repeat and maintain in working memory, compared to those with fewer nativelike features, because the sequencing of their letters or sounds is more predictable. In addition, newly learned words that are composed of high-probability patterns can benefit more from redintegration, the process of reconstructing a partially decayed short-term memory using long-term knowledge (Gathercole, Frankish, Pickering, & Peaker, 1999; Gupta & Tisdale, 2009; Schweickert, 1993). Thus, a word that was incompletely encoded may yet be accurately produced, reinforcing the target representation. To again use the French word *stylo* as an example, it contains common English sound sequences /sti/ and /lo/, plus the fourth-most common English onset bigram *st*, and the frequency of these features can facilitate accurate encoding and retrieval. These sublexical effects are distinct from the one-to-one whole-word associations that drive learning in the scaffolding model; this memory for a word as the sum of its parts is key to the accumulation account of vocabulary acquisition.

## The Current Study

In order to compare the scaffolding and accumulation models of word learning, we made use of a population with unique language experience by investigating acquisition of a third language (L3) in bilingual adults. Bilinguals necessarily divide their use and exposure between two languages. As a result, bilinguals know more unique words across both languages than monolinguals, both as children (Bialystok, Luk, Peets, & Yang, 2010) and as adults (Bialystok & Luk, 2012), but they use each language with lower frequency than a monolingual. Each of a bilingual's languages may share some neighbors and letter patterns with words in the target L3, but the way in which a bilingual's two languages compare and contrast with the target language is unique.

The manner in which the effects of both known languages combine leads to different predictions for L3 learning. If comparable learning benefits are observed both for L3 words that overlap with one language and L3 words that overlap with both of the two previously learned languages, this will provide support for the scaffolding account of word learning. By this account, word learning relies on a single form-related intermediary as an anchor to existing knowledge. For each novel word, quality anchors in either or both languages may be available, but one direct link is sufficient for learning benefits to be observed. In contrast, if target L3 words that overlap with both languages of a bilingual are learned better than those that overlap with a single language, this will provide support for the accumulation account of word learning. In the accumulation account, individual words benefit more from similarities that are common to both existing languages because of the features' higher overall frequency. The effect of wordlikeness on learning is thus additive across languages for each novel word.

The present study was designed to compare predictions of the scaffolding and accumulation models in the context of L3 orthographic vocabulary learning. We designed an artificial language containing novel written words that were paired with pictures of everyday objects. The use of pictures instead of translation equivalents allowed bilinguals to make use of both of their existing languages during learning. English-German bilinguals were taught the meaning and spelling of the novel words in a single experimental session. The novel words varied orthogonally in their similarity to existing English and German words, based on wordlikeness metrics comprising orthographic neighborhood size, orthotactic probability in each language, and judgments by an independent group of bilinguals. E+G+ words had high wordlikeness in both English and German (e.g., *nist* or *baft*), E+G– words had high English but low German wordlikeness (e.g., *sumb* or *gonk*), E–G+ words had low English but high German wordlikeness (e.g., *gach* or *kenf*), while E–G– words had low English and German wordlikeness (e.g., *gofp* or *kowm*). The scaffolding account predicts that bilinguals should learn the three wordlike types equivalently, and more than the unwordlike (E–G–) items. The accumulation account makes different predictions for the single-wordlike (E+G–, E–G+) and the double-wordlike (E+G+) items. The single-wordlike items should each be learned better than the E–G– baseline. The double-wordlike items should be learned better than either of the single-wordlike items, because they overlap to a greater degree across the bilingual's full lexicon. This critical comparison between double- and single-wordlike items should thus reveal whether effects of lexical similarity during novel language learning are additive across

**Table 1** Participant linguistic and cognitive backgrounds

| Measure | English | German |
|---|---|---|
| | | *M* (*SD*) [*Range*] |
| Age (years) | | 27.66 (6.19) [19.75–40.0] |
| Nonverbal intelligence[1] | | 110.70 (12.73) [86–128] |
| Phonological memory[2] | | 109.20 (8.94) [94–127] |
| | English | German |
| Speaking proficiency[3] | 9.42 (0.84) [8–10] | 8.10 (1.59) [5–10] |
| Listening proficiency[3] | 9.47 (0.77) [8–10] | 8.47 (1.35) [6–10] |
| Reading proficiency[3] | 9.53 (0.70) [8–10] | 8.16 (1.92) [3–10] |
| Composite proficiency[3] | 9.50 (0.72) [8–10] | 8.27 (1.52) [4.67–10] |
| Age of acquisition (years)[3] | 3.84 (5.04) [0–14] | 10.74 (7.69) [0–20] |
| Current usage (%)[3] | 75.47 (19.30) [40–100] | 16.18 (13.51) [0–50] |
| Vocabulary size[4] | 95.22 (4.99) [86.25–100] | 77.35 (14.58) [47.5–97.5] |

*Notes.* [1]Performance IQ standard score, Wechsler Abbreviated Scale of Intelligence (Psychological Corporation, 1999)
[2]Standard score, Comprehensive Test of Phonological Processing (Wagner et al., 1999)
[3]Language Experience and Proficiency Questionnaire (Marian et al., 2007), self-rated proficiency on a 1–10 scale
[4]LexTALE score on a 0–100 scale (Lemhöfer & Broersma, 2012).

multiple languages or whether learners associate novel words with a single language.

## Method

### Participants

Twenty English-German bilinguals participated for a small monetary compensation. Informed consent was obtained in accordance with the university's Institutional Review Board. After the experiment, participants completed the Language Experience and Proficiency Questionnaire (LEAP-Q, Marian, Blumenfeld, & Kaushanskaya, 2007) and the following three tests used to assess individual differences in word-learning performance: phonological memory (Comprehensive Test of Phonological Processing), consisting of digit span and nonword repetition subtests (Wagner, Torgesen, & Rashotte, 1999); English and German vocabulary size (LexTALE; Lemhöfer & Broersma, 2012); and nonverbal IQ (Wechsler Abbreviated Scale of Intelligence) consisting of block design and matrix reasoning subtests (Psychological Corporation, 1999). Descriptive statistics for each variable can be found in Table 1. The LexTALE assesses vocabulary knowledge based on lexical decision accuracy and is highly correlated with translation accuracy and other measures of vocabulary size. In the LEAP-Q, participants provided their speaking, listening, and reading proficiencies on a 1–10 scale, their age of acquisition, and their percentage of current usage for each language. Participants had been using English for an average of 23.9 years ($SD = 6.9$) and German for 18.2 years ($SD = 9.6$), and all bilinguals spent at least 1 year living in both an English-speaking and a German-speaking country. Participants learned English and German through a variety of mechanisms, primarily family exposure and school instruction. English was the first language for the majority of participants ($n = 14$), German was the first language for three participants, and three participants acquired English and German simultaneously. All but one participant were currently exposed to each language.

