FUZZY LEXICAL REPRESENTATIONS IN THE NONNATIVE MENTAL LEXICON

EDITED BY: Kira Gor, Denisa Bordag, Anna Chrabaszcz, Svetlana V. Cook
and Andreas Opitz
PUBLISHED IN: Frontiers in Psychology and Frontiers in Communication

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-413-0
DOI 10.3389/978-2-83250-413-0

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

FUZZY LEXICAL REPRESENTATIONS IN THE NONNATIVE MENTAL LEXICON

Topic Editors:
Kira Gor, University of Maryland, College Park, United States
Denisa Bordag, Leipzig University, Germany
Anna Chrabaszcz, University of Pittsburgh, United States
Svetlana V. Cook, University of Maryland, College Park, United States
Andreas Opitz, Leipzig University, Germany

Citation: Gor, K., Bordag, D., Chrabaszcz, A., Cook, S. V., Opitz, A., eds. (2022). Fuzzy Lexical Representations in the Nonnative Mental Lexicon. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-413-0

Table of Contents

TYPE Editorial
PUBLISHED 21 September 2022
DOI 10.3389/fcomm.2022.1027692

OPEN ACCESS

EDITED AND REVIEWED BY
Xiaolin Zhou,
Peking University, China

*CORRESPONDENCE
Kira Gor
kiragor@umd.edu

SPECIALTY SECTION
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

RECEIVED 25 August 2022
ACCEPTED 29 August 2022
PUBLISHED 21 September 2022

CITATION
Gor K, Cook S, Bordag D, Chrabaszcz A and Opitz A (2022) Editorial: Fuzzy lexical representations in the nonnative mental lexicon. *Front. Commun.* 7:1027692. doi: 10.3389/fcomm.2022.1027692

Editorial: Fuzzy lexical representations in the nonnative mental lexicon

Kira Gor[1]*, Svetlana Cook[2], Denisa Bordag[3,4], Anna Chrabaszcz[5,6] and Andreas Opitz[3]

[1]Graduate Program in Second Language Acquisition, School of Languages, Literatures, and Cultures, University of Maryland, College Park, MD, United States, [2]National Foreign Language Center, University of Maryland, College Park, MD, United States, [3]Faculty of Philology, Herder Institute, Leipzig University, Leipzig, Germany, [4]University of Haifa, Haifa, Israel, [5]Center for Language and Brain, HSE University, Moscow, Russia, [6]Department of Psychology, Kenneth P. Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA, United States

KEYWORDS

L2, nonnative, fuzzy, lexical representation, word recognition, lexicon

Editorial on the Research Topic
Fuzzy lexical representations in the nonnative mental lexicon

The call for contributions to the Research Topic "*Fuzzy lexical representations in the nonnative mental lexicon*" proposed that fuzziness, i.e., imprecise encoding of the word form and/or meaning that may lead to weak lexical competition and lexical confusion, is a pervasive property of the nonnative (L2) lexicon. The 12 publications that have appeared in the Research Topic have addressed and developed different aspects of fuzzy lexical representations (FLRs).

The article by Gor et al. introduces the fuzzy lexical representations hypothesis (FLRH) and identifies the causes of fuzziness in L2 LRs and its consequences for L2 word recognition and lexical processing, in general. It reviews a number of studies that show how imprecise and ambiguous (i.e., fuzzy) phonolexical encoding, or phonological encoding of lexical units leads to weak form-meaning connections and sometimes incorrect form-meaning mappings. As a result, fuzziness produces spurious semantic associations in L2 and weakened lexical competition. According to FLRH, if fuzziness applies to the phonological encoding of difficult L2 contrasts, it may never get resolved even despite significant amounts of input. Furthermore, phonolexical encoding of less familiar L2 words may be fuzzy even if no particularly problematic L2 phonemes or phonological contrasts are involved. This type of initial fuzziness may decrease with L2 input and growing proficiency. The article suggests that L2 models of auditory word recognition should account for the interaction of input with FLRs to be more realistic. Another theoretical contribution by Kapnoula discusses the sources of lexical inhibition and facilitation in lexical competition, as they are represented in L1 and L2-based models of lexical access. This paper argues that the L1-based approach focusing on inter-lexical connections (i.e., inhibition arising from lexical competition) and the FLR approach focusing on intra-lexical fuzziness are compatible with each other.

Three contributions address different aspects of fuzziness in L2 phonolexical encoding. The study by Llompart focuses on asymmetries in perceptually confusable L2 phonemes—both in accuracy and in the contribution of lexical and acoustic-phonetic factors—in auditory word recognition using a lexical decision task (LDT). The encoding of the dominant L2 category that has a counterpart in L1 is not strongly constrained by lexical frequency or phonological neighborhood density; at the same time, it has more clearly defined (i.e., less fuzzy) acoustic-phonetic properties. In contrast, the encoding of the non-dominant category is influenced by the lexical factors. Another study by Daidone and Darcy uses a LDT to investigate the role of individual differences in the quality of L2 phonolexical encoding with a set of nonnative phonological contrasts. The study reveals that independently measured vocabulary size of individual participants is by far the best predictor of accuracy in the LDT—the bigger the vocabulary size, the higher the accuracy, i.e., the smaller the fuzziness effects on L2 lexical encoding. Phonological short-term memory is another factor that affects the quality of phonological encoding of two L2 phonemes that do not have counterparts in L1.The article by Barrios and Hayes-Harb explores the possibilities and limitations of a LDT in establishing the locus of learner difficulty in lexical processing. It contrasts two cases—when the locus of learners' difficulty is phonological encoding vs. phonolexical encoding. Specifically, the study looks at the scenarios when the phonological contrast in question is neutralized, the non-dominant phoneme is substituted by the dominant (or native) phoneme, or the non-dominant phoneme is distinct from the dominant, but is yet non-target-like. This approach offers concrete ways to operationalize the varying types of phonological and phonolexical fuzziness.

Two articles broaden the scope of FLRH by addressing the L2 lexical encoding that goes beyond phonemic contrasts. The contribution by Pelzl et al. extends the predictions of FLRH to the domain of suprasegmental features. Using a picture-phonology matching task with concurrent event-related potentials (ERP) recordings, the authors examine whether L2 learners experience encoding or retrieval problems with lexical tones in Mandarin Chinese. The results demonstrate that L2 tone recognition falls under the umbrella of fuzziness in that sometimes despite having correct metalinguistic knowledge of tones, L2 listeners nevertheless experience difficulties with tone recognition. The authors argue that these differences reflect problems both at the encoding and the retrieval levels. In another study, Frederiksen investigates placement descriptions (e.g., "she put the cup on the table") of hearing L2 learners of American Sign Language (ASL). Results suggest that fuzzy semantic boundaries occur in cross-modal L2 acquisition as well, as indicated by the observation that L2 signers use a wider range of handshapes and use them less appropriately than native ASL controls. However, L2 signers' placement descriptions look much more similar to those of native ASL controls than expected based on the findings for written and spoken modalities. The high degree of iconicity and transparency of placement distinctions in the visual modality may facilitate L2 acquisition and reduce fuzziness.

The study by Baxter et al. addresses the crucial question of how to improve the quality of lexical encoding of novel L2 words, i.e., to reduce their fuzziness. An L2 vocabulary training experiment with children aged 9–10 demonstrates that orthographic and semantic confusability of novel words is reduced when orthographically and semantically similar words are contrasted during word learning. While the learning challenge in contrastive training is increased, the outcomes of contrastive training are improved compared to non-contrastive training.

Three articles are devoted to the grammatical encoding of L2 words. The contribution by Bordag and Opitz extends the construct of fuzziness to the encoding of grammatical properties of lexical representations. The study compares the reading speed for newly introduced lexical items to test how L1 and L2 readers establish mental representations of new words that belong to different grammatical categories while having the same orthographic form. As predicted, new lexical representations in L2 encode the grammatical information less precisely than in L1. In an eye-tracking-in-reading study, Nakamura et al. examine strategies in the online use of information about a verb's argument structure by L2 learners. They report that both L2 participants and L1 controls access subcategorization information to guide initial syntactic analysis. While the authors argue that their results support the hypothesis of a more general parsing strategy in both L1 and L2, i.e., the intransitivity overriding hypothesis, their results do not allow for more precise conclusions about the quality of lexical representations in L1 vs. L2. Whether the quality of lexical representations differs in L1 and L2 with respect to syntactic information, e.g., argument structure, remains one of the areas to be explored in future research. The visual priming study with present and past-tense verb forms by Wanner-Kawahara et al. provides evidence that L2 learners can develop connections in terms of morphological relationships that are similar to those of L1 while having fuzzy representations of L2 orthographic forms.

Finally, the article by Zhao and Li compares the results of a computational simulation model of a developing bilingual lexicon with an analysis of error patterns in real second language learners to establish the role of the age of onset of L2 learning in developing lexical representations. It shows that the early learning of L2 compared to L1 leads to the establishment of functionally distinct lexical representations. Conversely, when the learning of L2 occurs significantly later compared to L1, fuzzy L2 representations may be established. Because of the L1 structural primacy, L2 lexical units are forced to compete for space, and as a result, form dense overlapping representations with fuzzy boundaries between the entries.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

The work on this manuscript by KG was supported by the funding from the Slavic and East European Language Resource Center (SEELRC) at Duke University.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 May 2021
doi: 10.3389/fpsyg.2021.659852

Lexical and Phonetic Influences on the Phonolexical Encoding of Difficult Second-Language Contrasts: Insights From Nonword Rejection

*Miquel Llompart**

Chair of Language and Cognition, Department of English and American Studies, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany

Establishing phonologically robust lexical representations in a second language (L2) is challenging, and even more so for words containing phones in phonological contrasts that are not part of the native language. This study presents a series of additional analyses of lexical decision data assessing the phonolexical encoding of English /ɛ/ and /æ/ by German learners of English (/æ/ does not exist in German) in order to examine the influence of lexical frequency, phonological neighborhood density and the acoustics of the particular vowels on learners' ability to reject nonwords differing from real words in the confusable L2 phones only (e.g., *l[æ]mon, *dr[ɛ]gon). Results showed that both the lexical properties of the target items and the acoustics of the critical vowels affected nonword rejection, albeit differently for items with /æ/ → [ɛ] and /ɛ/ → [æ] mispronunciations: For the former, lower lexical frequencies and higher neighborhood densities led to more accurate performance. For the latter, it was only the acoustics of the vowel (i.e., how distinctly [æ]-like the mispronunciation was) that had a significant impact on learners' accuracy. This suggests that the encoding of /ɛ/ and /æ/ may not only be asymmetric in that /ɛ/ is generally more robustly represented in the lexicon than /æ/, as previously reported, but also in the way in which this encoding takes place. Mainly, the encoding of /æ/ appears to be more dependent on the characteristics of the L2 vocabulary and on one's experience with the L2 than that of its more dominant counterpart (/ɛ/).

Keywords: second language learning, lexical representation, speech perception, L2 lexicon, phonolexical encoding, lexical decision, nonword rejection, L1-accented input

Edited by:
Kira Gor,
University of Maryland, College Park, United States

Reviewed by:
Rachel Hayes-Harb,
The University of Utah, United States
Isabelle Darcy,
Indiana University Bloomington, United States

Correspondence:
Miquel Llompart
llompart.garcia@fau.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *28 January 2021*
Accepted: *27 April 2021*
Published: *31 May 2021*

Citation:
Llompart M (2021) Lexical and Phonetic Influences on the Phonolexical Encoding of Difficult Second-Language Contrasts: Insights From Nonword Rejection. Front. Psychol. 12:659852. doi: 10.3389/fpsyg.2021.659852

INTRODUCTION

A crucial part of second language (L2) learning is building a non-native lexicon. This can be a very challenging endeavor, especially when the L2 is learned later in life and in a non-immersion setting, as is the case for many learners of English around the world (e.g., Díaz et al., 2012). For this type of learners, much of the learning takes place in a formal instruction setting (i.e., classroom) and the L2 is rarely spoken outside of that environment. This rather constrained interaction with the

L2 has apparent negative consequences on the acquisition of non-native lexical items. First, the relatively impoverished input translates into reduced exposure to individual L2 words, which often prevents their robust integration into long-term memory (Gollan et al., 2008) and almost invariably results in smaller vocabulary sizes in the L2 when compared with the native language (L1; Nation, 2006). Secondly, for words that become part of the L2 lexicon, the scarcity of L2 input results in the newly established lexical representations being phonologically vague or "fuzzy" (Cook and Gor, 2015; Cook et al., 2016; Lancaster and Gor, 2016). This means that the encoding of phonetic categories into lexical representations (i.e., phonolexical encoding) is not as robust as that of native lexical items, which greatly contributes to L2 spoken word recognition being rather error-prone and characterized by spurious lexical competition (e.g., Weber and Cutler, 2004; Cook et al., 2016).

