A REASSESSMENT OF THE EFFECTS OF NEIGHBORHOOD DENSITY AND PHONOTACTIC PROBABILITY ON L2 ENGLISH WORD LEARNING

Tsung-Ying Chen, Yi-Sin Sie

National Tsing Hua University, Taiwan
chen.ty@mx.nthu.edu.tw, yisinsie@gmail.com

ABSTRACT

Phonotactic probability (PP) and neighborhood density (ND) were found to significantly affect L1/L2 (non-)word processing and learning. This study sets out to reassess their effects in the context of L2 word learning at the recognition level with a regression-based experimental design. Twenty-seven L1 Mandarin speakers learning L2 English were recruited to learn one hundred Mandarin (text)-English (audio) pairs and tested in two separate sessions. Results were analyzed using mixed-effects logistic regression with independent variables including L1/L2 PP and L1/L2 ND and L1 translations token frequency. L2 PP, L2 ND, and L1 frequency were found to negatively correlated with the accuracy within each test session as well as the consistency in accuracy across the two test sessions. These findings help shed lights on the nature of lexical effects in word processing/learning and the emergence of associations between L1 and L2 lexical entries.

Keywords: phonotactic probability, neighborhood density, token frequency, word learning, L2, Mandarin, English

Acknowledgements: We thank James Myers, Fabian Tomaschek, and Richard Wright for their valuable comments on this preliminary work.

1. INTRODUCTION

The processing of a lexical auditory form is influenced by the transitional probability of sounds in that form and the phonological similarity between that form and other lexical entries. The former is commonly referred to as phonotactic probability (PP) and the latter as neighborhood density (ND; e.g., number of phonological neighbors different in one single segment).

High-PP words were found to be recognized more rapidly and accurately than were low-PP words (Vitevitch and Luce [16, 17]). Storkel et al. [13] attributed these findings to frequently activated common sound paths that lead to a faster retrieval of high-PP lexical representations. Levelt and Wheeldon [6] and Levelt et al. [5], on the other hand, proposed a sublexical syllabary storing frequent sound combinations that facilitate speech planning of high-PP words.

Low-ND words are in general recognized more easily than high-ND words (Luce [8] and Luce and Pisoni [9]), as it could be more difficult to retrieve high-ND words when more lexical competitors are activated (e.g., Marslen-Wilson [10]). Word production nevertheless benefits from a denser neighborhood as a high-ND word is more likely than a low-ND word to be activated by its phonological neighbors before its lexical retrieval for production (e.g., Vitevitch [15]).

In both child and adult L1/L2 word learning, PP and ND were found to play distinctive roles, too. In Storkel and Hoover [14] and Storkel et al. [13], learners demonstrated an advantage in learning low-PP words in a story-telling context. This is presumably because uncommon sound combinations are more likely to be detected from the continuous speech. Experimental results from Jones [4] and Stamer and Vitevitch [12] suggest a high-ND advantage in L1/L2 word learning; if newly created lexical representations have many lexical neighbors, they could be activated more frequently and therefore more consolidated in the long run.

There are, however, theoretical and methodological reasons to reassess the PP and ND effects on L2 word learning. First, since bilingual lexical access is non-selective (e.g., de Groot et al. [3]), the studies of PP and ND effects in L2 word learning must also take L1 PP and ND into consideration. In addition, the non-selective nature could make it easier to develop new L2 lexical representations by linking them to their L1 counterpart (e.g., Lotto and de Groot [7]). Therefore, in this study we tested how L1 Mandarin EFL learners acquire L2 English auditory words paired with their L1 Mandarin gloss, and we factored in the token frequency of L1 translations to show its relationship with the L1/L2 PP and ND effects. Second, the ND effects on L1/L2 word learning in previous studies were explored at the production level either in a picture naming task or in a word repetition task, which indicated a high-ND advantage. However, since low-ND words are recognized more easily, we expect low-ND L2 words to be acquired better than high-ND ones, and the current study was designed to test this hypothesis. Finally, previous studies of L1/L2 word learning usually exposed their child and adult learners to a small set of novel words, which forced their analyses to dichotomize lexical measures into categorical variables, leading to a possible loss of statistical generalizations associated with intrinsically continuous variables (e.g., Baayen [1]; Cohen [2]). The current study adopted a regression-based experimental design to sidestep the above issue.

Our main research questions can be summarized as follows:

- What are the effects of L2 PP/ND of L2 novel words on L2 English word learning?
- What are the effects of L1 PP/ND of L2 novel words on L2 English word learning?
- What are the effects of L1 Mandarin gloss token frequency on L2 English word learning?

2. WORD LEARNING EXPERIMENT

Our word learning experiment includes a training phase and two test phases, in which L1 Mandarin EFL learners were required to learn 100 associations between L2 English auditory forms and their L1 Mandarin gloss and tested with a form-meaning mapping task.