## Materials

Forty-eight orthographic words following the CVCC syllable structure were created in a novel language named Colbertian (named after the comic Stephen Colbert to engage participants in the task). The English and German wordlikeness of the novel words was calculated as composite scores of English and German orthographic neighborhood size and orthotactic probability (sum of grams and sum of bigrams), which were estimated using CLEARPOND (Marian et al., 2012), and English and German word similarity judgments obtained from English-German bilinguals ($N = 10$, ratings on 0–5 scales, 0 = *totally unacceptable as a [language] word*, 5 = *perfectly acceptable as a [language] word*). To illustrate, for the novel word *nist*, the values for neighborhood size (English: 5, German: 6), sum of grams (English: 0.273, German: 0.286), sum of bigrams (English: 0.038, German: 0.031), and similarity judgments (English: 4.75, German: 4.50) were all *z*-transformed and then averaged to yield composite scores for each language (English: 0.922, German: 0.728). This approach enables categorical classifications that incorporate both objective and subjective metrics. The target words were then categorized based on the wordlike composite scores: 14 words had both high English and high German wordlikeness (E+G+; e.g., *nist* or *baft*); 10 had high English but low German wordlikeness (E+G–; e.g., *sumb* or *gonk*); 11 had low English but high German wordlikeness (E–G+; e.g., *gach* or *kenf*); and the remaining 13 had low English and German wordlikeness (E–G–; e.g., *gofp* or *kowm*). A full list of stimuli is available in Appendix S1 in the Supporting Information online. 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 picture names did not overlap orthographically or phonologically in English or German with their paired Colbertian words. The names of pictures used in each of the four conditions did not differ in lexical frequency, orthographic or phonological neighborhood size, or gram, bigram, phoneme, or biphone probabilities in English or German, based on calculations using CLEARPOND (Marian et al., 2012). Stimulus presentation and data collection (both accuracy and response time) were controlled by the experimental software MATLAB with the Psychophysics Toolbox version 3.0.9 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). The experimental computer was a 27-inch Apple iMac (Mid 2011) with a 3.1 GHz quad-core Intel Core i5 processor and 4 GB of 1333 MHz DDR3 ram running OSX 10.7.5.

**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 seconds. Following the exposure block, participants performed five blocks of word recognition with feedback and five blocks of word production with feedback, alternating between the two tasks.

*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 (e.g., target word *nist* with target picture *window* and foil pictures *glove*, *tree*, and *goat*). The three foil pictures 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 computer recentered the mouse cursor to the center of the screen at the onset of each trial. Participants were instructed to click on the correct picture, and accuracy and response time (RT) were recorded. After making a response, the three foils disappeared, and the target picture and written word remained onscreen for 1 second, followed by a 1-second intertrial interval. Because the word–picture pair remained visible during feedback, participants could use this opportunity to relearn the correct association.

*Word Learning: Production*

In 48 production trials, a single random target picture (e.g., *window*) was presented in the center of the screen; participants were instructed to type the name of the picture in Colbertian and their response and RT were recorded. After making a response, the picture and the participant's response remained on the screen, and the correct name of the target was printed below the participant's response for 1 second, followed by a 1-second inter-trial interval. Because the correct answer was provided as feedback, participants were able to relearn the 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 in the recognition task was calculated as the proportion of correct responses in each testing block. RT was automatically measured from the onset of the visual display to the time at which the participant clicked on one of the four pictures. In the production task, responses were automatically scored by the computer, yielding two measures of accuracy. Exact accuracy was calculated as the proportion of responses in each block where participants' response exactly matched the correct target name. Partial accuracy assigned fractional points to incorrect responses that demonstrated partial knowledge of the correct target name. Each letter in the correct position received 0.25 points (for a max score of 1). Each extra letter over four incurred a 0.25 point penalty. RTs were automatically measured from the onset of the target picture to the time that the participant submitted their response by pressing the Enter key on the keyboard. RT analyses were performed on correct responses only in order to control for accuracy differences between blocks and conditions. In addition, outlier RTs within each combination of block and condition were identified (threshold = $M + 2SD$s) and replaced with the mean value plus two standard deviations.

Each participant's acquisition rate was calculated for each word as the number of repetitions required until the participant demonstrated mastery of the word. Mastery was defined as the first block in which all subsequent responses for the word were also correct. For example, a correct answer in all five blocks received an acquisition rate of 1, indicating that the single exposure block was sufficient for learning to occur. A correct answer in blocks one, four, and five received an acquisition rate of 4, because a correct answer was recorded in block four and all subsequent responses. Acquisition rates were analyzed using a linear mixed-effects regression model that included fixed effects of Englishlikeness and Germanlikeness, random intercepts of subject and item, and random by-subjects slopes for Englishlikeness and Germanlikeness, using the lmer function in the lme4 package in the statistical package R (Baayen, Davidson, & Bates, 2008). Model comparisons were performed using a likelihood ratio test, and *p* values for parameter estimates were obtained using Satterthwaite's approximation of degrees of freedom (Goodnight, 1976; Schaalje, McBride, & Fellingham, 2002).

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), a technique specifically designed to analyze change over time. Analyses were performed in the statistical package R using the lmer function in the lme4 package. 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 in each task were fit with two model levels. The Level-1 submodels captured the effect of time on changes in the dependent measure over the course of training using second-order orthogonal polynomials. In these models, the intercept term describes the overall height of the curve over the specified time window (i.e., all five testing blocks), the linear term reflects the overall slope, and the quadratic term reflects the curvature (i.e., change in slope across the time window). The Level-2 submodels capture the effects of experimental manipulations and individual differences on each of the time terms present in the Level-1 model through a combination of population means, fixed effects, and random effects. In the current study, the fixed effects corresponded to experimental manipulations of English wordlikeness and German wordlikeness. The random effects captured individual deviances from the global mean and condition means. The effects of individual differences in language background, nonverbal IQ, and phonological memory on learning were assessed by correlating each measure with individuals' random effect estimates obtained from the accuracy and RT models for the recognition and production tasks.