An additional obstacle for the establishment of robust L2 lexical representations is that learners are bound to face difficulties while trying to master the phonology of the non-native language. In particular, L2 phonological contrasts that are not part of the L1 are very often the source of perceptual difficulties. This is the case, for example, of the English distinction between /r/ and /l/ for native speakers of Japanese (Goto, 1971; Bradlow et al., 1999) and the vowel contrast between /ɛ/ and /æ/ for L1-German learners of English (Llompart and Reinisch, 2017, 2019a, 2020; Eger and Reinisch, 2019a,b), which is the object of the present study. Both /r/-/l/ and /ɛ/-/æ/ are instances of what Best and Tyler (2007) labeled as single-category assimilations in their model of L2 phonology learning; that is, a scenario in which two L2 phones are perceived as being perceptually close to one and the same L1 phone. It has been repeatedly shown that perceptual difficulties with L2 contrasts in single-category-assimilation relationships lead to representational imprecisions for words containing these contrasts (e.g., Broersma, 2012; Llompart and Reinisch, 2019b). Importantly, these imprecisions are long-lasting in that they appear to be in place even after L2 speakers have already learned to perceive the phonetic differences between the L2 phones (Díaz et al., 2012; Darcy et al., 2013; Amengual, 2016; Llompart, 2021). For example, Llompart (2021) provided evidence of a weak encoding of the /ɛ/-/æ/ contrast into English words even by German learners of English who had had extensive experience with the L2 and were able to distinguish between the two vowels in a phonetic identification task.

A task that has recurrently been used to assess the phonological robustness of lexical representations in late L2 learners is lexical decision involving real words and "mispronounced" nonwords. In such a task, words of the L2 are auditorily presented either in their canonical form or containing systematic phonological substitutions that transform them into nonwords. Participants are then asked to decide whether the items presented are real words in the L2 (Díaz et al., 2012; Darcy et al., 2013; Darcy and Thomas, 2019; Llompart and Reinisch, 2019b; Melnik and Peperkamp, 2019, 2021). Lexical decision tasks of this type have helped shed light on several issues concerning the phonolexical encoding of challenging L2 contrasts. First, lexical decision data have served to support the finding of previous visual-world eye tracking studies (Weber and Cutler, 2004; Cutler et al., 2006) that the encoding of these challenging contrasts is asymmetric and modulated by the goodness-of-fit of the L2 categories to the closest L1 category. As discussed by Cutler et al. (2006), the better-fitting L2 category in the contrast (i.e., more similar to the L1 category) is thought to be dominant and more robustly encoded into the corresponding L2 words than the worse-fitting alternative, whose encoding is generally less precise. For /ɛ/-/æ/, /ɛ/ is attributed this dominant role, to the point that /æ/ has been re-labeled in previous studies as not-/ɛ/ or *'/ɛ/ to emphasize its weaker phonolexical encoding (Llompart and Reinisch, 2017; see also Hayes-Harb and Masuda, 2008). Lexical decision data from Dutch (Simon et al., 2014) and German learners of English (Llompart and Reinisch, 2019b; Llompart, 2021) has contributed to characterizing this asymmetry by showing that learners are more sensitive to vowel substitutions when the target vowel should be /ɛ/ (e.g., *l[æ]mon) than in contexts in which it should be /æ/ (e.g., *dr[ɛ]gon). Secondly, recent research using this paradigm has examined the role that individual differences within the learner population may play with regard to phonolexical encoding. Here findings suggest that a more robust encoding of difficult L2 contrasts relates to learners' phonetic categorization ability for that particular contrast (Silbert et al., 2015; Simonchyk and Darcy, 2017; Darcy and Holliday, 2019) as well as to their L2 vocabulary size (Daidone, 2020; Llompart, 2021).

What has not received much attention in this particular body of literature, however, is the role that item-specific properties, both lexical and phonetic, may play on learners' ability to accept real words containing confusable L2 phones and successfully reject nonwords that differ from real words in those particular phones. This is the case even though an examination of such properties could be crucial for our understanding of the influence that lexical factors may have on the phonolexical encoding of phones in challenging L2 contrasts, an issue that is not well-understood yet. As a first step in this direction, the present study presents a series of additional analyses on lexical decision data from German learners of English (as reported in Llompart and Reinisch, 2019b, and Llompart, 2021) aimed at assessing the effects of L2 words' lexical frequency, phonological neighborhood density and the acoustics of the critical vowel on learners' ability to reject nonwords containing /ɛ/-/æ/ mispronunciations (e.g., *l[æ]mon, *dr[ɛ]gon). While the role that these factors may play with respect to accuracy in real word acceptance is also a question of theoretical interest, real word acceptance rates were not assessed in this study because of learners' ceiling performances with real /ɛ/- and /æ/-words in Llompart and Reinisch (2019b) and Llompart (2021)[1].

For responses to mispronounced nonwords, it is in principle expected that both the lexical properties of the items presented and the acoustic image of the auditory stimuli corresponding to these items should influence learners' lexicality decisions.

[1]To illustrate the ceiling effects for real word acceptance, percentages of correct responses to real word stimuli by vowel (/ɛ/-words and /æ/-words) for the different groups tested in Llompart and Reinisch (2019b) and Llompart (2021) are provided as **Supplementary Table 1**. In addition, **Supplementary Table 2** outlines the results of a model with the exact same structure as the base model reported in the "Results" section but run on responses to real words instead of mispronounced nonwords. Unlike the model in the Results section, the model for real words does not show significant differences in accuracy as a function of vowel or learner group.

Concerning lexical frequency, nonwords built on high-frequency words could be expected to be harder to reject than nonwords based on lower-frequency words. Lexical decision tasks as the ones described above rely on the well-documented Ganong effect (Ganong, 1980) to bias participants' responses toward considering the stimuli "real words." Ganong (1980) created a stimulus that was ambiguous between /t/ and /d/, appended the same stimulus to *-ask* and *-ash* and asked listeners to categorize the initial phone as /t/ or /d/ in each context. Listeners were found to be more likely to categorize it as /t/ in the ?ask context and as /d/ in the ?ash context, thus showing that lexical knowledge guides speech perception when the signal is acoustically ambiguous. Following from this, in the present experimental paradigm, in which listeners are presented with items like *dr[ɛ]gon and asked whether they are real words of English or not, they are expected to be more likely to answer "word" than "nonword" whenever acoustic information is not enough for them to be certain of the identity of the substituted or mispronounced phone. Crucially, this attraction toward "word" responses should be stronger the more frequently listeners have encountered the words that served as a base form for the nonwords in the L2 (Coltheart et al., 1977; Andrews, 1996; Perea et al., 2005; but see Politzer-Ahles et al., 2020).

Like lexical frequency, phonological neighborhood density is also known to have a major impact on lexical access, and more specifically, on lexicality decisions. However, for the task examined here, it is unclear whether high phonological neighborhood densities should aid or prevent the accurate rejection of mispronounced nonwords. On the one hand, given that higher phonological neighborhood density tends to hinder auditory word recognition by enhancing lexical competition (Luce and Pisoni, 1998; Vitevitch, 2002a,b), higher densities could be expected to bias listeners toward "word" responses for the nonwords in a similar way as high lexical frequencies should. On the other hand, one could alternatively predict that higher neighborhood densities may boost accurate nonword rejection. Higher phonological neighborhood densities should almost invariably mean larger clusters of similar-sounding words containing the same L2 target phone. Hence, especially for difficult L2 phonological contrasts, the existence of multiple phonologically similar word forms with a specific L2 category in the contrast (and not the other) may be beneficial for the establishment of robust links between the corresponding phonetic category and L2 word forms (see Llompart, 2021). Because of this, a scenario in which accuracy in nonword rejection increases as a function of the number of neighboring words with the same target category a given form has is also plausible.

Regarding the acoustic properties of the relevant L2 phones in the nonwords, lexical decision tasks generally use stimuli in which the mispronunciations were elicited naturally, and this is the case in Llompart and Reinisch (2019b) and Llompart (2021), the studies that provided the dataset to be analyzed here. By design, the use of naturally elicited stimuli means that the target phones must show some variation in their acoustics, most likely related to the surrounding phones (e.g., Strange et al., 2007) and to inherent within-speaker variation. Hence, a relevant question here, and one which has not yet been addressed, is how sensitive learners are to fine-grained acoustic variation in a task where they are mainly asked to focus on the lexicality of the stimuli. In principle, one could predict that, for nonwords with systematic vowel substitutions, the more acoustically distinctive the substitution, and thus the further from the canonical vowel the "mispronounced" vowel is, the easier it should be for learners to detect the mismatch and reject these nonwords. While this is a likely possibility for mispronunciations involving L2 phones not leading to perceptual difficulties, it is less clear that this should be the case for nonwords containing perceptually confusable non-native phones like /ɛ/ and /æ/ for German learners of English. Since the phonetic categories for these phones are most likely not as well-defined, learners may fail to use between-item variation in the acoustics of the stimuli as a cue to their judgments (Díaz et al., 2012; Llompart and Reinisch, 2019a).

Finally, it is worth noting that, in Llompart and Reinisch (2019b) and Llompart (2021), there were two different types of mispronounced nonwords for the L2 contrast of interest (/ɛ/-/æ/). These were items in which /ɛ/ was substituted by [æ] (e.g., *l[æ]mon) and items in which /æ/ was substituted by [ɛ] (e.g., *dr[ɛ]gon). Crucially, these two types differ in two important respects. The first is the difference in goodness of fit to the closest L1 category of the canonical vowel (i.e., /ɛ/ > /æ/) and the mispronounced vowel realizations (i.e., [ɛ] > [æ]), which, as was discussed above, is bound to have consequences on the perception and lexical encoding of these phones. The second key difference is that, for learners in a non-immersion setting, the two mispronunciation types are not equally likely to conform to their experience in their day-to-day L2-learning environment. Germans learning English in Germany are extremely likely to be exposed to instances in which /æ/ is produced with acoustic properties more closely aligning with /ɛ/ (e.g., h[ɛ]ppy, pl[ɛ]n, dr[ɛ]gon) in the speech of fellow learners and perhaps even English teachers (see Eger and Reinisch, 2019a; Llompart and Reinisch, 2019a for acoustic data), whereas the opposite pattern (e.g., st[æ]p, l[æ]mon) is very unlikely to occur. Therefore, these critical disparities warrant the additional question as to whether the effects of lexical frequency, phonological neighborhood density and vowel acoustics could differ between the two types of mispronounced nonwords examined.

MATERIALS AND METHODS

Participants

Data from 116 participants were included in all analyses presented in the Results section. Thirty-seven participants were the German learners of English (19 females, mean age = 25.32, *SD* = 4.37) included in the analyses of Llompart and Reinisch (2019b). These participants were students at the Ludwig Maximilian University of Munich (LMU) who had grown up in German monolingual households, had not spent more than 6 months in an English-speaking country and were not enrolled in courses administered by the English department of the university. The remaining 79 participants were the two groups of learners tested in Llompart (2021). The first group consisted of 49 German learners of English studying at the Friedrich Alexander University Erlangen-Nuremberg (FAU; 35 females, mean age = 24.22, *SD* = 4.26) who were recruited according to the same criteria as the previous group. The second group consisted of 30 English professionals and university students studying to become English professionals also recruited at FAU (17 females, mean age = 28.5, *SD* = 12.32). These were either language instructors at the university's Language Center (*N* = 5) or students enrolled in the BA and MA programs offered by the Department of English and American Studies (*N* = 25). In the present study, and following Llompart (2021), the first two groups will be commonly referred to as "intermediate" German learners of English and the last group will be henceforth referred to as "advanced" German learners of English. Detailed information on self-reported proficiency and language use measures for these participants, elicited by means of language background questionnaires, can be found in Llompart and Reinisch (2019b) and Llompart (2021)[2].

