# IS THE EFFECT OF L2 VOCABULARY SIZE ON LEXICAL ENCODING MODULATED BY L1-L2 SIMILARITY?

Miquel Llompart[13], Brian Rocca[2], Isabelle Darcy[2]

[1]Universitat Pompeu Fabra, Spain, [2]Indiana University, USA, [3]Friedrich Alexander University, Germany
miguel.llompart@upf.edu, brocca@iu.edu, idarcy@indiana.edu

## ABSTRACT

Previous research suggests that second language (L2) vocabulary size is a key predictor of how robust the phonological encoding of challenging L2 sounds into lexical representations is. Here we test whether this is also the case when the native language (L1) and the L2 are orthographically and lexically dissimilar. Korean learners of English completed an auditory lexical decision task (LDT) with phonological substitutions, a categorization task assessing their perception of the English /ɛ/-/æ/ distinction, and a vocabulary test. Results showed that vocabulary scores predicted /ɛ/-/æ/ nonword (e.g., m[ɛ]tch) rejection rates in the LDT task and that this effect was larger for participants with better perception skills for the target contrast. These results further support the role of L2 vocabulary size as a major determiner of phonological improvements in the representation of L2 words and speak against an account merely based on the mediating role of L1-L2 orthographic and lexical similarity.

**Keywords**: L2 learning, speech perception, lexical encoding, vocabulary size, L1-L2 similarity

## 1. INTRODUCTION

Learning a second language (L2) requires familiarizing oneself with a phonological inventory containing sounds and sound contrasts that are not part of the native language (L1) while simultaneously being exposed to vast amounts of L2 words. Therefore, learners need to acquire the ability to perceptually identify and produce the L2 categories that are not part of the L1 inventory. At the same time, they must create lexical representations for many newly-learned words and store them in long-term memory.

While there is much research on perceptual improvements in L2 learning, less is known about the development of L2 lexical representations and about how perception and word learning interact at different stages in the acquisition process. A recurrent finding of recent studies has been that, even in cases in which a robust perceptual identification of challenging L2 categories is in place, learners still struggle during spoken word recognition if recognition is dependent on such sounds [1, 2]. This is thought to be due to a combination of increased processing demands in word recognition tasks, unreliable acoustics-to-lexicon mappings and phonologically imprecise lexical representations [3, 4].

Regarding the phonological form of lexical representations, it has been suggested that improvements in perceptual ability are crucial and necessary for an accurate encoding of L2 sounds at the lexical level (henceforth *lexical encoding*), yet not sufficient on their own [5]. Crucially, recent studies have additionally shown that, at least for proficient learners, the precision of lexical encoding is connected to their lexical knowledge (operationalized as vocabulary size), above and beyond sheer perceptual ability [5, 6]. In [5], the lexical encoding of English /ɛ/ and /æ/ by German learners of English was assessed, showing that, while the ability to perceptually identify the vowels in a phonetic task predicted lexical encoding for learners of intermediate proficiency, it was vocabulary size that did so for the most advanced learners in the sample. Similarly, in [6], vocabulary size was the main predictor of the lexical encoding of the challenging Spanish segments /r/, /ɾ/ and /d/ for English speakers of intermediate-to-high proficiency in Spanish.

These findings have been crucial to highlight the need for a deeper understanding of the interplay between phonetic, phonological and lexical knowledge in L2 speech learning. In particular, more research is needed to determine the extent to which the relationship between vocabulary and lexical encoding holds across different combinations of L1s and L2s. This is especially relevant because the findings of previous studies were obtained for languages with a common script and a large overlap in their vocabularies (i.e., German-English and English-Spanish). Consequently, it cannot be ruled out that the effect of L2 vocabulary size is actually directly mediated to some extent by the L1 vocabulary and/or by the phonological and orthographic similarities between the L1 and L2 vocabularies.

Considering this, in the present study we test L1-Korean intermediate to advanced learners of English on their perception and lexical encoding of the /ɛ/-/æ/ contrast, which is known to also pose difficulties for this population (see [3]), and measure their vocabulary size in an attempt to replicate the findings in [5]. Crucially, this means that now the L1 and L2 of these learners are not only expected to share fewer cognates and phonologically similar words than in [5] (see [7]), but also have different orthographic scripts. This constitutes a stringent test for the effects of L2 vocabulary previously observed.

For the German learners in [5], L1-L2 lexical similarity could have contributed to a tight coupling between L2 vocabulary and lexical encoding both indirectly through the degree of congruency in grapheme-to-phoneme correspondences (GPC) across languages with the same script and directly through the existence of many phonologically-similar lexical items. L1-congruent GPCs can be beneficial in the establishment of phonological contrasts in the lexicon [8]. In a similar way, cognates can highlight the existence of particular contrasts if a similar phonological contrast can already be inferred from the cognates' L1 form.