The base Level-2 model included all time terms, fixed effects of English and German proficiency, and random effects of participant and participant-by-condition on all time terms. Additional Level-2 models were built that added three fixed effects of English wordlikeness (E+/E–), German wordlikeness (G+/G–), and their interaction to each time variable in turn. A significant improvement in model fit (a chi-square test on the change in model fit using -2Log Likelihood) indicates an effect of condition on independent properties of the curve (i.e., height, slope, or curvature). Parameter-specific *p* values were estimated by using a normal approximation, treating the *t* value from the model as a *z* value (see Barr, Levy, Scheepers, & Tily, 2013). Production errors were analyzed to determine how learners attempted to fill in gaps in their memory for the novel words. All Colbertian words followed the CVCC structure, and each of the three consonants (i.e., letter positions 1, 3, and 4) were analyzed separately. Errors at the vowel position were not analyzed, because Colbertian only allowed five letters in that position, preventing good model fits to errors. At a single position, the frequency with which each letter in the alphabet was used incorrectly across all items and participants in all blocks was calculated. These usage frequencies were compared to orthotactic probabilities for Colbertian, English, and German that were calculated as in Vitevitch and Luce (2004). Colbertian positional segment probabilities 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. English and German positional segment probabilities 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 or German words, respectively.

At each consonant position, a base linear model of time (testing block) was created on the error frequency of each letter in the alphabet (letters that were never produced or that had zero frequency in that position in English, German, or Colbertian were excluded to enable model comparisons across languages). The effects of adding each language's frequency to the base model were assessed by change in goodness of fit using likelihood ratio tests. The difference in predictive ability across languages was assessed using a nonnested model comparison approach (Merkle, You, & Preacher, 2016; Vuong, 1989).

## Results

### Word Recognition

As expected, participants became faster and more accurate over time. There were significant effects of adding linear, $\Delta LL = 6.08$, $\chi^2(1) = 12.15$, $p < .001$, and quadratic, $\Delta LL = 7.72$, $\chi^2(1) = 33.41$, $p < .001$, time terms to the base model of accuracy, and participants' accuracy improved from 61.6% ($SD = 19.0$) in the first block to 98.4% ($SD = 3.7$) in the fifth block. Recognition accuracy in each block and condition is available in Appendix S2 in the Supporting Information online. The words that were known by block five were acquired after 1.89 exposures ($SD = 1.12$); the acquisition rate was not affected by English or German wordlikeness. For RT, there was a significant effect only of the linear time term, $\Delta LL = 18.9$, $\chi^2(1) = 12.15$, $p < .001$, on the base model, and RTs became faster over time, from 3,325 milliseconds ($SD = 593$) in the first block to 2,208 milliseconds ($SD = 493$) in the fifth block.

The word recognition accuracy model was improved by adding the wordlike predictors to the intercept, $\Delta LL = 4.52$, $\chi^2(3) = 8.10$, $p < .05$. Englishlikeness

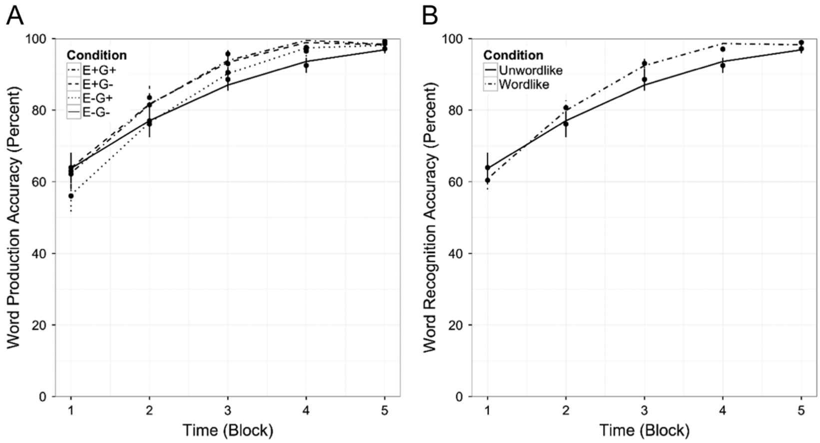


**Figure 1** Accuracy of novel word recognition. Dots and vertical lines mark observed data and standard error; lines are best-fit quadratic growth curve models. Panel A: Englishlikeness increased overall accuracy (intercept height). Panel B: Wordlikeness collapsed across languages affected line curvature, with wordlike items (dotted line) improving more with training and approaching an asymptote at ceiling before unwordlike items (solid line).

raised the overall height of the curve (*Estimate* = .036, *SE* = .016, $p < .05$), reflecting consistently higher accuracy for the Englishlike words compared to non-Englishlike words of 3.6% over the course of training (as illustrated in Figure 1A). With the E+G+, E+G–, and E–G+ conditions combined as a single factor, there was a significant improvement in model fit on the quadratic term, $\Delta LL = 1.16$, $\chi^2(1) = 4.33$, $p < .05$. Wordlikeness changed the curvature of the line (*Estimate* = –.058, *SE* = .027, $p < .05$), reflecting greater increases in accuracy for all wordlike items compared to the unwordlike condition early in training, with the wordlike curve reaching ceiling performance earlier (as shown in Figure 1B).

For recognition RT (depicted in Figure 2), there was a significant improvement to the base model by adding English and German wordlikeness to the intercept, $\Delta LL = 22.16$, $\chi^2(3) = 44.32$, $p < .001$. English wordlikeness reduced RT relative to the baseline (*Estimate* = –.406, *SE* = .061, $p < .001$), and there was a marginal decrease in RT by German wordlikeness (*Estimate* = –.109, *SE* = .061, $p < .1$). Novel words that resembled English were thus correctly identified 406 milliseconds faster than baseline words, whereas words that resembled German were identified 109 milliseconds faster than baseline,