Materials

All participants took part in the same lexical decision task. In this task, real words of English as well as nonwords created by applying systematic phonological substitutions to real words were presented auditorily and participants had to decide whether each stimulus was a real word of English or not. As described in Llompart and Reinisch (2019b) and Llompart (2021), materials included 304 English unique words, of which 52 contained the phones in the difficult L2 contrast /ɛ/-/æ/. The remaining 252 words involved 5 contrasts (/i/-/ɪ/, /ɔ:/-/u/, /p/-/t/, /k/-/m/ and /b/-/v/) that were expected to be unproblematic for native German speakers. Importantly, half of the words were selected to appear in the task as canonically produced, while the other half was presented as nonwords in which the phones in the relevant contrasts were exchanged. Hence, the sets of canonically produced words and mispronounced nonwords contained different lexical items. For /ɛ/-/æ/ this meant that 13 words with /æ/ appeared with /æ/ produced as [æ] and 13 different words were presented with /æ/ mispronounced as [ɛ] (h[æ]mmer vs. *dr[ɛ]gon). The same manipulation held for items with /ɛ/ (d[ɛ]sert vs. *l[æ]mon), and the same procedure was also applied to filler contrasts. While for the critical items the target was always the first stressed vowel, for fillers the position of the critical phones in the word could vary. All 304 words were recorded by a male speaker of Southern British English in their correct form and half of the items, that is, those designed to appear as nonwords, were also recorded with the suitable phonological substitutions.

[2]The elicited measures of language use differed across studies, and are thus not fully comparable. By contrast, proficiency measures for English comprehension and spoken English were the same for the two studies, albeit elicited using slightly different scales. In spite of this, when the means for all three groups of participants as reported in the original studies are transformed into a common 0–10 scale (0 = very poor; 10 = very good), it is observable that the intermediate learners in Llompart and Reinisch (2019b) and Llompart (2021) provided very similar scores for both English comprehension (2019: 6.74; 2021: 6.5) and spoken English (2019: 5.5; 2021: 5.88), while the advanced learners in Llompart (2021) self-scored their abilities as higher for both measures (English comprehension: 7.55; spoken English: 6.72).

Procedure

Participants were tested either in a sound-attenuated booth or a quiet room at their respective universities. The lexical decision task was implemented in Psychopy 2 (in Llompart and Reinisch, 2019b; v. 1.83.01) or Psychopy 3 software (in Llompart, 2021; v. 3.0.2; Peirce et al., 2019). Auditory stimuli were presented over headphones at a comfortable listening level. Before the start of the task, participants were instructed that they would listen to a native English speaker say English words and invented words that could in some cases sound similar to English words. Their task was to indicate, for each item, whether they considered it to be a real word in English. On each trial, two boxes were shown on the screen, a green one with "word" written on it on the left-hand side and a red one with "not a word" written on it on the right-hand side, and an auditory stimulus was presented. Participants had to press "1" on a numeric keyboard (in Llompart and Reinisch, 2019b) or the leftmost button of a response pad (in Llompart, 2021) to indicate that the auditory stimulus was a real word, and "0" or the rightmost key of the response pad if they considered that the stimulus was not a real word. There was no time limit for participants' responses. The 304 items were presented in a randomized order. Before the start of the task, participants were presented with 10 practice trials in order to familiarize them with the procedure. It took participants approximately between 15 and 20 minutes to complete the task.

RESULTS

All analyses focused on /ɛ/- and /æ/-nonwords only, that is, the 13 items containing /ɛ/ → [æ] mispronunciations (e.g., *l[æ]mon) and the 13 items containing /æ/ → [ɛ] mispronunciations (e.g., *dr[ɛ]gon), respectively. Lexical frequencies and phonological neighborhood densities for these items were calculated in order to assess whether these lexical factors modulated participants' responses to the nonwords in the lexical decision task. Lexical frequency was assessed through the Zipf-scale frequency measures of Subtlex-UK (van Heuven et al., 2014) and neighborhood density was determined by consulting CLEARPOND (Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities; Marian et al., 2012). In addition, for the acoustic stimuli corresponding to these items, the F1 and F2 values (in hertz) of the critical vowels at vowel midpoint were extracted using a Praat script (Boersma and Weenink, 2009) so that the potential impact of the acoustics of the mispronounced vowel could also be examined. The difference score between F2 and F1 (F2–F1) was then calculated for each item in order to be able to use only one value per item in the analyses. In British English, [ɛ] is known to have a lower F1 and a higher F2 than [æ] (Deterding, 1997; Llompart and Reinisch, 2017). Therefore, the F2–F1 difference should always be higher for [ɛ] than [æ]. This was the case in the present stimuli, as the mean F2–F1 of the /æ/-nonwords, in which the first vowel was produced like [ɛ], was 1,187 Hz (*SD* = 91), and the mean F2–F1 of the /ɛ/-nonwords (i.e., with /ɛ/ produced like [æ]) was 568 Hz (*SD* = 103). The F2–F1 value of the critical vowel in each of the /ɛ/- and /æ/-nonwords is provided in **Table 1**, together with the lexical frequency and phonological neighborhood density of the word from which the nonword was derived. Correlational analyses over the set of 26 nonwords showed that lexical frequency and phonological neighborhood density were not correlated with each other [$r(24) = 0.08$, $p = 0.69$] and neither of them was significantly correlated with the F2–F1 values of the critical vowels either [lexical frequency: $r(24) = -0.28$, $p = 0.17$; phonological neighborhood density: $r(24) = -0.23$, $p = 0.27$].

Prior to any analyses, lexical decision data corresponding to responses to /ɛ/- and /æ/-nonwords were first trimmed by excluding all trials that contained nonwords based on words with which participants were not familiar. This was assessed by means of a word familiarity questionnaire administered after the lexical decision task. Only 26 trials were excluded on these grounds (0.86% of all /ɛ/- and /æ/-nonword trials). Before directly assessing the influences of lexical frequency, phonological neighborhood density and vowel acoustics on learners' responses, data were first submitted to a generalized mixed-effects regression model with a logistic linking function (lme4 package 1.1–23 in R version 3.6.3; Bates et al., 2015) on accuracy data with Vowel [/ɛ/ (produced as [æ]; *l[æ]mon) - /æ/ (produced as [ɛ]; *dr[ɛ]gon)] and Group (intermediate in Llompart and Reinisch (2019b), intermediate in Llompart (2021) and advanced in Llompart (2021)) as variables of interest. This model was devised to be used as the base model on which the effects of lexical frequency, phonological neighborhood density and vowel acoustics were to be subsequently tested (see below).

TABLE 1 | Lexical frequency, phonological neighborhood density, and F2–F1 values of the critical vowels for each of the /ɛ/- and /æ/-nonwords analyzed in the present study.

nonword item	Lexical frequency (Subtlex-UK)	Phonological neighborhood density (CLEARPOND)	Vowel acoustics (F2–F1 in hertz)
ch[æ]rry	4.23	11	710
ch[æ]ss	3.91	18	592
d[æ]sk	4.34	6	685
dr[æ]ss	4.85	9	621
fr[æ]sh	4.89	4	513
h[æ]lth	5.14	9	356
h[æ]lp	5.75	12	433
l[æ]gend	4.34	1	615
l[æ]mon	4.46	5	546
l[æ]sson	4.46	5	471
s[æ]ntence	4.34	1	592
w[æ]ther	5.12	12	578
y[æ]llow	4.84	10	679
ch[ɛ]nnel	4.51	6	1,086
dr[ɛ]gon	4.28	1	1,251
f[ɛ]ctor	4.5	7	1,189
f[ɛ]ctory	4.59	0	1,232
fl[ɛ]g	4.42	11	1,189
g[ɛ]llery	4.28	2	1,151
h[ɛ]bit	4.04	4	1,291
l[ɛ]mp	4.09	14	1,118
r[ɛ]mp	3.67	16	1,055
sc[ɛ]ndal	4.22	3	1,332
spl[ɛ]sh	4.13	3	1,045
st[ɛ]ndard	4.72	2	1,228
th[ɛ]nk	5.87	11	1,259

The base model had Response (0 = incorrect, 1 = correct) as categorical dependent variable. Vowel was contrast coded such that /ɛ/ was coded as −0.5 and /æ/ as 0.5. Group was re-coded as two linearly independent contrasts which will be henceforth referred to as "Proficiency" and "Study." "Proficiency" was coded to capture differences in accuracy between the two groups of intermediate learners and the group of advanced learners. Hence, trials for the former two groups were coded with −0.25, and trials corresponding to the latter were coded as 0.5. "Study" was included to assess potential differences between the two intermediate groups of learners, who were recruited as part of two different studies and in two different institutions but following the same recruiting requirements. Data from the intermediate learners in Llompart and Reinisch (2019b) were coded as −0.5, data from the intermediate learners in Llompart (2021) were coded as 0.5, and data from the advanced participants in the same study were coded as 0. Proficiency and Study were not allowed to interact but the interactions between each of these predictors and Vowel were included. The random effects structure consisted of random intercepts for Participants and a random slope for Vowel over Participants. Random intercepts for Items were not included because Item co-varied with lexical frequency, phonological neighborhood density and vowel acoustics (i.e., each item had one value for each variable) and would thus be problematic for the additional analyses examining their effects.

The model revealed significant effects of Vowel ($b = -0.98$; $z = -9.68$; $p < 0.001$) and Proficiency ($b = 1.58$; $z = 5.66$; $p < 0.001$). The effect of Study was not significant ($b = -0.03$; $z = -0.14$; $p = 0.89$) and neither were the interactions between Vowel and Proficiency and Vowel and Study (both $p > 0.1$). Hence, listeners were found to be more accurate with /ɛ/ → [æ] mispronunciations (/ɛ/-nonwords, e.g., *l[æ]mon; $M = 50.67\%$ correct, $SD = 50.01$) than with /æ/ → [ɛ] substitutions (/æ/-nonwords, e.g.,*dr[ɛ]gon; $M = 31.52\%$ correct, $SD = 46.47$) across the board, and learners labeled as advanced in Llompart (2021) were more accurate overall ($M = 58.61\%$ correct, $SD = 49.28$) than the two groups of intermediate learners (2019: $M = 35.5\%$ correct, $SD = 47.88$; 2021: $M = 34.6\%$ correct, $SD = 47.59$), whose nonword rejection accuracies were almost identical.

After this, separately for each of the three predictors of interest (i.e., lexical frequency, phonological neighborhood density and vowel acoustics), forward stepwise model comparisons (Zhang et al., 2020) were conducted between i) the base model described above (random-effects structure: Vowel|Participant) and a model including random slopes for one of the predictors over Participants (e.g., Vowel + Frequency|Participant), and ii) between the model including the random slopes only and a model including these random slopes plus an interaction term with Vowel over Participants (e.g., Vowel*Frequency|Participant). Comparisons were conducted by means of log-likelihood tests using the anova() function in R. These comparisons assessed whether the additional complexity of the random-effects structure improved the models' fit. In particular, the comparisons between the base model and the models with only random slopes were performed to ascertain whether lexical frequency, phonological neighborhood density and vowel acoustics modulated participants lexicality responses across the board, while the comparisons between the models with and without the interaction terms determined whether the effects were qualified by the type of nonword items (/ε/-nonwords vs. /æ/-nonwords).