2.1. Materials

The one hundred target L2 English concrete nouns were selected from a low-frequency subset of the English CELEX corpus with a COBUILD frequency ranged from 1 (N = 74) to 55 tokens per million words. Each target L2 English word was also paired with another low-frequency English word sharing initial phonemes as a competitor in the test session (e.g., canary vs. canard). Among the one hundred competitors, 54 had a COBUILD frequency of 1 token per million words. Therefore, the target words and their competitors were unlikely known by our participants prior to the experiment. The auditory form of target words and competitors were retrieved from the Web version of the *Oxford Advanced Learner's Dictionary*. Silent sections were removed, and the volume was normalized.

The L1 Mandarin translation for each target word was obtained from Google Translate, which was validated by the authors speaking L1 Mandarin as their L1. L1 Mandarin translations longer than three characters were modified to have only two or three characters to minimize the word length effect on the learning of the target word pairs.

2.2. Procedure

All participants began with the training session presenting the one hundred L1-L2 word pairs in random order. On each training trial, the auditory form of the target word was presented via a headphone, and its L1 Mandarin gloss was presented at the center of a laptop screen simultaneously. Each trial lasted for four seconds before the training session moved to the next trial automatically. The one hundred pairs were repeated three times during the training session to familiarize our participants with the learning materials.

Following the training session were two test sessions administered on two separate days to test the memory consolidation effect and the consistency in participants' responses. In both test sessions, each trial began with an eye-fixation cross appeared at the center of a laptop screen for 500 ms. The target and its competitor were then presented auditorily via a headphone in random order on each trial. Immediately after the end of the second auditory input, the L1 Mandarin gloss appeared at the center of the screen in Traditional Chinese. The participants were instructed to match the first or the second auditory form to the L1 gloss by pressing the 'S' key (first) or the 'L' key (second) as quickly as possible. The experiment moved to the next trial following a valid response or after four seconds without any response.

After both test sessions were completed, we administered the Boston Naming Test to estimate the L2 vocabulary size of each participant.

2.3. Participants

Twenty-seven adult native speakers of Taiwan Mandarin learning L2 English were recruited to the current study. The first four participants were excluded from the current study due to technical issues that interrupted their training or test sessions. Another participant was excluded as the only participant completing the two test sessions with an interval longer than 48 hours. Among the remaining twenty-two participants, fifteen were males and seven were females. Their age ranged from 21 to 29 years old with a mean of 26.1 years old (sd = 2.5) and their sleep duration between the two test sessions fell into a range between 3.5 to 10 hours (mean = 7.9, sd = 1.6). Their performance in the Boston Naming Test suggests a group mixed with beginner and intermediate L2 English learners (range = 5-35 points, mean = 18.3, sd = 8.6). No participant reported any learning or language impairment.

3. RESULTS

3.1. Accuracy

Binomial tests suggest an above-chance accuracy in both test sessions (first: 71.6%, $p < .001$; second: 68.6%, $p < .001$). It is thus safe to conclude that our participants were engaged in the word learning task.

Our first main focus is on the effect of lexical measures and test session on response accuracy. PP was quantified as the mean conditional bigram probability and ND as PLD20 (mean edit distance from the twenty nearest neighbors). L2 PP and ND were calculated with the English CELEX corpus. L1 PP and ND were calculated with the CC-CEDIT dictionary ignoring tones. L1 Mandarin gloss token frequencies were extracted from the Academia Sinica Balanced Corpus of Modern Chinese. The ND measures and the L1 token frequencies were log-transformed, and all lexical measures were z-scored.

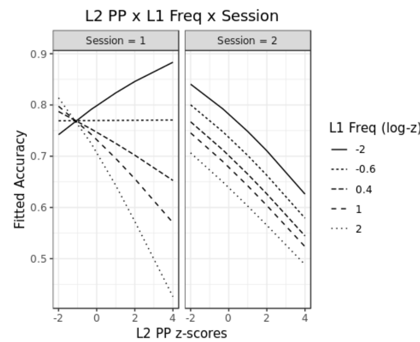
In our mixed-effects logistic regression, the binary response accuracy ('correct' vs. 'incorrect') served as the dependent variable, which was regressed against all lexical measures and two categorical variables session ('first' vs. 'second') and target-competitor order ('target first' vs. 'target last'). The interactions between L1 measures, L2 measures, and categorical variables were also tested. The random intercepts of Participant and Item were included. The model was then simplified by removing insignificant predictors L1 PP and ND for a better data fit.

Table 1: Summary of main effects excluding interactions in mixed-effects logistic regression on response accuracy; * = $p < .05$

	β	SE	z	p
Intercept	0.743	0.131	5.68	< .001*
L2 ND	0.4	0.099	4.04	< .001*
L1 Freq	-0.207	0.073	-1.99	.004*
L2 PP × L1 Freq × Session	0.124	0.065	1.89	.059

The predictors and interaction of interest are summarized in Table 1. Crucially, both a sparser L2 ND (high PLD20 = most distant from phonological neighbors) and a higher L1 token frequency predict a significantly lower accuracy rate. L2 PP, L1 token frequency, and session marginally interact with each other as visualized in Figure 2. In Session 1, a high L2 PP and a low L1 token frequency give rise to a higher accuracy. Otherwise, participants in general benefited from a low L2 PP regardless of session and L1 token frequency.