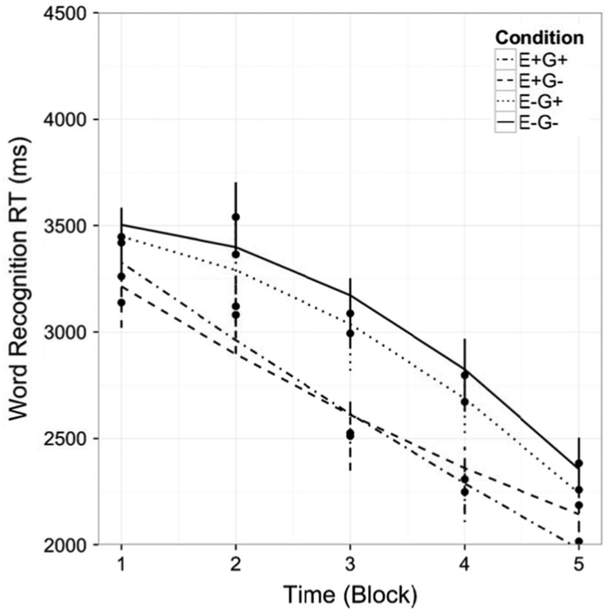


**Figure 2** Response time for novel word recognition. Dots and vertical lines mark observed data and standard error; lines are best-fit quadratic growth curve models. Both English and German wordlikness decreased overall response time across blocks (intercept height), but the two factors did not have an additive effect: Double-wordlike items (E+G+, dot-dash line) were no faster than single-wordlike conditions (Englishlike only, dashed line; Germanlike only, dotted line).

with no interaction between the two factors. These wordlike increases were stable across training, even as RTs globally decreased by 1,117 milliseconds from blocks one to five.

## Word Production

Accuracy improved from 10.3% ($SD = 14.8$) to 66.3% ($SD = 26.1$) over blocks one through five. Production accuracy in each block and condition is available in Appendix S3 in the Supporting Information online. The words that were known by block five were learned after 2.95 exposures ($SD = 1.33$). Acquisition rates were analyzed using a linear mixed-effects regression model that included fixed effects of Englishlikeness and Germanlikeness, random intercepts of subject and item, and random by-subject slopes for English and German wordlikeness. The random slopes did not significantly improve model fit, $\chi^2(5) = 2.61$, *ns*, and thus the model with random intercepts only is reported. Adding English and German wordlikeness to the base model significantly improved model fit, $\Delta LL = 14.10$, $\chi^2(3) = 28.14$, $p < .001$. Relative to the E–G– baseline (*Estimate* = 3.58 exposures, *SE* = 0.18), Englishlikeness (*Estimate* = –.66, *SE* = .13, $p < .001$), and Germanlikeness (*Estimate* = –.41, *SE* = .14, $p < .01$) each decreased the number of exposures needed to learn the word. However, the interaction between English- and Germanlikeness canceled out their additive effect (*Estimate* = .45, *SE* = .18, $p < .05$), yielding no additional learning benefit for double- compared to single-language wordlikeness.

In the model of changes to exact novel word production accuracy over time, there were significant effects of adding linear, $\Delta LL = 22.52$, $\chi^2(1) = 45.03$, $p < .001$, and quadratic, $\Delta LL = 4.70$, $\chi^2(1) = 9.40$, $p < .01$, time terms to the base model. Production RTs for correct responses improved from 3,361 milliseconds (*SD* = 1,672) to 2,998 milliseconds (*SD* = 901) over blocks one through five, and there was a significant effect of the linear time term, $\Delta LL = 4.50$, $\chi^2(1) = 7.01$, $p < .01$, on the base RT model.

The production accuracy model (depicted in Figure 3) was improved by adding English and German wordlikeness to the intercept, $\Delta LL = 25.82$, $\chi^2(3) = 51.64$, $p < .001$, and to the quadratic term, $\Delta LL = 4.89$, $\chi^2(3) = 9.78$, $p < .05$. The overall height of the curve was increased by English (*Estimate* = .22, *SE* = .02, $p < .001$) and German wordlikeness (*Estimate* = .09, *SE* = .02, $p < .001$), and there was a significant interaction between the two terms (*Estimate* = –.12, *SE* = .03, $p < .001$). The combination of the two terms revealed that whereas Englishlikeness improved accuracy by 21.8% and Germanlikeness improved accuracy by 8.7% relative to the unwordlike baseline, they combined nonadditively, as the E+G+ double-wordlike condition was only 18.8% above baseline. Additionally, the benefits of English and German wordlikeness were not equivalent, as the height of the learning curve for E–G+ words was significantly lower than both the E+G– words (*Estimate* = .13, *SE* = .02, $p < .001$) and the E+G+ words (*Estimate* = .10, *SE* = .02, $p < .001$).

The curvature of the learning gains over time (i.e., the quadratic term in the model) was significantly affected by Englishlikeness (*Estimate* = –.10, *SE* = .03, $p < 0.01$), but not by Germanlikeness (*Estimate* = –.05, *SE* = .03, *ns*). The baseline quadratic term was also not significant (*Estimate* = –.03, *SE* = .03, *ns*), and together, these results indicate that whereas accuracy gains between blocks were nearly linear for baseline words, Englishlikeness had a nonlinear effect on change in accuracy over time, with the largest accuracy gains between blocks occurring earlier during training. Word production RTs were not affected by wordlikeness in English or German and varied considerably across participants.

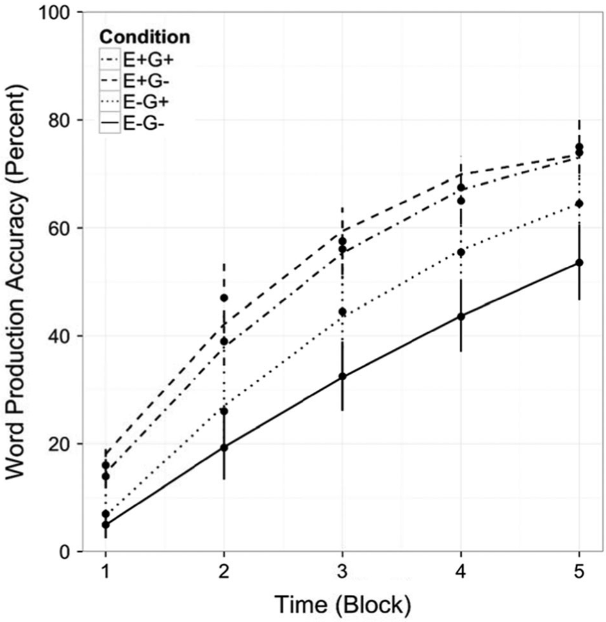


**Figure 3** Accuracy of novel word production. Dots and vertical lines mark observed data and standard error; lines are best-fit quadratic growth curve models. Both English and German wordlikness increased overall accuracy (intercept height), but the two factors did not have an additive effect. Both Englishlike conditions (E+G–, dashed line; E+G+, dot-dash line) had higher accuracy than Germanlike only (E–G+, dotted line).