This analytical procedure was selected because it allowed for the examination of item-specific effects in an independent way while still taking into account the population-level effects that had already been reported in the previous studies. By analyzing whether allowing the model to account for variation caused by participants' diverging sensitivities to lexical frequency, neighborhood density and vowel acoustics improved the fit of the model to the actual data, it could be determined whether these properties of the individual items affected participants' responses without having to deal with drawbacks that would have been unavoidable if these predictors had simply been added to the fixed-effects structure of the model. First, this approach avoided that the effects of item-specific properties were knowingly overestimated, as it would have been the case if they had been analyzed as the sole fixed effects, disregarding thus the effects that both target vowel and differences between learner groups have been shown to have in previous studies. Secondly, and very much relatedly, this procedure also prevented that the contributions of the item-specific measures investigated were obscured by the robust effects of the aforementioned variables.

Results of the model comparisons between the base model and three separate models including random slopes for Lexical Frequency, Phonological Neighborhood Density and Vowel Acoustics, respectively, over Participants showed that the addition of a random slope for Lexical Frequency over Participants improved the model's fit [$\chi^2(3) = 8.50$, $p < 0.05$], and so did adding a slope for Vowel Acoustics [$\chi^2(3) = 36.61$, $p < 0.001$]. By contrast, adding a slope for Neighborhood Density did not result in an improvement [$\chi^2(3) = 1.44$, $p = 0.70$]. Furthermore, comparisons between the models with random slopes only and models including an interaction term with Vowel revealed that the interaction terms between Vowel and Lexical Frequency over Participants [$\chi^2(4) = 17.52$, $p < 0.01$] and between Vowel and Neighborhood Density [$\chi^2(4) = 28.80$, $p < 0.001$] improved the fit of the respective models. The model including an interaction between Vowel Acoustics and Vowel over Participants had severe convergence issues that rendered it uninterpretable. However, a comparison involving simplified models in which the non-significant interactions between Vowel and Proficiency and Vowel and Study were removed from the fixed-effects structure showed that adding an interaction term between Vowel Acoustics and Vowel over Participants to the random-effects structure considerably improved the fit of the simplified model with random slopes for Vowel Acoustics only [$\chi^2(4) = 29.50$, $p < 0.001$].

Based on the significant improvements in model fit stemming from the addition of interaction terms to the random-effects structure, data were split by Vowel and the effects of adding random slopes for Lexical Frequency, Neighborhood Density and Vowel Acoustics were quantified for each vowel separately by comparing a base model with only random intercepts (1| Participant) to models with random slopes for Lexical Frequency, Phonological Neighborhood Density and Vowel Acoustics, respectively, over Participants (e.g., Frequency|Participant). For /æ/-nonwords (e.g., *dr[ε]gon), slopes for Lexical Frequency [$\chi^2(3) = 21.88$, $p < 0.001$] and Neighborhood Density [$\chi^2(3) = 26.07$, $p < 0.001$] over Participants improved the model's fit, while a slope for Vowel Acoustics did not [$\chi^2(3) = 1.00$, $p = 0.80$]. For /ε/-nonwords (e.g., *l[æ]mon), the opposite pattern emerged. A random slope for Vowel Acoustics over Participants substantially improved the model's fit [$\chi^2(3) = 63.29$, $p < 0.001$] while slopes for Lexical Frequency [$\chi^2(3) = 0.40$, $p = 0.94$] and Neighborhood Density [$\chi^2(3) = 1.75$, $p = 0.62$] did not do so. These results align perfectly with the patterns observed in the raw data presented in **Figure 1**, which provides scatterplots of accuracy in nonword rejection for /æ/-nonwords (top row) and /ε/-nonwords (bottom row) as a function of lexical frequency (left), neighborhood density (center) and vowel acoustics (right). Regression lines and correlation coefficients (i.e., *r*) are also provided to better outline the relationships between these variables.

Summarizing, model comparisons showed that nonword rejection accuracy for items containing mispronunciations of confusable L2 phones was modulated across the board by both the lexical frequency of the items and the acoustics of the critical vowels. However, the significant interactions and subsequent follow-up analyses indicated that the relative contributions of lexical frequency, neighborhood density and vowel acoustics differed between /æ/-nonwords and /ε/-nonwords. For the former, lower lexical frequencies and higher phonological neighborhood densities contributed to higher accuracies, whereas the F2–F1 values of the critical vowels did not strongly relate to nonword rejection accuracy (see **Figure 1**, top row). For /ε/-nonwords, higher accuracies in nonword rejection were only associated to lower F2–F1 values (i.e., more [æ]-like) for the critical vowels (see **Figure 1**, bottom row). Similar scatterplots to those in **Figure 1** but with data split by group are provided in **Figure 2**. An examination of **Figure 2** additionally suggests that the asymmetric patterns for the two types of mispronounced nonwords are highly consistent across the three groups of participants included in the sample.

DISCUSSION

The present study examined the effects of item-specific properties both related to the organization of the L2 lexicon and to the acoustics of the confusable L2 categories on rejection accuracy for nonwords only differing from real words in the phones of a difficult L2 phonological contrast. A series of additional analyses of lexical decision data from German learners of English (Llompart and Reinisch, 2019b; Llompart, 2021) were conducted to assess the effects of i) the lexical frequency of the L2 (non)words presented, ii) their phonological

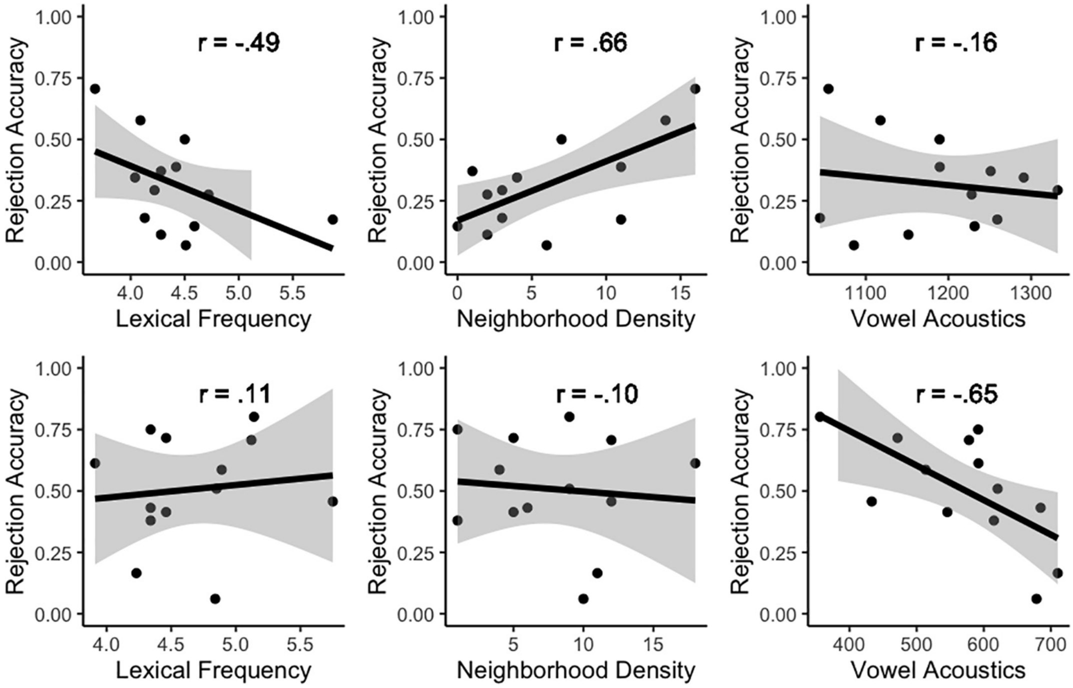


FIGURE 1 | Scatterplots of accuracy in nonword rejection for /æ/-nonwords (top row) and /ɛ/-nonwords (bottom row) as a function of lexical frequency (left), neighborhood density (center), and vowel acoustics (right). Regression lines and correlation coefficients (i.e., *r*) are provided for illustration purposes.

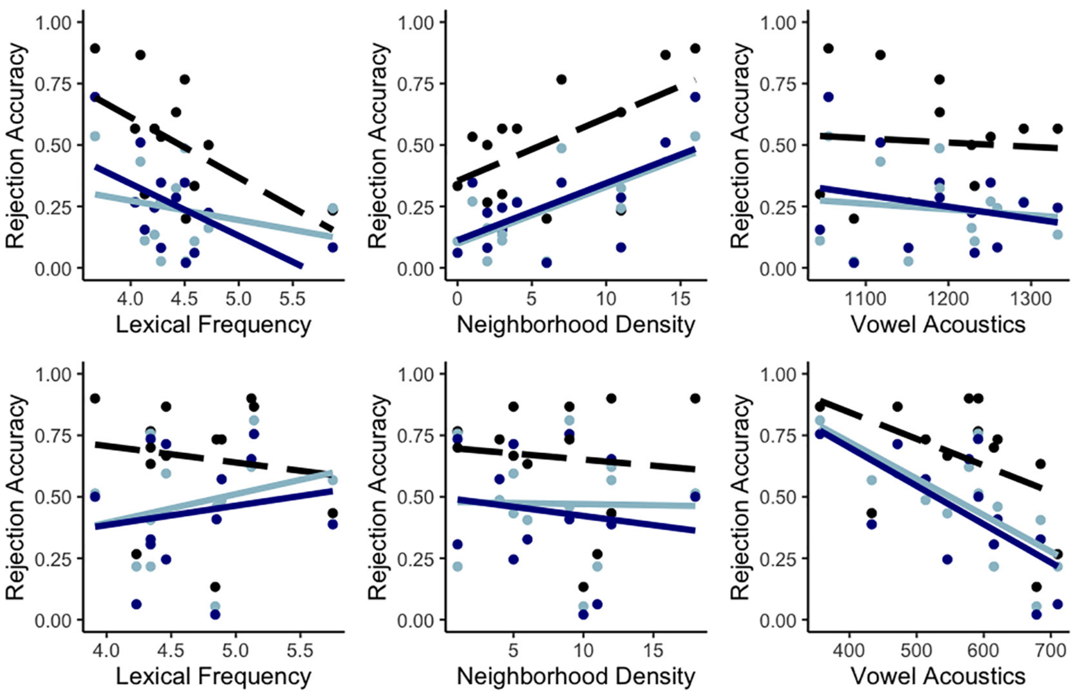


FIGURE 2 | Scatterplots of accuracy in nonword rejection for /æ/-nonwords (top row) and /ɛ/-nonwords (bottom row) as a function of lexical frequency (left), neighborhood density (center), and vowel acoustics (right) with data split by group. Advanced learners in Llompart (2021) are in black, intermediate learners in Llompart (2021) are in dark blue and intermediate learners in Llompart and Reinisch (2019b) are in light blue. Regression lines (advanced learners in dashed line, intermediate learners in solid lines) are provided for illustration purposes.

neighborhood densities, and iii) the spectral image of the critical L2 phones, on learners' ability to reject nonwords containing /ɛ/-/æ/ mispronunciations. These are factors that have not been considered in previous research but whose thorough investigation could improve our understanding of how lexical properties modulate the phonolexical encoding of phones in challenging L2 contrasts, as well as of the extent to which learners are sensitive to fine phonetic detail regarding the phones in such contrasts when engaging in lexical retrieval tasks. Even though the results of the present study should be interpreted with caution, as they stem from a limited set of L2 (non)words targeting just one L2 contrast and one learner population, they constitute a first stepping stone toward a better characterization of these issues, which are further discussed below.