For instance, for /ɛ/ and /æ/, which in English are most often orthographically represented as <e> and <a>, the GPC of German for these same symbols would point towards the existence of a phonological contrast. In German <e> and <a> are mostly linked to the phonological categories /e/ or /ɛ/, and /a/, respectively. These similarities in GPC would be particularly obvious, and possibly also particularly useful for learning, in phonologically similar words such as cognates and near cognates. For example, *panda* is written in the same way in English and German and, in both languages, the first vowel, albeit different in English (/æ/) and German (/a/), would sound different from that of the word *pencil* (/ɛ/).

Therefore, if one expects that L1-L2 lexical similarity mediates the effect of L2 vocabulary on lexical encoding, this should be much weaker, if not non-existent, for Korean learners of English. If these learners, whose L1 is orthographically represented through a different script and its lexical similarity to English is considerably lower than that of German [7], do indeed show an effect of vocabulary size on lexical encoding, this will be taken as crucial evidence for a major role of L2 lexical knowledge in the process of establishing phonologically-robust L2 lexical representations.

## 2. METHODS

### 2.1. Participants

Thirty-five Korean learners of English (7 males) were recruited to participate in the experiment. To be included, participants had to be native speakers of Korean as well as intermediate or advanced L2 learners of English. To meet the proficiency criterion, participants had to have taken at least one course at a university or college that used English as the primary language of communication. Table 1 provides the participants' mean age, age of onset, years spent in an English-speaking country, current English usage at home and at university/work, and self-reported proficiency for English comprehension and speaking.

| Measure | Mean (SD) |
|---|---|
| Age | 29.4 (5.4) |
| Age of onset (speaking) | 8.8 (5.1) |
| Years in English-speaking country | 5.4 (4.9) |
| Current L2 use at home | 2.3 (1.7) |
| Current L2 use at university/work | 4.7 (1.6) |
| English comprehension | 4.0 (0.8) |
| Spoken English | 3.9 (0.8) |

**Table 1:** Learner group characteristics. Note: L2 use ratings were given on a 1-6 scale (1=no English, 6=only English); proficiency ratings were given on a 1-5 scale (1=no ability, 5=perfect).

### 2.2. Materials and procedure

The three experimental tasks in this study were programmed in PsychoPy 3 [9], and participants were tested individually via the experimental platform Pavlovia. All participants wore headphones, used a keyboard, and met with one of the researchers remotely on Zoom during the experiment.

#### *2.2.1. Lexical decision task*

This task assessed the lexical encoding of English words containing /ɛ/ and /æ/. Participants were auditorily presented with real words and nonwords with phonological substitutions. The materials were 328 English words. 64 of these words were mono- or disyllabic words that contained the critical /ɛ/-/æ/ contrast—32 for each vowel. The rest of the words targeted other contrasts that can be considered fillers for the purposes of this study. Half of the /ɛ/-/æ/ items were real words and half nonwords. Nonwords were created by switching the two target sounds in each contrast. For example, an /æ/ real word "match" was transformed into the /æ/ nonword "m[ɛ]tch". The stimuli were recorded by a native speaker of American English. First and second formant values for /ɛ/-/æ/ items are provided in Table 2.

| | Mean F1 (SD) | Mean F2 (SD) |
|---|---|---|
| /æ/ nonword | 881 (54) | 1947 (100) |
| /æ/ real word | 1060 (69) | 1854 (87) |
| /ɛ/ nonword | 1024 (48) | 1845 (64) |
| /ɛ/ real word | 842 (50) | 1958 (101) |

**Table 2**: Vowel formants for /ɛ/-/æ/ real words and nonwords.

On each trial, participants heard an auditory stimulus (i.e., a word or a nonword) and indicated whether it was a real word or not by pressing 1 or 0 on the keyboard, such that nonword responses were always provided with the dominant hand.

### *2.2.2. Perceptual categorization task*

A two-alternative forced-choice categorization task (2AFC) on the steps of a *bet-bat* continuum was used to assess the sharpness of participants' perceptual categories for /ɛ/ and /æ/. The words "bet" and "bat" were recorded by the same speaker who recorded the words for the LDT. These natural productions were then used to create a 21-step continuum in Matlab using the STRAIGHT morphing algorithm [10].

On each trial, participants saw a picture representing "bet" on the left side of the computer screen and a picture representing "bat" on the right and heard one of the steps of the continuum. They were asked to press 1 or 0 to indicate whether they heard "bet" or "bat", respectively. The 21 steps from the continuum were presented 10 times (210 trials) in a pseudorandomized order.

### *2.2.3. Vocabulary test*

The vocabulary component of Shipley-2 [11] was administered to assess participant's English vocabulary. This is a multiple-choice test with 40 items of differing lexical frequencies. A word is presented in capitals and four response options are provided. Participants are asked to select the response that they believe is a synonym of the capitalized word. The items were presented in the fixed order of the test.