## Errors and Partial Accuracy

Partial word accuracy (illustrated in Figure 4) captures additional information about incremental gains in word knowledge. We found significant effects of wordlikeness on the intercept, $\Delta LL = 6.47$, $\chi^2(3) = 12.93$, $p < .01$, and linear terms, $\Delta LL = 6.54$, $\chi^2(3) = 13.07$, $p < .01$. Englishlikeness decreased overall height of the curve (*Estimate* = –.03, *SE* = .01, $p < .05$), indicating that Englishlike words tended to be acquired as complete units, and were less likely to be partially produced. For the slope of the curve, Englishlikeness (*Estimate* = –.09, *SE* = .02, $p < .001$) and Germanlikeness (*Estimate* = –.06, *SE* = .02, $p < .05$) each led to negative slopes compared to baseline, and there was a significant nonadditive interaction between the two terms (*Estimate* = .09, *SE* = .03, $p < .01$). In combination with the gains in exact accuracy over time, these results suggest that wordlike items that were partially known in early blocks transitioned to completely correct words (and thus no longer contributed

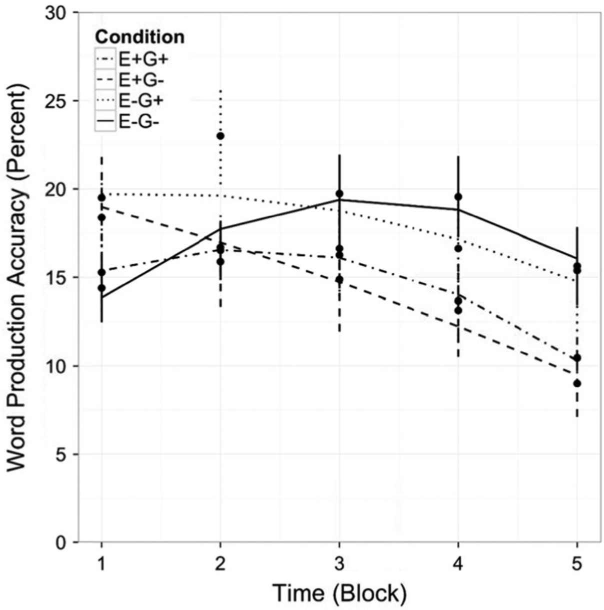


**Figure 4** Partial accuracy of novel word production. Dots and vertical lines mark observed data and standard error; lines are best-fit quadratic growth curve models. English and German single-wordlike conditions (dashed and dotted lines, respectively) decreased in partial accuracy over time as partially correct answers transitioned to correct responses.

to partial accuracy scores), whereas partial knowledge across all unwordlike items increased over time, reflecting a longer transitionary period from partial to full knowledge.

Single-letter production errors were analyzed with a 5 (block) × 4 (letter position) within-subjects analysis of variance; mean errors in each block and position are available in Appendix S4 in the Supporting Information online. Errors decreased from each block to the next, indicating continuous gains in performance through a significant main effect of block, $F(4, 76) = 40.1$, $p < .001$ (pairwise comparisons all $p$s < .05, Holm correction). In addition, errors were unevenly distributed across letter positions in a word, based on a significant main effect of letter-position, $F(3, 57) = 30.85, p < .001$. The vowel had fewer errors than all consonants, and the first consonant had fewer errors than the final two consonants, suggesting a primacy effect in memory for word onsets (all $p$s < .05, Holm correction). There was no interaction between block and position, $F(12, 228) = 1.52$, *ns*.

To examine how well participants extracted Colbertian letter distributions, the effects of Colbertian, English, and German letter frequencies on the types of errors that participants produced were reviewed. We compared the fit of linear models that included only time to models that included time and either Colbertian, English, or German letter frequency. Each of the three consonant positions within a word (i.e., letter positions 1, 3, and 4) was analyzed separately (the vowel position was not analyzed, because Colbertian only allowed five letters in this position, preventing good model fits based on Colbertian letter frequencies). In the first consonant position, Colbertian frequency and English frequency each improved fit compared to the base model, but German frequency did not: Colbertian model Likelihood Ratio (LR) = 42.38, $p < .001$; English model LR = 8.19, $p < .001$; German model LR = 0.25, *ns*. The Colbertian model was a significantly better fit to the data than either the English (Vuong's nonnested model comparison, $z = 2.94$, $p < .01$) or the German ($z = 3.61$, $p < .001$) models, while English was a better fit than German ($z = 2.18$, $p < .05$). At the second consonant position, each of Colbertian (LR = 34.40, $p < .001$), English (LR = 25.45, $p < .01$), and German (LR = 33.28, $p < .001$) frequencies improved fit over the base model, but no model was a better fit than any other (all *p*s > .05). In the final consonant position, neither Colbertian, English, nor German frequencies affected errors, though there was a marginal effect of Colbertian frequency on model fit (LR = 8.22, $p < .1$).

In sum, the frequency with which letters appeared in the Colbertian vocabulary had a pronounced effect on the types of errors participants made at word onset, indicating sensitivity to letter distributions within the novel language. English was not as good a predictor of onset errors, and German even less so (given the high correlation between English and German letter distributions at word onset for four-letter words, $R^2 = .61$, $p < .001$, these results suggest application of English-specific knowledge, instead of common letter patterns). Errors at the final two letters were no better predicted by Colbertian frequency than either English or German; the evidence for learning and use of Colbertian distributional probabilities was primarily at word onset.

**Individual Differences**

We used the random-effect terms in the accuracy and RT models to obtain individual effect sizes for each analysis. These random effects quantify how much individual participants' performances deviated from the group mean in curve height, slope, and curvature. In order to assess the effects of these predictors on word learning, we correlated the random effects with individuals' language proficiency, age of acquisition, current exposure, and language balance (the relative differences between both languages), as well as nonverbal IQ and phonological memory.