Before actually gauging the effects of lexical frequency, neighborhood density and vowel acoustics in the present study, however, a first analysis was conducted to assess differences in accuracy as a function of vowel, or item type (/ɛ/-nonwords vs. /æ/-nonwords), and learner group. This analysis was conducted to confirm previous findings with a larger dataset and, most importantly, so that the model could then be used as a baseline to quantify the effects of the lexical and phonetic predictors of interest at a later stage. Results showed that learners were better at accurately detecting /ɛ/ → [æ] mispronunciations (/ɛ/-nonwords; e.g., *l[æ]mon) than /æ/ → [ɛ] mispronunciations (/æ/-nonwords; e.g., *dr[ɛ]gon) and that the group of advanced learners included in the analyses outperformed the two groups labeled as intermediate learners. This replicates the findings of previous studies showing nonword rejection asymmetries for words with difficult L2 phonological contrasts (Darcy et al., 2013; Simon et al., 2014; Llompart and Reinisch, 2019b; Melnik and Peperkamp, 2019, 2021) and proficiency and usage effects in nonword rejection for this type of items (Sebastián-Gallés et al., 2005; Amengual, 2016; Llompart, 2021). In addition to this, another relevant finding was that accuracy rates for the two intermediate learner groups, who were recruited and tested at different universities but by means of the same recruiting procedure, were found to be extremely similar. This evidences that the samples from Llompart and Reinisch (2019b) and Llompart (2021) were comparable and speaks in favor of the high reliability of this experimental paradigm when used with late L2 learners and applying systematic recruiting requirements.

The main question was, however, whether lexical frequency, phonological neighborhood density and vowel acoustics influenced nonword rejection accuracy on top of the previously mentioned effects. This was assessed by manipulating the presence or absence of random slopes for the three variables, as well as interaction terms between them and vowel, in the random-effects structure of the models while the fixed-effects structure remained constant. In that respect, results revealed that both the lexical properties of the target items and the acoustics of the critical vowels contributed to characterizing the variation observed for nonword rejection, albeit differently for the two types of items examined. For /æ/-nonwords (i.e., /æ/ → [ɛ] mispronunciations), lexical factors had robust modulating effects: First, nonwords whose real word counterparts had lower frequencies were more easily rejected than those that had higher frequencies. Secondly, nonwords based on words with more lexical neighbors were more easily rejected than those with fewer neighbors (see **Figures 1**, **2**, top row). In contrast, for /ɛ/-nonwords (i.e., /ɛ/ → [æ] mispronunciations), accurate rejection for individual items was tightly related to the acoustics of the critical vowel (/ɛ/ produced as [æ]), as higher rejection rates were associated to more extremely [æ]-like spectral articulations of /ɛ/ (see **Figures 1**, **2**, bottom row). Therefore, results showed clear asymmetries between /ɛ/-nonwords and /æ/-nonwords for the two lexical factors as well as for vowel acoustics.

With regard to vowel acoustics, the fact that it modulated the rejection of /ɛ/-nonwords (e.g., *l[æ]mon) but not /æ/-nonwords (e.g., *dr[ɛ]gon) indicates that L2 learners were indeed sensitive to small differences in the acoustic properties of the critical vowels when judging the lexicality of words and similar-sounding nonwords, but only when the mispronunciations in the latter went in one particular direction. A possible explanation for this asymmetry is that the more robust encoding of /ɛ/ (vs. /æ/) into the lexical representation of L2 words leads not only to higher accuracies when rejecting items in which the vowel is mispronounced, as already shown (Simon et al., 2014; Llompart and Reinisch, 2019b), but also to an enhanced attentiveness to how large (or small) the mismatch between the expected category and the acoustics of the input is. Building on the same argument, the lack of a relationship between vowel acoustics and rejection of *dr[ɛ]gon-type mispronunciations could be attributed to the "fuzzier" representation of /æ/ in L2 words containing this vowel. This would make L2 learners more tolerant of mispronunciations, and thus less accurate in their judgments, while also reducing their sensitivity to the magnitude of the mismatch between the input and the canonical vowel. In addition, note that, for the L2 contrast of interest, critical differences in peripherality between the two vowels could have also contributed to this asymmetric pattern. Given that /æ/ is more peripheral than /ɛ/ in the English vowel space, mispronunciations involving a substitution of the less peripheral vowel by the more peripheral one may have been more salient than the opposite type, enhancing the effect that small acoustic differences in the more peripheral region of the vowel space could have on learners' perception and subsequent decisions (Polka and Bohn, 2003, 2011).

The second asymmetry observed involved lexical frequency and phonological neighborhood density, which were found to only influence rejection accuracy for /æ/-nonwords. For lexical frequency, a potential explanation is that it only played a role for /æ/ → [ɛ] mispronunciations because these are very often encountered in German-accented English and learners most likely had had experience with items of this kind (Eger and Reinisch, 2019b; Llompart and Reinisch, 2019a, 2020). The effect of lexical frequency could thus be explained by the fact that the more frequent (non)words with /æ/ presented in the task, like *thank*, may have repeatedly been heard as *th[ɛ]nk in the speech of fellow L1-speakers, while less frequent words like *habit* probably not as much. Consequently, this would have led to learners being more likely to consider *th[ɛ]nk a real English word than *h[ɛ]bit. For /ɛ/-nonwords, on the contrary, as the mispronunciations in these items (/ɛ/ → [æ]) are not a typical marker of L1-accented speech, the amount of exposure to L1-accented input would not be expected to make a difference, and this would explain the lack of an effect of lexical frequency for these items. Since detailed information about the learners' L2 input would be needed to be able to properly assess whether L2 input characteristics could indeed be the source of this asymmetry, this explanation remains in need of further research at this point.

Finally, the effect of phonological neighborhood density for /æ/-nonwords indicates that, for the most problematic category in the contrast (i.e., /æ/), the existence of clusters of phonological neighbors containing the same target vowel made it more likely that learners spotted the corresponding mispronunciations[3]. This suggests that high phonological neighborhood densities may support the accurate phonolexical encoding of the vowel into particular L2 lexical representations, probably by strengthening the connection between the challenging non-native phonetic category and the clustered lexical items. For words containing /ε/, phonological neighborhood density may not be as crucial because of the dominant role of /ε/ in the phonological contrast and its relatively easier perceptual identification (Weber and Cutler, 2004; Cutler et al., 2006).

All in all, the present study provides a first approximation to the issue of how the lexicon and speech perception intertwine in the phonolexical encoding of difficult L2 contrasts from an item-centered perspective. Challenging L2 phonological contrasts introduce an additional level of "fuzziness" to L2 lexical representations, which are known to already be fuzzy because of the inherent characteristics of L2 learning itself (Cook and Gor, 2015; Cook et al., 2016; Lancaster and Gor, 2016). Previous studies have shown that, for non-native phonological contrasts in which the two L2 phones differ in how well they match L1 categories, the difficulties brought about by such phones are not symmetric (Weber and Cutler, 2004; Cutler et al., 2006; Darcy et al., 2013; Simonchyk and Darcy, 2017, 2018; Melnik and Peperkamp, 2019, 2021). This study contributes to this literature by suggesting that these asymmetries may also extend to the way in which phonolexical encoding takes place. Based on the present results, the encoding of the best-fitting or dominant L2 category (i.e., /ε/) appears not to be strongly constrained by lexical properties of specific L2 lexical items such as lexical frequency and phonological neighborhood density. This, in addition to the effects of vowel acoustics observed for /ε/-nonwords, suggests that, for this category, encoding may be more directly linked to learners' phonetic perception of the contrast. Note that this idea accounts well for the results of Llompart and Reinisch (2019b), who found that it was only for responses to (non)words with phonological /ε/ (and not /æ/) that a relationship with learners' perceptual flexibility in a distributional learning task could be found. In contrast, for the worse fitting, non-dominant category (i.e., /æ/), lexical decision data suggests that the level of success at phonologically encoding the non-native phonetic category into lexical representations is influenced by higher-level lexical properties that situate these items within the learners' vocabulary, and possibly relates to their familiarity with native and non-native input. Hence, the phonolexical encoding of /æ/ could be hypothesized to operate to a larger extent in a piecemeal manner (Lieven et al., 1997; Pine and Lieven, 1997) modulated by the learners' experience with the L2 and even with particular L2 words (Llompart, 2019, 2021). Future research including larger item samples and, ideally, also examining data from other experimental paradigms that tap into lexical retrieval will now be essential to ascertain to what extent the insights gained from nonword rejection in this study are robust and generalizable.

[3]As an additional check for phonological neighborhood density, it was assessed whether densities per item differed from these reported in **Table 1** when only neighbors containing the same target vowel were considered (vs. all neighbors). A correlational analysis showed that total neighborhood densities and neighborhood densities only including same-vowel neighbors were almost perfectly correlated [$r(24) = 0.98$, $p < 0.001$].

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://osf.io/u7syd/ (Open Science Framework).

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ML was solely responsible for the conception of the study, the analysis and interpretation of the data, and the drafting of the manuscript.

FUNDING

This work was funded by an Alexander von Humboldt Professorship (ID-1195918) awarded to Ewa Dąbrowska, Chair of Language and Cognition at the Department of English and American Studies of Friedrich Alexander University Erlangen-Nuremberg.

ACKNOWLEDGMENTS

I would like to thank Ewa Dąbrowska for her support of this research, Eva Reinisch for her comments on a previous version of the manuscript and the two reviewers for their helpful suggestions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.659852/full#supplementary-material

REFERENCES

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Appl. Psycholinguist.* 37, 1221–1251. doi: 10.1017/S0142716415000557

Andrews, S. (1996). Lexical retrieval and selection processes: effects of transposed-letter confusability. *J. Mem. Lang.* 35, 775–800. doi: 10.1006/jmla.1996.0040

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67:1406. doi: 10.18637/jss.v067.i01

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language experience in second language speech learning: In honor of James Emil Flege*, eds O. S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34. doi: 10.1075/lllt.17.07bes

Boersma, P., and Weenink, D. (2009). *Praat: doing phonetics by computer. Computer program* URL: http://www.praat.org

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Percept. Psychophys.* 61, 977–985. doi: 10.3758/BF03206911

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Lang. Cognit. Process.* 27, 1205–1224. doi: 10.1080/01690965.2012.660170

Coltheart, M., Davelaar, E., Jonasson, J. T., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance, volume VI*, ed. S. Dornick (New Jersey: Erlbaum), 535–556.

Cook, S. V., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *Ment. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *J. Phon.* 34, 269–284. doi: 10.1016/j.wocn.2005.06.002

Daidone, D. (2020). *How Learners Remember Words In Their Second Language: The Impact Of Individual Differences In Perception, Cognitive Abilities, And Vocabulary Size*. Ph.D. thesis, Bloomington: Indiana University.

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Ment. Lex.* 8, 372–420. doi: 10.1075/ml.8.3.06dar

Darcy, I., and Holliday, J. J. (2019). "Teaching an old word new tricks: Phonological updates in the L2 mental lexicon," in *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference*, eds J. Levis, C. Nagle, and E. Todey (Ames: Iowa State University), 10–26.

Darcy, I., and Thomas, T. (2019). When blue is a disyllabic word: perceptual epenthesis in the mental lexicon of second language learners. *Biling. Lang. Cogn.* 22, 1141–1159. doi: 10.1017/S1366728918001050

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *J. Int. Phon. Assoc.* 27, 47–55.

Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access. *Learn. Individ. Differ.* 22, 680–689. doi: 10.1016/j.lindif.2012.05.005

Eger, N. A., and Reinisch, E. (2019a). The impact of one's own voice and production skills on word recognition in a second language. *J. Exp. Psychol. Learn. Mem. Cogn.* 45, 552–571. doi: 10.1037/xlm0000599

Eger, N. A., and Reinisch, E. (2019b). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Stud. Second Lang. Acquis.* 41, 179–200. doi: 10.1017/S0272263117000377

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *J. Exp. Psychol. Hum. Percept. Perform.* 6, 110–125. doi: 10.1037/0096-1523.6.1.110

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9, 317–323. doi: 10.1016/0028-3932(71)90027-3

Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980

Lancaster, A., and Gor, K. (2016). Abstraction of phonological representations in adult nonnative speakers. *Proc. Linguistic Soc. Am.* 1:24. doi: 10.3765/plsa.v1i0.3725

Lieven, E. V. M., Pine, J. M., and Baldwin, G. (1997). Lexically-based learning and early grammatical development. *J. Child Lang.* 24, 187–219. doi: 10.1017/S0305000996002930

Llompart, M. (2019). *Bridging The Gap Between Phonetic Abilities And The Lexicon In Second Language Learning*. Ph.D. thesis, Munich: Ludwig Maximilian University Munich.