## 3. RESULTS

Before entering any analyses, data from 5 participants were excluded, leaving a total of 30 participants. One participant was excluded because their accuracy on the LDT control items was more than 2.5 SDs below the group mean and two more were excluded because part of their data did not save correctly. Two other participants were excluded because their categorization function in the perception task was the opposite from what was expected (i.e., had clearly negative slopes), suggesting that they had reversed the response keys. LDT items that were responded to with a mean accuracy 2.5 SDs below the mean were also removed from the analysis for all participants. Finally, items which individual participants identified as unfamiliar in a subsequent questionnaire were also excluded on an individual basis.

For the LDT, the average accuracy for /ɛ/ and /æ/ real words was 95% (SD=22%) and 97% (SD=18%), respectively, while the average score for /ɛ/ and /æ/ nonwords was 28% (SD=45%) and 14% (SD=35%), respectively. Figure 1 shows the mean accuracy by participant by condition.

Perception was measured by calculating the steepness of the /ɛ/-/æ/ categorization curve in the 2AFC task. Following [5], individual slopes were calculated by submitting the categorization data to a generalized linear mixed-effects regression model (GLMM) with a logistic linking function (lme4 package) with Response (coded as 0 and 1) as the categorical dependent variable, an intercept term, and a random slope for Continuum step over Participants. The slope coefficient for each participant was extracted from the model. This coefficient quantifies the increase in log-odds of a "bat" response as a function of an increase of one unit for continuum step. Hence, the higher the slope coefficient, the steeper the slope of the categorization function is. The average slope on the perception task was 0.46 (SD=0.26) with a range of -0.1 - 1.08.

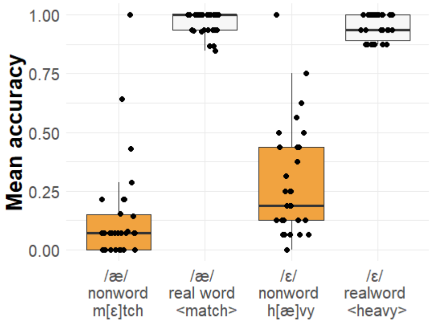


**Figure 1**: Mean accuracy by participant in the LDT as a function of lexical status and vowel.

For the vocabulary test, the number of correct responses out of 40 was calculated for each participant. The average percentage correct was 72.5% (SD=0.14) with a range of 47.5-92.5%. Figure 2 shows the performance for each participant in the perception and vocabulary tasks.

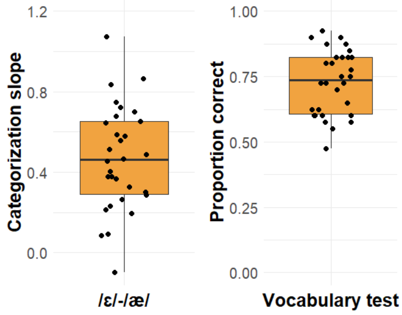


**Figure 2**: /ɛ/-/æ/ categorization slopes (left) and vocabulary test scores (right) by participant.

Because of the clear ceiling effects with real words (see Figure 1), statistical analyses on LDT data were limited to /ɛ/-/æ/ nonword trials. These data were submitted to a GLMM with a logistic linking function.[i] The categorical dependent variable was Response (0=incorrect, 1=correct), and the independent variables were Perception score, Vocabulary size, Vowel, and the interaction between Perception score and Vocabulary size. The variable Vowel was contrast-coded with /æ/ as -0.5 and /ɛ/ as 0.5 and Perception score and Vocabulary size were centered and scaled. The random-effects structure included random intercepts for Participants and Items. A random slope for Vowel over Items was not included because it did not improve the model's fit. The results of the model are provided in Table 3.

| Predictor | *b* | std error | *z* | *p* |
|---|---|---|---|---|
| Intercept | -2.13 | 0.28 | -7.68 | <.005 |
| Vowel | 1.24 | 0.38 | 3.30 | <.005 |
| Perception | 0.00 | 0.28 | 0.02 | 0.99 |
| Vocabulary | 0.97 | 0.24 | 4.14 | <.005 |
| Perception x Vocabulary | 0.56 | 0.25 | 2.27 | < .05 |

**Table 3.** Results of the GLMM on the effects of Vowel, Perception, Vocabulary, and the Perception x Vocabulary interaction on LDT nonword rejection accuracy.

The model revealed significant effects of Vowel and Vocabulary as well as a significant Perception x Vocabulary interaction. Hence, participants were significantly more accurate in rejecting /ɛ/ nonwords than /æ/ nonwords and, crucially, having a larger vocabulary size led to higher accuracy in the LDT. Interestingly, the effect of vocabulary was modulated by an additive interaction with perception such that the effect of vocabulary size was larger for participants with better perception abilities. This interaction is illustrated in Figure 3.