Language proficiency balance (i.e., the relative difference in English and German self-rated proficiencies) was associated with individual differences in learning rate for the Englishlike compared to Germanlike words in the production task. Specifically, higher proficiency in English relative to German was associated with a faster learning rate (larger random-effect term for slope) for the Englishlike relative to the Germanlike words. Conversely, higher relative German proficiency was associated with a faster learning rate for the Germanlike compared to the Englishlike words ($r = .48$, $R^2 = .22$, $p < .05$).

Overall language proficiency was associated with individual differences in recognition accuracy. English vocabulary size (LexTALE score) was correlated with a higher intercept ($r = .65$, $R^2 = .42$, $p < .01$) and a shallower slope ($r = -.62$, $R^2 = .39$, $p < .01$) of the accuracy curve. English and German self-rated proficiencies were each marginally correlated with higher intercepts (English proficiency $r = .42$, $R^2 = .18$, $p < .1$; German proficiency $r = .41$, $R^2 = .22$, $p < .1$) and shallower slopes (English proficiency $r = -.47$, $R^2 = .17$, $p < .05$; German proficiency $r = -.43$, $R^2 = .19$, $p = .05$). In each case, English or German knowledge increased accuracy; the shallower slopes are a consequence of participants reaching ceiling performance in the task. No other linguistic characteristics were related to learning performance.

Higher nonverbal IQ was correlated with more negative-trending slopes in the word production RTs ($r = -.52$, $R^2 = .27$, $p < .05$) and greater curvature ($r = .50$, $R^2 = .25$, $p < .05$). These results suggest that participants with higher nonverbal IQs reduced their production RTs at a faster rate between blocks than participants with a lower IQ. Memory capacity did not correlate with learning measures.

## Discussion

In the current study, we found that lexical similarity to a single known language improved bilinguals' learning of novel written words as much as simultaneous overlap with two known languages. These results indicate that L3 vocabulary learning benefits from each language and that bilinguals can flexibly transfer L1 and L2 knowledge to the L3, when appropriate, at early stages of instruction. The lack of an additive learning benefit for words with close lexical neighbors and familiar patterns in both languages suggests that early vocabulary transfer may occur through a process of linking novel words to anchors in a single language. This process most closely resembles the scaffolding model of word learning, with limited evidence for the accumulation model. In addition, the pattern by which learners attempted to infer unknown words' spellings was influenced by statistical distributions of letter frequency within the novel language, indicating that sensitivity to L3 sublexical patterns begins at early stages of L3 acquisition.

**Acquisition of Word Form and Meaning**

The current study used two tasks, word recognition and production, to assess different aspects of novel word learning. The recognition task probed the formation of word–meaning links without imposing additional memory demands for word form. Successful acquisition (defined as a consistently accurate response for a word) took on average less than two exposures ($M = 1.89$). In contrast, the production task required recall of the L3 written word form from only a picture prompt and assessed partial as well as full word knowledge. Form acquisition required one additional exposure ($M = 2.95$), reflecting the increased difficulty of storing and retrieving a novel word's correct form. By designing novel words whose forms overlapped with letter and word knowledge in either one, both, or neither of the two languages (English and German), we were able to investigate how two languages interact to affect vocabulary learning in a L3.

Accuracy in the recognition task was relatively high even after a single exposure to a word–picture pair. In the first block, participants recalled roughly 30 of the 48 pairs correctly (61.6%, with chance performance at 25%) and quickly approached ceiling performance. English, but not German, wordlikeness increased overall accuracy across training relative to baseline words. In addition, participants with larger English vocabularies performed better in the task, with higher accuracies across training. Marginal correlations between individuals' proficiency in either language and accuracy suggest the possibility of a generalized vocabulary size benefit on recognition learning. For word recognition, there is more support for an effect of English similarity than of German; this difference between languages reflects the overall higher English proficiency in our sample. Because the novel words were presented in their entirety during the recognition task, accurate performance depended not on memory for word forms, but on the link between form and meaning. This association between form and meaning may be less sensitive to language-specific knowledge than the memory for word forms, which may be why the effects of English and German wordlikeness were more pronounced in the production task when compared to recognition.

The combined factor describing wordlikeness in either language, however, had a significant effect on model curvature. Whereas the wordlike items were actually responded to less accurately than unwordlike items in the first testing block, the wordlike items rose steeply in accuracy with additional training and reached ceiling performance earlier than unwordlike items. The early advantage for unwordlike items is consistent with models of word learning that trigger learning upon encountering a word not found in the existing lexicon; this process occurs more readily for unique words with no near neighbors (Carpenter & Grossberg, 1987; Gupta & MacWhinney, 1997; Storkel et al., 2006). Improving storage and retrieval of a novel word, though, occurs at a faster rate for wordlike items, which benefit more from overlap with existing lexical items, once they are detected as novel and learning begins.

The production task, in contrast, was designed specifically to probe participants' memories for the actual written forms in the artificial language. With only a single picture prompt, the participant's task was to type the matching word from memory. In this task, we saw evidence for strong effects of wordlikeness both in English and in German. All three wordlike conditions (E+G+, E+G–, and E–G+) had higher curve heights than the baseline (E–G–) words, indicating higher accuracy throughout the experiment. However, the combination of English and German wordlikeness was not additive, as the E+G+ words were no different from the better of the two single-wordlike conditions (E+G–). This pattern of results provides support for the scaffolding account, by which the novel words received a benefit to learning if they overlapped with at least one of bilinguals' known languages, but received no additional benefit for overlapping both languages.

**Vocabulary Transfer From L1 and L2 to L3**

Bilingual L3 learners appear to be especially sensitive to perceived crosslinguistic overlap and will transfer vocabulary knowledge from their two languages preferentially based on typological similarity, regardless of other factors like age of acquisition (Cenoz, 2003; Jarvis & Odlin, 2000; Llama, Cardoso, & Collins, 2010; Ringbom, 2001). English and German are closely typologically related to each other, and words in the novel language were designed to draw on this similarity, maximizing opportunities for transfer. As a result, we would expect participants in this study to transfer knowledge from the language of overlap for E+G– or E–G+ words, and either language for E+G+ words. In post-experiment debriefings, 95% of participants reported a strategy for learning the words' spellings by creating a mental image bridging the picture and novel word by means of a similarly spelled existing word. This strategy resembles the successful keyword learning method (Shapiro & Waters, 2005). Participants reported relying on a mix of English and German anchors. In most reported cases, participants used an English keyword for Englishlike words and a German keyword for Germanlike words. For example, to learn that the novel E+G– word *sumb* meant "fork," one participant imagined having to count the total number of forks in a drawer (i.e., the *sum*). Another participant reported learning that the novel E–G+ word *kenf* meant "goat" by thinking of a goat eating *senf*, the German word for mustard. Words that resembled both English and German, like the word *duch* meaning "eyeglasses," were sometimes learned through an English keyword (a *duck* wearing glasses), but other times a German keyword (using glasses to read a *buch*, the German word for book). By not limiting themselves to a single language of transfer, learners were able to maximize the benefits from their existing knowledge.