Llompart, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Biling. Lang. Cogn.* 24, 481–496. doi: 10.1017/S1366728920000656

Llompart, M., and Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1040–1056. doi: 10.1037/xhp0000383

Llompart, M., and Reinisch, E. (2019a). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Lang. Speech* 62, 594–622. doi: 10.1177/0023830918803978

Llompart, M., and Reinisch, E. (2019b). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Biling.Lang. Cogn.* 22, 1085–1100. doi: 10.1017/S1366728918000925

Llompart, M., and Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 1590–1610. doi: 10.1037/xlm0000832

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36.

Marian, V., Bartolotti, J., Chabal, S., and Shook, A. (2012). CLEARPOND: cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS One* 7:e43230. doi: 10.1371/journal.pone.0043230

Melnik, G. A., and Peperkamp, S. (2019). Perceptual deletion and asymmetric lexical access in second language learners. *J. Acoust. Soc. Am.* 145, EL13–EL18. doi: 10.1121/1.5085648

Melnik, G. A., and Peperkamp, S. (2021). High-Variability Phonetic Training enhances second language lexical processing: evidence from online training of French learners of English. *Biling. Lang. Cogn.* 24, 497–506. doi: 10.1017/S1366728920000644

Nation, I. (2006). How large a vocabulary is needed for reading and listening?. *Can. Mod. Lang. Rev.* 63, 59–82. doi: 10.3138/cmlr.63.1.59

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., et al. (2019). PsychoPy2: experiments in behavior made easy. *Behav. Res. Methods* 51, 195–203. doi: 10.3758/s13428-018-01193-y

Perea, M., Rosa, E., and Gómez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Percept. Psychophys.* 67, 301–314. doi: 10.3758/BF03206493

Pine, J. M., and Lieven, E. V. M. (1997). Slot and frame patterns and the development of the determiner category. *Appl. Psycholinguist.* 18, 123–138. doi: 10.1017/S0142716400009930

Politzer-Ahles, S., Lee, K. K., and Shen, L. (2020). Ganong effects for frequency may not be robust. *J. Acoust. Soc. Am.* 147, EL37–EL42. doi: 10.1121/10.0000562

Polka, L., and Bohn, O.-S. (2003). Asymmetries in vowel perception. *Speech Commun.* 41, 221–231. doi: 10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. *J. Phon.* 39, 467–478. doi: 10.1016/j.wocn.2010.08.007

Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The influence of initial exposure on lexical representation: comparing early and simultaneous bilinguals. *J. Mem. Lang.* 52, 240–255. doi: 10.1016/j.jml.2004.11.001

Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., and Tare, M. (2015). Non-native phonemic discrimination, phonological short term memory, and word learning. *J. Phon.* 50, 99–119. doi: 10.1016/j.wocn.2015.03.001

Simon, E., Sjerps, M. J., and Fikkert, P. (2014). Phonological representations in children's native and non-native lexicon. *Biling. Lang. Cogn.* 17, 3–21. doi: 10.1017/S1366728912000764

Simonchyk, A., and Darcy, I. (2017). "Lexical encoding and perception of palatalized consonants in L2 Russian," in *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, eds M. O'Brien and J. Levis (Ames: Iowa State University), 121–132.

Simonchyk, A., and Darcy, I. (2018). The effect of orthography on the lexical encoding of palatalized consonants in L2 Russian. *Lang. Speech* 61, 522–546. doi: 10.1177/0023830918761490

Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., and Nishi, K. (2007). Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *J. Acoust. Soc. Am.* 122, 1111–1129. doi: 10.1121/1.2749716

van Heuven, W. J. B., Mandera, P., Keuleers, E., and Brysbaert, M. (2014). Subtlex-UK: a new and improved word frequency database for British English. *Q. J. Exp. Psychol.* 67, 1176–1190. doi: 10.1080/17470218.2013.850521

Vitevitch, M. S. (2002a). Influence of onset density on spoken-word recognition. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 270–278. doi: 10.1037/0096-1523.28.2.270

Vitevitch, M. S. (2002b). Naturalistic and experimental analyses of word frequency and neighborhood density effects in slips of the ear. *Lang. Speech* 45, 407–434. doi: 10.1177/00238309020450040501

Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0

Zhang, H., Carlson, M. T., and Diaz, M. T. (2020). Investigating the effects of phonological neighbours on word retrieval and phonetic variation in word naming and picture naming paradigms. *Lang. Cogn. Neurosci.* 35, 980–991. doi: 10.1080/23273798.2019.1686529

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fcomm.2021.689423

Check for updates

Advanced Second Language Learners of Mandarin Show Persistent Deficits for Lexical Tone Encoding in Picture-to-Word Form Matching

Eric Pelzl[1,2]*, *Ellen F. Lau*[1], *Taomei Guo*[3] *and Robert M. DeKeyser*[1]

[1]*University of Maryland, College Park, MD, United States,* [2]*The Pennsylvania State University, University Park, PA, United States,* [3]*Beijing Normal University, Beijing, China*

People who grow up speaking a language without lexical tones typically find it difficult to master tonal languages after childhood. Accumulating research suggests that much of the challenge for these second language (L2) speakers has to do not with identification of the tones themselves, but with the bindings between tones and lexical units. The question that remains open is how much of these lexical binding problems are problems of *encoding* (incomplete knowledge of the tone-to-word relations) vs. *retrieval* (failure to access those relations in online processing). While recent work using lexical decision tasks suggests that both may play a role, one issue is that failure on a lexical decision task may reflect a lack of learner confidence about what is *not* a word, rather than non-native representation or processing of known words. Here we provide complementary evidence using a picture-phonology matching paradigm in Mandarin in which participants decide whether or not a spoken target matches a specific image, with concurrent event-related potential (ERP) recording to provide potential insight into differences in L1 and L2 tone processing strategies. As in the lexical decision case, we find that advanced L2 learners show a clear disadvantage in accurately identifying tone mismatched targets relative to vowel mismatched targets. We explore the contribution of incomplete/uncertain lexical knowledge to this performance disadvantage by examining individual data from an explicit tone knowledge post-test. Results suggest that explicit tone word knowledge and confidence explains some but not all of the errors in picture-phonology matching. Analysis of ERPs from correct trials shows some differences in the strength of L1 and L2 responses, but does not provide clear evidence toward differences in processing that could explain the L2 disadvantage for tones. In sum, these results converge with previous evidence from lexical decision tasks in showing that advanced L2 listeners continue to have difficulties with lexical tone recognition, and in suggesting that these difficulties reflect problems both in encoding lexical tone knowledge and in retrieving that knowledge in real time.

Keywords: second language, Mandarin, lexical tone, ERPs, speech perception, Chinese, fuzzy lexical representations

Edited by:
Anna Chrabaszcz,
University of Pittsburgh, United States

Reviewed by:
Seth Wiener,
Carnegie Mellon University,
United States
Wanting Huang,
The University of Hong Kong, China

***Correspondence:**
Eric Pelzl
pelzlea@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *31 March 2021*
Accepted: *02 June 2021*
Published: *22 June 2021*

Citation:
Pelzl E, Lau EF, Guo T and DeKeyser RM (2021) Advanced Second Language Learners of Mandarin Show Persistent Deficits for Lexical Tone Encoding in Picture-to-Word Form Matching. Front. Commun. 6:689423. doi: 10.3389/fcomm.2021.689423

INTRODUCTION

People who grow up speaking a language without lexical tones typically find it difficult to master tonal languages after childhood. They may confuse or misidentify tones in speech early on (e.g., Wang et al., 1999), and they often end up with a large store of *fuzzy* second language (L2) tone word representations, that is, mental lexical representations with missing, incorrect, or uncertain tone representations (Pelzl et al., 2020). This outcome is not surprising, given that F0 (pitch) is used for many things in non-tonal languages (stress, intonation, emphasis, singing), but does not differentiate one word from another.

By far the most studied L2 tone language is Mandarin Chinese (for a review, see Pelzl, 2019). Mandarin has four citation tones, differentiated primarily by F0 height and contour (Howie, 1974; Ho, 1976). Relative to a speaker's own vocal pitch range, Mandarin Tone 1 is high and level; Tone 2 rises from mid to high; Tone 3 is low; and Tone 4 falls from high to low. Along with consonants and vowels, these tones serve to uniquely identify each syllable-sized unit (typically a morpheme or word) of spoken Mandarin.

Misidentification of Mandarin tones is common among naïve and novice learners (e.g., Wang et al., 1999; Alexander et al., 2005; Bent et al., 2006; Huang and Johnson, 2010; So and Best, 2010). For more experienced learners, tone identification and categorization abilities improve and many individuals approach native levels on categorization and identification tasks (Ling and Grüter, 2020; Pelzl, 2019; Shen and Froud, 2016; Tsukada and Han, 2019; Zou et al., 2017). Nevertheless, similarities between F0 contours among Mandarin tones can lead to confusions among some tones (e.g., in isolated syllables, Tone 3 may have a dipping contour leading it to resemble Tone 2). Such confusions can persist into intermediate and advanced levels of L2 proficiency (Lee et al., 2010; Hao, 2012; Tsukada and Han, 2019).

Given that tone identification is already a challenge, it is not surprising that using tones to differentiate words in Mandarin is also difficult for many novice learners (Wong and Perrachione, 2007; Chandrasekaran et al., 2010; Chang and Bowles, 2015). Perhaps less expected is that the same difficulties appear to persist into more advanced stages of learning, even for many learners who have achieved strong categorization or identification abilities (Han and Tsukada, 2020; Ling and Grüter, 2020; Pelzl et al., 2019; Pelzl et al., 2020). We will refer to this type of difficulty as an L2 *tone word* difficulty, that is, it is not necessarily about tones alone, but about how representations of the tone categories are bound to the lexical representations in long-term memory.

This L2 tone word difficulty may best be understood as *phonolexical* in nature. A difficult L2 phonological contrast (tones) impacts the learner's mental representations of the relevant lexical units, leading to fuzzy representations of lexical tones. While this situation is similar to documented segmental learning challenges in other L2 contexts (Díaz et al., 2012; Darcy et al., 2013; Chrabaszcz and Gor, 2014; Cook and Gor, 2015; Amengual, 2016; Cook et al., 2016; Gor and Cook, 2020; Llompart and Reinisch, 2020), L2 tone word difficulties may also differ in important ways. For instance, learning a novel L2 vowel contrast may require a learner to add new categories in their phonological vowel space—a challenge addressed by many models of L2 phonological learning (Best and Tyler, 2007; Escudero and Boersma, 2004; Flege, 1995). However, using vowels to differentiate meaningful lexical units is already a given. For tones, non-tonal language speakers must not only learn to categorize F0 contrasts on syllable-sized units, they must also learn to apply these new tone categories to the process of word recognition. This is a functional leap that may not come easily.

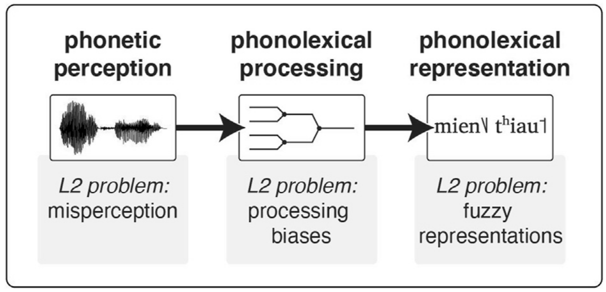


FIGURE 1 | Word recognition "pipeline" depicting three sites of potential difficulty in L2 (tone) word recognition: phonetic perception, phonolexical processing, and phonolexical representation.

When it comes to the fuzzy L2 lexical representations that result from such difficulties, tone word representations are much like those of purely segmental representations. They will vary from low to high quality, and this is likely to be closely related to a learner's familiarity with those words (Diependaele et al., 2013; Gor and Cook, 2020). An individual word's representation may have tones that are uncertain, incorrect, or completely missing (Pelzl et al., 2020). Each of these problems has the potential to impede fluent spoken word recognition. In this study, we set out to examine how this happens in more detail, specifically asking 1) whether tone word recognition errors will persist even when we control for fuzzy L2 word knowledge, and 2) whether L2 listeners' neural responses for correctly recognized words will display early sensitivity to tones.