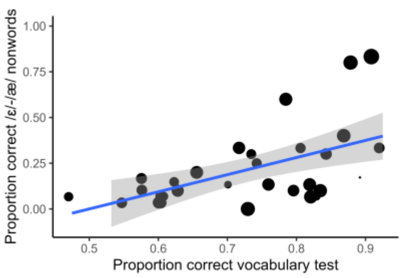


**Figure 3**: Scatterplot of the relationship between vocabulary size and /ɛ/-/æ/ nonword rejection in the LDT. Dot size conveys perceptual accuracy, where a larger dot means a steeper categorization slope.

## 4. DISCUSSION

In this paper, we aimed at a conceptual replication of [5] with an L2 learner group whose native language (i.e., Korean) is orthographically and lexically less similar to English than German. The main question we asked is whether an effect of L2 vocabulary size on the lexical encoding of two sounds that constitute a difficult L2 contrast could also be found for this new population. Results showed that, indeed, scores in an English vocabulary test predicted Korean learners' performance in a lexical decision task assessing their lexical encoding of English /ɛ/ and /æ/. In addition, this effect was qualified by an additive interaction with perception, indicating that the effect of vocabulary size was larger for learners with better perceptual abilities for the target contrast.

The across-the-board effect of vocabulary size therefore replicates the findings in [5] and [6] and suggests that it is not only when L1 and L2 are highly similar in terms of the orthographic and phonological forms of their words that knowledge of L2 vocabulary plays a major role in the development of phonologically-robust L2 lexical representations. In fact, a coarse comparison by means of Pearson product-moment correlations between vocabulary scores and LDT nonword rejection scores reveals an *r* of .55 for the Korean learners of English here and one of .39 for the whole sample of native German speakers in [5]. This indicates that a lower L1-L2 lexical similarity does not entail a looser relationship between vocabulary and lexical encoding. If anything, the coefficients appear to suggest the opposite. Hence, our results for Korean learners indirectly provide evidence against the possibility that an effect of L2 vocabulary is just a by-product of how existing vocabulary in the native language mediates the acquisition of L2 lexical items.

The second key finding of this study is the significant interaction between vocabulary and perception (see Figure 3). This interaction points towards the idea that a larger vocabulary size in the L2 is particularly beneficial when accompanied by sufficient ability in perceptual categorization for the target contrast. These results fit nicely with the proposal in [5] that learners need to reach a threshold of perceptual acuity (what could perhaps be referred to as "*good enough perception*") before they can leverage their L2 lexical knowledge and undergo substantial improvements in their lexical encoding.

In sum, the present study showed that L2 vocabulary predicts how learners phonologically encode challenging L2 categories into lexical representations even when the L1 and L2 of the learners are much more dissimilar than in previous studies on the same topic. This starkly challenges the concern that this relationship may simply surface because it is overwhelmingly mediated by the relationship between L1 and L2 words in terms of orthography and phonology, and strengthens the notion that L2 vocabulary is indeed a key predictor of essential improvements in the phonological form of L2 words containing challenging L2 sounds.

## 5. REFERENCES


[1] Darcy, I., Daidone, D., Kojima, C. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *The mental lexicon* 8(3), 372-420.

[2] Llompart, M., Reinisch, E. 2021. Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics* 88, 101080.

[3] Barrios, S., Hayes-Harb, R. 2021. L2 processing of words containing English /æ/-/ɛ/ and /l/-/ɹ/ contrasts, and the uses and limits of the auditory lexical decision task for understanding the locus of difficulty. *Frontiers in Communication* 6, 144.

[4] Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., Scott, J. H. 2012. Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research* 28(1), 5-40.

[5] Llompart, M. 2021. Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition* 24(3), 481-496.

[6] Daidone, D., Darcy, I. 2021. Vocabulary size is a key factor in predicting second-language lexical encoding accuracy. *Frontiers in psychology* 12, 2769.

[7] "Similar languages," ezglot.com. https://www.ezglot.com/most-similar-languages.php (accessed December 15, 2022)

[8] Escudero, P., Hayes-Harb, R., Mitterer, H. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36(2), 345-360.

[9] Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J. 2019. PsychoPy2: experiments in behavior made easy. *Behavior Research Methods*. 10.3758/s13428-018-01193-y

[10] Kawahara, H., Masuda-Katsuse, I., De Cheveigne, A. 1999. Restructuring speech representations using a pitch-adaptive time–frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech communication* 27(3-4), 187-207.

[11] Shipley W. C., Gruber, C.P., Martin, T.A., Klein, A.M. 2009. *Shipley-2 manual*. Western Psychological Services.


---

[i] The dataset and the code to reproduce the main analyses of this paper are available at https://osf.io/e637k/ (Open Science Framework)