The keyword strategy enabled participants to make use of words in either language as appropriate, but it only required them to make a single, strong association. A larger pool of possible anchors to select from should increase the possibility that a strong keyword candidate is found. In this study, though, whereas the E+G+ words had conceivably twice as large a pool as the single-overlap items, no learning advantage for E+G+ words was found. This apparent discrepancy can be resolved by considering the meanings of the Colbertian words that participants learned. All of the pictures paired with Colbertian words referred to high-frequency, easily imageable objects, which are easier to learn by keywords than less imageable concepts (Shapiro & Waters, 2005). Thus, even a small pool of possible anchors is sufficient for learners to create a strong, memorable association. For abstract or less imageable concepts, the number of related words may have a larger effect on learning. In this case, novel words with large neighborhoods either within or across languages should be learned more easily using keywords than words with smaller neighborhoods.

While both languages provided benefits to memory for novel word forms, the sizes of the effects were not equivalent. Across all participants, English wordlikeness had a larger effect on overall production accuracy compared to German wordlikeness and led to a slightly different curvature over the course of training. These patterns are consistent with participants' proficiency asymmetry reflecting English dominance. Although participants were highly proficient in both languages, all participants were currently living and working in the United States, and had slightly higher English proficiency. L2 proficiency can affect the degree to which it influences L3 learning (Hammarberg, 2001; Tremblay, 2006); accordingly, the largest accuracy gains were seen for English wordlikeness, and overgeneralization errors were influenced more by English than German. At the individual level, however, we found that learning patterns were influenced by relative proficiencies in English and German. Bilinguals with higher proficiency in English learned to produce the Englishlike words at a faster rate than the

Germanlike words, while the opposite was true for those with higher proficiency in German. This difference may reflect either the relative ease of acquisition of individual words or an attention allocation strategy that prioritized words resembling the learner's dominant language.

Accuracy alone does not tell the whole story, however, because words can be partially learned at one point and only later recalled completely. As a result, partial knowledge provides a useful metric for investigating the development of novel word knowledge over time. In the current study, whereas partial accuracy for the unwordlike baseline increased over training, English and German wordlikeness each reversed this trend, leading to decreasing partial accuracy scores over time. This decrease reflects a transition from partial to full knowledge, as a portion of the incompletely recalled words in one block was likely to be recalled accurately in the next. In addition to these effects of wordlikeness on slope, we saw that English, but not German, similarity decreased overall height of the curve. The lack of a German similarity effect reflects the high partial accuracy scores for the E–G+ condition. This finding complements the exact-accuracy results, which showed that curve height for the two Englishlike word conditions was significantly higher than the E–G+ condition. Due to lower German proficiency, the bilinguals in our sample were more likely to make single-letter errors in their productions of the Germanlike words, but when given credit for their partial word knowledge, their knowledge of the Germanlike words more closely approached the other word types.

In addition to the effect of letter patterns in English and German, evident in partially correct responses, post hoc error analyses revealed an effect of letter patterns characteristic to the novel language on learning. The novel language was designed to be broadly similar to both English and German—all words were constrained to CVCC structure and pronounceable based on English and German phonotactics. However, Colbertian letter frequency diverged in several ways from English and German. A learner who is sensitive to Colbertian letter distributions could augment their word memory by filling in gaps with probable letters. By examining errors in production, it is possible to detect evidence of learners' gap-filling heuristic based on their overgeneralization of learned letter frequencies. Bilinguals' pattern of errors at word onset was predicted better by Colbertian frequencies than either English or German, suggesting that participants were sensitive to and actively utilized letter frequencies in the new language to supplement their memory for individual words. Both children and adults are adept at learning statistical regularities from input within a short time span, even when they are not explicitly aware of learning such regularities (Newport & Aslin, 2004; Saffran, Johnson, Aslin, & Newport, 1999). Here, we observed evidence for similar learning of letter frequencies within a novel language.

### Individual Differences and Learning

Although there were consistencies in how participants were affected by wordlikeness in the novel language, there was also notable variability in individual performance, and thus we explored how individual differences may impact learning. The relative difference between an individual's English and German proficiency increased the learning rate for words that resembled the dominant language. Absolute proficiency levels in each language reflected consistent differences in learning of surface form–meaning associations. Nonverbal intelligence was associated with greater speed increases between blocks in producing the novel words. This speed increase may have been the result of fluent productions reinforcing the correct sequence, facilitating the next production. Participants with lower general intelligence may have been less able to utilize this production feedback process. Short-term memory is important for storage and recall of short novel sequences, and previous work found that memory capacity in monolinguals was correlated with the learning of novel written words, regardless of overlap (Bartolotti & Marian, 2014, in press). However, no such relationship was found in the current study. As bilinguals generally perform better on tests of short-term memory than monolinguals (Majerus et al., 2008; Papagno & Vallar, 1995; Service, Simola, Metsänheimo, & Maury, 2002), it may be that bilinguals in this study had sufficient capacity to meet the demands of the single-word learning context, and thus individual differences in memory capacity were not a primary factor in predicting learning outcomes.

## Conclusion

To conclude, we found evidence for effects of wordlikeness in each of a bilingual's two languages on L3 orthographic word learning. Memory for word forms was often improved by linking novel vocabulary to existing lexical anchors in either of a bilingual's already known languages, depending on the similarity of the novel word to lexical patterns in English and German. Importantly, a novel word's similarity to both of a bilingual's known languages does not provide an additional learning benefit beyond similarity to a single language, suggesting that orthographic knowledge does not necessarily combine additively during L3 learning. These results provide support for the scaffolding account of word learning, in which existing language knowledge provides a framework upon which novel words in another language can be built, accelerating early stages of language acquisition.