Three General Explanations for L2 Tone Word Processing Difficulties

To provide some theoretical context for L2 difficulties with tone word recognition, we begin with a rough sketch of three basic ways in which it might break down (for more detail, see Pelzl et al., 2019, Pelzl et al., 2020). We use the metaphor of a pipeline to capture the way these issues feed one into another (**Figure 1**).

First, an L2 listener may have difficulty in what we are calling *phonetic perception,* that is, accurately perceiving the unfamiliar sounds of the L2. Specific to tones, many beginning and novice L2 learners regularly misidentify or confuse similar tone categories when recognizing words or syllables (e.g., Chandrasekaran et al., 2010; Chang and Bowles, 2015; Wang et al., 1999; Wong and Perrachione, 2007). Our own previous research suggests many advanced learners develop excellent tone identification abilities for monosyllables (Pelzl et al., 2019), and other studies have also found strong—if not completely nativelike—L2 categorization of tones among more advanced learners (Ling and Grüter, 2020; Shen and Froud, 2016, 2019; Zou et al., 2017). While overall impressive, such results are not a claim for flawless L2 tone perception. Though not examined in detail here, tone identification results from the participants in the current study show that advanced L2 listeners may struggle to identify tones when there is a following syllable (for details, see Pelzl, 2018), showing accuracy that is notably lower than L1 participants for the same context. So then, weaknesses in tone identification remain a possible cause of persistent L2 tone word recognition difficulties. If listeners cannot faithfully perceive tones in the acoustic-phonetic signal, they will have difficulty using and encoding them. In this case, the breakdown occurs near the beginning of the word recognition pipeline (**Figure 1**). The "substance" (the perceived speech sounds) that enters the pipeline is already problematic. This could well have knock-down effects leading to fuzzy L2 lexical representations (cf. Matusevych et al., 2021).

Second, an L2 listener may have difficulty processing the perceived speech sounds as lexical cues. We are calling this *phonolexical processing*. This is the real-time process that links the perceived phonetic signal to phonolexical representations encoded in long-term memory. It roughly corresponds to the phonological step of the phonetic-phonological-lexical continuum (Wong and Perrachione, 2007; Chan and Leung, 2020), but we wish to stress the *lexical* aspect of this process, along with the phonological. For L2 learners, years of experience attending only to the important phonetic features of their L1 might interfere in real-time word recognition. Such ideas have long been part of L2 theories, under a variety of terms (e.g., *cue competition* MacWhinney and Bates, 1989; *selective perception routines* Strange, 2011; *perceptual attention* Chang, 2018). In the case of L2 tone learning, listeners might privilege segmental information over tonal information. This need not be all or nothing; contextual factors or a learner's wider knowledge about the language might impact when and how tones are used. For example, Wiener et al. (2018), Wiener et al. (2020) have shown that when speech is produced by several different speakers (as opposed to just one), L2 learners tend to rely more on their experiential knowledge of syllable + tone co-occurrence probabilities (i.e., which tones are most likely with which syllables), and less on the acoustic-phonetic signal itself. When L2 processing strategies do not give appropriate weight to tones, it can lead to errors or inefficiencies during lexical retrieval. With respect to our pipeline, in this case the function of the pipe itself is the problem. Accurately perceived spoken tone words enter, but key features get siphoned off before they reach their destination.

Finally, an L2 listener may have difficulty with *phonolexical representations*, that is, encoding the lexical units of speech in their mental lexicon. This difficulty may lead to a variety of issues for tone representations: they may be entirely missing, incorrect, only known with some degree of uncertainty, or even confidently incorrect (Pelzl et al., 2020). If a given representation is not faithful to the actual form of the spoken tone word, this has the potential to lead to a variety of problems during lexical access. If lexical activation is very strict, the appropriate lexical unit might fail to be activated due to misaligned tone representations. If lexical activation is more lenient, inappropriate competitors could become active during lexical competition (Broersma and Cutler, 2008; Broersma, 2012). This representational account puts the problem at the end of the pipeline. The perceived speech sounds enter and run through the pipe smoothly, but the destination is incorrect.

Each of these difficulties is likely to make its own contribution to the fuzziness of L2 tone word representations. In previous work (Pelzl et al., 2019), we found that advanced L2 Mandarin learners (native English speakers) displayed near-native abilities on a challenging tone identification task, suggesting excellent phonetic perception of tones. In that task, we used syllables clipped out of disyllabic words that had been produced in continuous speech. Despite their strong performance when identifying those tones in isolation, when we presented the disyllabic words themselves—many of which contained the very same spoken syllables as used in the identification task—most L2 learners performed below chance in rejecting tonal nonword competitors of real Mandarin words (e.g., nonword *fang4*zi*/fɑŋ4tsɹ/derived from real word *fang2zi* "house"). This extreme difficulty was only seen for tone nonwords, not vowel nonwords (e.g., nonword *feng2zi*/fəŋ2tsɹ/). Taken together, the excellent tone identification paired with chance performance for tone nonwords suggested to us that auditory tone perception is unlikely to be the primary source of L2 tone word recognition difficulty. A follow-up study used more clearly produced (non)words and once again found a strong difference between tone and vowel nonwords (Pelzl et al., 2020).

The Present Study

Although our previous studies provided compelling evidence of advanced L2 Mandarin learners' tone word difficulties, the use of lexical decision as the primary test may have painted a particularly dismal picture. Rejection of a nonword in a lexical decision task requires a lexical search on the part of listeners (i.e., to confirm that a nonword does *not* exist in the lexicon). So then, a person's failure on any given lexical decision trial may reflect their lack of confidence about what is *not* a word, rather than their fuzzy representation or processing of the targeted real word.

To specifically target L2 learners' knowledge and recognition of lexical tone in real words, we conducted a picture-phonology matching experiment with native Mandarin speakers and advanced L2 learners of Mandarin (cf. Desroches et al., 2009). In the picture-phonology matching task, people see an image (noodles) and then hear a word that either matches or does not match the image. The auditory targets are real words (*mian4tiao2* "noodles"), or nonwords with a mismatching vowel (*men4tiao2*) or tone (*mian1tiao2*). Unlike lexical decision, a picture-phonology matching trial requires only knowledge of the specific word targeted in a trial. If the listener can successfully bring that word to mind, their task is simply to determine whether the auditory stimulus matches it or not. Here we allowed a full 1.75 s of picture-viewing time before the onset of the word, to provide L2 listeners with plenty of time to bring the word to mind. Thus, if the listener knows the pictured word and can faithfully perceive the spoken prompt, they should be able to confidently reject the mismatching nonwords. Another motivation for the picture-phonology matching task is to place L2 word recognition into a meaningful—albeit very simple—context. While tests of isolated word recognition can be a useful tool for understanding lexical processes, most words do not occur in isolation; typically, contextual cues help listeners create expectations about what words they expect to hear. The current paradigm mirrors this real-world situation, but uses a simple picture context to avoid complications that can arise in interpreting how much of a prior sentence context L2 learners have access to (see Pelzl et al., 2019 for discussion).

Our primary question was whether, when a prior visual context was provided in this way, advanced L2 learners of Mandarin would still show the same kind of disadvantage with lexical tone information vs. vowel information that we observed in the experiments on lexical decision. If the disadvantage we observed in prior experiments was primarily due to learners' lack of confidence about what tone words they *haven't* heard, then it might disappear when the task is focused only on determining the match of a picture to a known word. However, if the disadvantage is due to phonolexical encoding or processing of tone, then it should persist under the current conditions. We were also interested in gaining more insight into how much of any tone disadvantage observed is due to phonolexical encoding vs. processing. Therefore, we conducted an offline vocabulary knowledge post-test so that we could determine whether L2 listeners persist in (incorrectly) accepting tone mismatches even when they have correct and confident knowledge of the relevant words.

More exploratorily, during the picture-phonology matching experiment we also collected concurrent electrophysiological responses in order to look for signs of differential processing of lexical tone in native and L2 listeners that might explain different behavioral performance. Because the smaller number of incorrect behavioral responses in this paradigm are insufficient for ERP analysis, we focused on examining the ERPs from trials that had a correct behavioral response. Although these represent cases in which the L2 learners, like the native listeners, succeeded in accepting real words or rejecting tone/vowel mismatches, this real-time neural data could suggest differential approaches to lexical processing that could explain the profile of errors observed in the L2 learners. In the next section, we briefly review some background on the ERP responses that might provide such clues.

The Phonological Mismatch Negativity and Late Positive Component Responses in Event-Related Potential Research

The picture-phonology matching task sets up strongly constraining lexical expectations. Prior ERP research suggests that native speakers performing such a task with words and near-neighbor nonwords are likely to show modulation in three ERP components: the phonological mismatch negativity (PMN), the N400, and a late positive component (LPC). However, because the N400 component overlaps with the other two and can be modulated by nonword status itself as well as real word expectation (cf. Newman and Connolly, 2009), we chose to focus on the PMN and LPC responses in the current study.

The phonological mismatch (or mapping) negativity typically occurs between 200 and 400 ms after stimulus onset and is hypothesized to index neural responses to unexpected/mismatching phonological content in words, relative to expected words (Connolly and Phillips, 1994; Desroches et al., 2009; Newman and Connolly, 2009; see also discussion of the "N200" in e.g., Brunellière and Soto-Faraco, 2015; Van Den Brink et al., 2001). The PMN has been consistently observed in previous ERP research of Mandarin spoken words (Zhao et al., 2011; Malins and Joanisse, 2012), although it has not always been overtly analyzed or labeled as such (Liu et al., 2006; Pelzl et al., 2019). Of particular relevance is the study by Malins and Joanisse (2012), which used a picture-word paradigm with single syllable Mandarin words. In their study, all auditory stimuli were real words, and they manipulated the relation to pictures so that either consonants, vowels, tones, or complete syllables matched/mismatched the evoked word. They found significant PMN and N400 effects for all mismatch types. An MEG study has linked the PMN to activity generated in anterior left auditory cortex (Kujala et al., 2004). Within the EEG literature, PMN peaks have appeared variably across anterior, central, and posterior electrode sites. In the present case, because our nonwords differ from real words only with respect to a tone or a vowel in the first syllable, we expect that PMN responses will be evoked in native speakers as soon as the departure from the target word becomes apparent.

Along with PMN responses, we also expect to see strong late positive components (LPCs) in native speakers. In sentence processing experiments, late positivities are often classified as P600s and are hypothesized to reflect reanalysis or repair processes when people are confronted by infelicitous syntax (Gouvea et al., 2010; Kaan and Swaab, 2003; Osterhout and Holcomb, 1992), though similar effects have been observed for lexical violation (e.g., Romero-Rivas et al., 2015; Schirmer et al., 2005) and phonological mismatches (e.g., Schmidt-Kassow and Kotz, 2009). Importantly, we observed LPCs in our previous sentence processing ERP study when L1 listeners detected tone and rhyme mismatches in nonwords (Pelzl et al., 2019). Although, not a sentence processing study, similar effects—though not analyzed—are also apparent in the later portion of waveforms for vowel and tone mismatches in Malins and Joanisse (2012, p. 2037 Figure 1). Thus, we expect to find LPCs in response to picture-phonology mismatches in the present case. These effects are often described as indexing error detection, repair, reanalysis, or reorientation processes and may be related to more general (i.e. non-linguistic) processing mechanisms (Coulson et al., 1998; Sassenhagen and Bornkessel-Schlesewsky, 2015; Sassenhagen et al., 2014; for a review, see; Leckey and Federmeier, 2019).

What differences might we expect to see in L1 and L2 ERP responses in our analysis of trials with accurate behavioral responses? Given that the LPC essentially indexes the attentional processes that lead to decisive rejections, we expected this component to align fairly well with behavioral responses across groups, such that both L1 and L2 listeners would show larger LPCs for correctly rejected tone mismatches and vowel mismatches relative to correctly accepted matching words.