Figure 1: Marginal interaction between L2 PP, L1 token frequency, and session



3.2. Consistency

Our second analysis concentrates on the consistency in response accuracy across the two test sessions. Word learning is considered successful only if learners could choose the correct target L2 auditory form from the same target-competitor pair in both test sessions.

The consistency served as the dependent variable coded with a binary contrast between 'consistently correct' and 'others', which was regressed against L1 and L2 lexical measures in mixed-effects logistic regression. Session and target-competitor order were excluded since it was a cross-session analysis per se, and target-competitor order was largely inconsistent across the two sessions. All interactions between L1 and L2 measures were included with participant and item as random intercepts. L1 PP and ND were again removed for being insignificant predictors to better fit the data.

Table 2: Summary of main effects excluding interactions in mixed-effects logistic regression on the consistency in response accuracy; * = $p < .05$

	β	SE	z	p
Intercept	0.187	0.155	1.21	.228
L2 PP	-0.203	0.102	-1.99	.046*
L2 ND	0.336	0.102	3.3	< .001*
L1 Freq	-0.245	0.101	-2.43	.015*

The model summary suggests a significant effect of L2 PP, L2 ND, and L1 token frequency, which are summarized in Table 2. A higher L2 PP results in a lower consistency, a sparser neighborhood leads to a higher consistency, and a higher L1 gloss token frequency negatively influences the consistency in correct responses.

4. DISCUSSION

There are four major findings in the current study. First, our study successfully replicated the low-PP advantage found in previous L1/L2 word processing and learning studies. Second, we found an expected low-ND advantage with a task that required L2 English learners to recognize target L2 auditory word forms and match them to their L1 gloss. Third, L1 Mandarin PP and ND did not interfere with the L2 English word learning. Finally, a low L1 Mandarin gloss token frequency facilitated the development of the associations between L2 English auditory forms and its L1 translation.

While L2 PP and ND played their respective roles in L2 English word learning, the difference in their effect size is worth noting; L2 ND was a strong predictor in both accuracy and consistency analyses, whereas the significant L2 PP effect only emerged in the latter and just reached the significance at the level of $\alpha = .05$. One possibility of the weaker PP effect is that all L2 English target words were *new* to our participants, who could not take advantage of novel word *detection*. An alternative explanation would be L1 Mandarin speakers' lower sensitivity to holistic phonological similarities than to phonotactic information (see Myers [11] and references cited therein).

The strong negative correlation between L1 gloss token frequency and response accuracy and consistency might seem counterintuitive given a well-established effect that high-frequency words are retrieved more easily. Nevertheless, high-frequency L1 lexical entries could be strongly associated to their L1 phonological forms as well as other existing L2 phonological forms, which inhibit the emergence of new L1-L2 links.

5. CONCLUSION

The current study re-examined the effects of PP and ND in the context of L2 English word learning to not only better understand the nature of the lexical effects on word processing but also their interactions with other lexical factors involved in the development of a bilingual lexicon driven. Our preliminary results suggest that while the PP and ND effects seem robust, conclusions cannot be made without considering L1 lexical factors that could potentially contribute to the L2 word learning process as well.

REFERENCES

1. Baayen, R. H. (2010). A real experiment is a factorial experiment? *The Mental Lexicon*, *5*(1), 149-157.
2. Cohen, J. (1983). The Cost of Dichotomization. *Applied Psychological Measurement*, *7*(3), 249-253.
3. de Groot, A. M. B., Delmaar, P., & Lupker, S. J. (2000). The Processing of Interlexical Homographs in Translation Recognition and Lexical Decision: Support for Non-Selective Access to Bilingual Memory. *The Quarterly Journal of Experimental Psychology Section A*, *53*(2), 397-428.
4. Jones, S. E. (2018). Adult Word Learning as a Function of Neighborhood Density. *Languages*, *3*(1), 5.
5. Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-38.
6. Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*(1-3), 239-269.
7. 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*(1), 31-69.
8. Luce, P. A. (1986). *Neighborhoods of Words in the Mental Lexicon. Research on Speech Perception*. Technical Report No. 6. Bloomington, IN.
9. Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and Hearing*, *19*(1), 1-36.
10. Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1-2), 71-102.
11. Myers, J. (2016). Meta-megastudies. *The Mental Lexicon*, 11(3), 329–349.
12. Stamer, M. K., & Vitevitch, M. S. (2012). Phonological similarity influences word learning in adults learning Spanish as a foreign language. *Bilingualism: Language and Cognition*, *15*(3), 490-502.
13. 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*(6), 1175-1192.
14. Storkel, H. L., & Hoover, J. R. (2010). Word learning by children with phonological delays: Differentiating effects of phonotactic probability and neighborhood density. *Journal of Communication Disorders*, *43*(2), 105-119.
15. Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 735-747.
16. Vitevitch, M. S., & Luce, P. A. (1998). When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, *9*(4), 325-329.
17. Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic Phonotactics and Neighborhood Activation in Spoken Word Recognition. *Journal of Memory and Language*, *40*(3), 374-408.