These findings have broad implications for adult language instruction by demonstrating the learning benefits to be gained by utilizing areas of overlap between a native and novel language. Although foreign language instruction has long placed a premium on total immersion, there is growing evidence that careful utilization of a learner's native language can benefit learning (Lin, 2015), particularly when capitalizing on the similarities and differences across languages (Laufer & Girsai, 2008). Adult language users have spent years developing a finely honed linguistic system that uniquely shapes lifelong learning in any language they speak. This experience can be an asset when overlap across languages is used to accelerate early stages of language acquisition. Bilingual L3 instruction in particular can further benefit from utilizing overlap between the novel language and each existing language as appropriate. Vocabulary acquisition is a critical, but potentially difficult aspect of language learning, and crosslinguistic similarities can be a powerful component of language instruction to increase learners' success.

Final revised version accepted 10 April 2016

## References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412. doi:10.1016/j.jml.2007.12.005

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford Univeristy Press.

Bailey, T., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, *44*, 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*, 1–27. doi:10.1177/00238309050480010101

Barcelona European Council. (2002). *Presidency conclusions* (Vol. *35*). Retrieved from http://data.consilium.europa.eu/doc/document/ST-6993-2002-INIT/en/pdf

Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals*, *37*, 200–208. doi:10.1111/j.1944-9720.2004.tb02193.x

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278. doi:10.1016/j.jml.2012.11.001

Bartolotti, J., & Marian, V. (2014). Wordlikeness and novel word learning. In *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 146–151). Austin, TX: Cognitive Science Society.

Bartolotti, J., & Marian, V. (in press). Interaction of prior orthographic knowledge and lexical form during novel language learning. *Applied Psycholinguistics*.

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., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition*, *15*, 397–401. doi:10.1017/S136672891100040X

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, *13*, 525–531. doi:10.1017/S1366728909990423

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. 401–424). Amsterdam: Elsevier.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897x00357

Brooks-Lewis, K. A. (2009). Adult learners' perceptions of the incorporation of their L1 in foreign language teaching and learning. *Applied Linguistics*, *30*, 216–235. doi:10.1093/applin/amn051

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*, 977–990. doi:10.3758/BRM.41.4.977

Carpenter, G. A., & Grossberg, S. (1987). A massively parallel architecture for a self-organizing machine neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, *37*, 54–115. doi:10.1016/s0734-189x(87)80014-2

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*, 71–87. doi:10.1177/F13670069030070010501

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*, 1–25. doi:10.1111/j.0023-8333.2005.00294.x

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*, 164–194. doi:10.1093/applin/aml015

European Commission Special Barometer. (2006). *Europeans and their languages*. Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_en.pdf

Gallup Organization. (2001). Gallup Poll, March 2001, #12 [Data Set].

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

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 84–95. doi:10.1037/0278-7393.25.1.84

Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73–100). Hove, UK: Psychology Press.

Goodnight, J. H. (1976). The new general linear models procedure. In *Proceedings of the First Annual SAS Users Group International Conference* (pp. 1–39). Cary, NC: SAS Institute.

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

Gupta, P., & Tisdale, J. (2009). Word learning, phonological short-term memory, phonotactic probability and long-term memory: Towards an integrated framework. *Philosophical Transactions of the Royal Society of London, B*, *364*, 3755–3771. doi:10.1098/rstb.2009.0132

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (Vol. 31, pp. 21–41). Bristol, UK: Multilingual Matters.

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430. Retrieved from http://eric.ed.gov/?id=EJ626518

Jarvis, S., & Odlin, T. (2000). Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition*, *22*, 535–556. doi:10.1017/S0272263100004034

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

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language*, *33*, 630–645. doi:10.1006/jmla.1994.1030

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36* (ECVP Abstract Supplement).

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174. doi:10.1006/jmla.1994.1008

Laufer, B. (2000). Task effect on instructed vocabulary learning: The hypothesis of "involvement." In *Selected Papers from AILA 1999 Tokyo* (pp. 47–62). Tokyo, Japan: Waseda University Press.

Laufer, B., & Girsai, N. (2008). Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics*, *29*, 694–716. doi:10.1093/applin/amn018

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343. doi:10.3758/s13428-011-0146-0

Liao, P. (2006). EFL learners' beliefs about and strategy use of translation in English learning. *RELC Journal*, *37*, 191–215. doi:10.1177/0033688206067428

Lin, A. M. Y. (2015). Conceptualising the potential role of L1 in CLIL. *Language, Culture and Curriculum*, *28*, 74–89. doi:10.1080/07908318.2014.1000926

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*, 39–57. doi:10.1080/14790710902972255

Long, L. L., & Shaw, R. J. (2000). Adult age differences in vocabulary acquisition. *Educational Gerontology*, *26*, 651–664. doi:10.1080/03601270050200644

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

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*, 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

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940–967. doi:10.1044/1092-4388(2007/067).

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods*, *21*, 151–163. doi:10.1037/met0000038

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

Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*, 59–81. doi:10.1353/cml.2006.0049

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.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology*, *48A*, 98–107. doi:10.1080/14640749508401378

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

Ringbom, H. (2001). Lexical transfer in L3 production. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 59–68). Clevedon, UK: Multilingual Matters.

Rivers, W. P., Robinson, J. P., Harwood, P. G., & Brecht, R. D. (2013). Language votes: Attitudes toward foreign language policies. *Foreign Language Annals*, *46*, 329–338. doi:10.1111/flan.12048

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

Schaalje, G. B., McBride, J. B., & Fellingham, G. W. (2002). Adequacy of approximations to distributions of test statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics*, *7*, 512–524. doi:10.1198/108571102726

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199–227). Cambridge, UK: Cambridge University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. doi:10.1177/1362168808089921

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*, 26–43. doi:10.1111/j.1540-4781.2011.01146.x

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*, 168–175. doi:10.3758/BF03202729

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

Service, E., Simola, M., Metsänheimo, O., & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, *14*, 383–408. doi:10.1080/09541440143000140

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

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

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

Tremblay, M. (2006). Cross-linguistic influence in third language acquisition: The role of L2 proficiency and L2 exposure. *Cahiers Linguistiques d'Ottawa*, *34*, 109–119.

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, & 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.

## Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

**Appendix S1.** List of Stimuli.
**Appendix S2.** Recognition Task Accuracy.
**Appendix S3.** Production Task Accuracy.
**Appendix S4.** Production Errors by Letter Position and Block.