However, if the L1 and L2 speakers arrive at those correct responses in different ways, we might expect to see differences across groups in the earlier PMN response. One possibility is that L2 speakers have incomplete encoding of lexical tone such that they are unable to fully retrieve it to form a prediction for the upcoming speech input. This would predict that the L2 group would show a PMN for vowel mismatches relative to matching real words, but not for tone mismatches. Another possibility is that L2 speakers use a different processing strategy across the board: they may not be able to use the picture to generate a detailed phonological prediction in the same way as native speakers do, which might manifest as an absence of PMN effects in all conditions. Such a pattern would not directly account for the tone disadvantage, but might point to differences in processing that indirectly contribute to phonolexical encoding or processing problems.

In summary, the picture-phonology matching experiment aims to create a scenario where L2 listeners are given strong odds of success in recognition of tone mismatches in a lexical context, and, by recording ERPs aims to examine L2 neural responses to the tone and vowel cues as they occur.

MATERIALS AND METHODS

Participants

We recruited 19 native English speakers, all of whom had achieved advanced levels of proficiency in spoken Mandarin Chinese (**Table 1**).[1] One was excluded due to early onset of learning (age 7), and one was removed from analyses due to excessive artifacts in EEG data. This left 17 advanced L2 participants. All participants passed two screening measures (yes/no vocabulary test and Can-do self-assessment). The measures and criteria were the same as used in Pelzl et al. (2019), Pelzl et al. (2020). Due to the difficulty of finding sufficient L2 participants, one L2 participant was accepted despite a slightly lower vocabulary score (65.7) than criterion (70). Additionally, all participants completed a tone identification task, testing their ability to identify tones produced by four different talkers that were presented either in isolated syllables, or on the first syllable of a disyllabic target (contextualized syllables). Due to space constraints, we do not present the full details of the tone identification here (see Pelzl, 2018).

Twenty-four native Chinese speakers also completed the experiment (average age = 26.1). One was excluded due to equipment failure, and three were excluded due to excessive EEG artifacts, leaving twenty L1 participants for all analyses presented below.

All participants gave informed consent and were compensated for their time.

TABLE 1 | Background information, screening measures, and tone identification scores for L2 participants (n = 17).

	Mean (sd)	Range
Age at testing	25.8 (4.9)	18–38
Age of onset	17.5 (4.0)	11–25
Semesters of formal study	9.0 (5.0)	3–20
Years in immersion	3.5 (2.7)	0.7–9
Total years learning	8.3 (3.8)	3–19
Can-do self-assessment (%)	82.7 (7.6)	72.8–96.8
Vocabulary self-assessment (%)	88.2 (9.6)	65.7–100
Tone identification accuracy (%): *overall*	85.8 (7.7)	71.9–99.2
isolated syllables	89.5 (4.2)	81.2–98.4
contextualized syllables	82.1 (12.3)	57.8–100

Task and Stimulus Design

In the picture-phonology matching experiment, participants saw a picture followed either by a word that matched the picture or by a nonword that mismatched the pronunciation of the word evoked by the picture.

Critical stimuli were based on a set of 96 disyllabic real words[2]. All were highly frequent imageable nouns, chosen so that a corresponding picture could be matched to each one (e.g., *mian4tiao2* ‘noodles’). Words were first sought in beginning and intermediate levels of the popular L2 Mandarin textbook series *Integrated Chinese*. Additional words were chosen based on frequency in the SUBTLEX-CH corpus (Cai and Brysbaert, 2010) and the intuitions of the first author, an L2 Mandarin speaker and former Mandarin teacher.

In order to make pictures as easily identifiable as possible, photographic images were used[3] The majority of images were taken from two freely available picture databases (BOSS: Brodeur et al., 2010; Ecological SVLO: Moreno-Martínez and Montoro, 2012), with additional images culled from other free photo repositories (e.g., Wikimedia commons). A small number of difficult to find images were purchased from Adobe Stock, and two more images were created specifically for the experiment. An example image is shown in **Figure 2**. All images were placed on a white background. No attempt was made to control colors or luminosity as the neural response to the presentation of the images was not of interest. Instead we aimed to make images as recognizable as possible.

To assure that images would evoke the intended words, two rounds of picture norming were conducted. In each round, ten native Mandarin speakers generated Chinese words for 132 images. Images that were judged to perform inadequately in

[1]This experiment was part of a larger study that was the first author’s dissertation (Pelzl, 2018). A brief overview of the full design is included in **Supplementary Appendix A**. Participants are the same as those described in Pelzl et al. (2020), though there were some participant exclusions in the current dataset due to excessive EEG artifacts in the picture-phonology task.

[2]There was no overlap in target words between the current study and the lexical decision stimuli reported in Pelzl et al. (2020).

[3]For the words *tian1shi3* ‘angel’ and *mo2gui3* ‘devil’, computer generated 3-D cartoon images were used, as no angels or demons were available for photos.

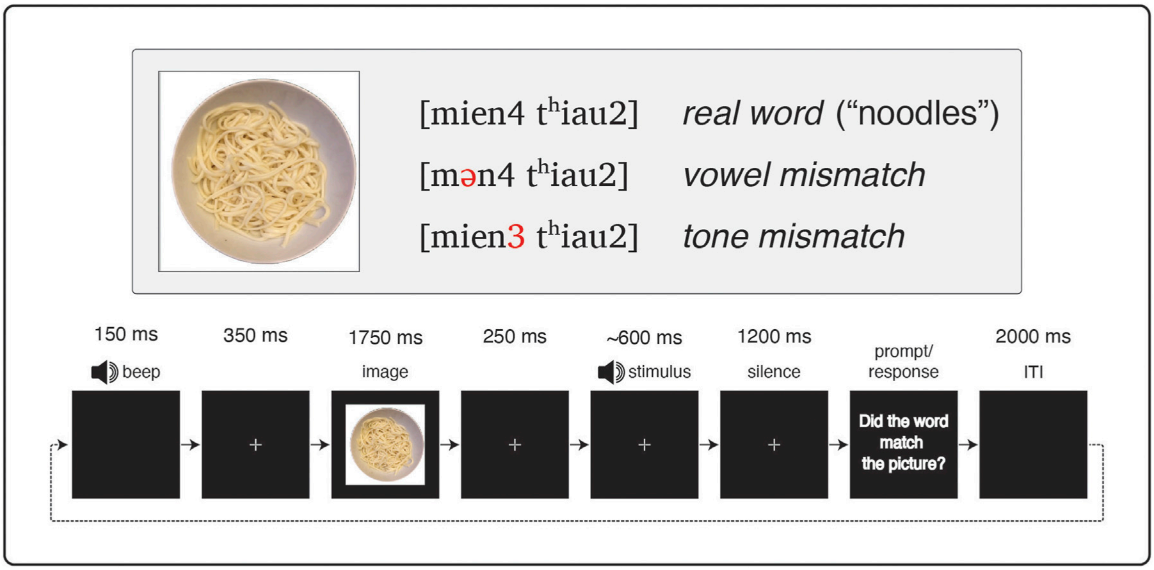


FIGURE 2 | Example stimulus set and trial parameters for picture-phonology matching task.

the first round (less than 70% generation of the target word, or generation of problematic competitor words) were replaced and a second round of norming was conducted with a new group of ten people. The end result was a set of 96 critical images that had an average word generation rate of 86%, though a handful of items (7 total) had rather low naming rates (under 50%). Future work might try to replace either those words or images. Images for filler items were also overall highly identifiable.

The real words were further manipulated to create two types of nonwords. The first syllable of the nonwords mismatched the real word counterpart with respect to either a tone or a vowel. For example, the real word *mian4tiao2* /mien4t^{h}iau2/ became the vowel nonword *m**e**n4tiao2* /mən4t^{h}iau2/ and the tone nonword *mian**3**tiao2* /mien3t^{h}iau2/. All possible tone combinations and manipulations were balanced across words and nonwords. For vowel mismatches, the syllable rhyme was changed by switching, adding, or deleting a single vowel sound (monophthong), though in a handful of cases, changes affected multiple vowel sounds (diphthongs). As much as possible, repetition of first syllables was avoided across stimuli, though some exceptions had to be made to accommodate the limited availability of imageable words that were likely to be known by L2 participant (all stimuli are available in the **Supplementary Appendix B**).

These procedures resulted in a total of 96 critical real word/vowel nonword/tone nonword triplets. An additional 16 real words with accompanying images were selected as fillers. Given all the constraints noted above, it was not possible to limit selection of fillers to words with a balanced occurrence of tones, and many filler items had neutral tones on the second syllable.

Three lists were constructed to balance images and words across participants. For each list, four unique pseudo-random presentation orders were prepared, with conditions balanced so that no more than three trials in a row would require the same response type (yes or no).

We also designed an offline vocabulary post-test. For each L2 participant, the test included all real word counterparts for vowel and tone mismatched nonwords encountered during the picture-phonology matching task (64 words total; words that occurred in the 'real word' condition were not tested). Each item provided Chinese characters and toneless Pinyin. Participants supplied tones (numbers 1-4 for each syllable), an English definition, and a confidence rating from 0–3 for both the tones and the definition of each item. Participants were informed that the 0–3 scale had the following meaning: *0 = I don't recognize this word; 1 = I recognize this word, but am very uncertain of the tones/meaning; 2 = I recognize this word, but am a bit uncertain of the tones/meaning; 3 = I recognize this word, and am certain of the tones/meaning*. This scale remained visible as a reference through the duration of the test. For any tones or definitions they did not know, participants were instructed to leave the answer blank and supply "0" for confidence.

Procedures

Thirty-six participants (24 L1 and 11 L2) were tested in the lab at Beijing Normal University (BNU). Seven additional L2 participants were tested under conditions as similar as possible in the lab at the University of Maryland (UMD). Each participant was seated in front of a computer monitor and fit with an EEG cap. Auditory stimuli were presented using a single high-quality audio monitor (JBL LSR305) placed centrally above the computer monitor.

This experiment was conducted as part of a larger study (see Pelzl, 2018), it followed a lexical decision task (reported in Pelzl et al., 2020), and was itself followed by a picture-word matching task that examined N400 responses for clear lexical violations (details in Pelzl, 2018). For the picture-phonology matching experiment, participants began by completing eight practice items with stimuli not included in the experiment, and then completed 112 picture-phonology matching trials. Stimuli were presented in seven blocks of 16 trials, with self-paced breaks between each block. Trial parameters are illustrated in **Figure 2**. The beginning of each trial was signaled with a 'beep', followed by a fixation cross. After 350 ms, a picture was displayed. Then, after 1.75 s the image was replaced by a fixation cross. Still 250 ms later the auditory stimulus was presented, followed by 1.2 s of silence at which point the fixation cross was replaced by a question prompt: "Did the word match the picture?" (or equivalent in Chinese for L1 participants). After the participant's response, there was a 2 s pause before the next trial began. The entire picture-phonology matching experiment lasted approximately 15 min.

The long display time for the images (1.75 s) was determined after piloting and with the logic that, for this experiment we wanted to maximize the opportunity for L2 learners to recognize images and their associated words. This design allows (but does not compel) participants to utilize explicit knowledge of tones in retrieving target items. The design serves as a proof-of-concept for this approach, testing L2 ability to utilize tone cues under near optimal circumstances.

After completion of the ERP experiments, participants completed the offline vocabulary test.

Vocabulary Posttest

The offline vocabulary posttest produced four data points for each mismatching nonword trial that an L2 participant encountered: an accuracy score for the tones and definition they supplied, a confidence rating for the tones, and a confidence rating for the definitions. Accuracy was scored correct (1) or incorrect (0). For example, if the real word target was 面条 *miantiao*, the only correct response would be "42" (*mian4tiao2* "noodles"). Any deviations from these two tones would be marked incorrect. Definitions were scored similarly using a list of acceptable definitions generated prior to scoring. Confidence ratings were recorded as a number from 0 to 3. One participant's vocabulary test data was lost due to a coding error, leaving a total of 1,024 trials (64 per participant) for the analysis.

Electroencephalogram Recording

Raw electroencephalogram (EEG) was recorded continuously at a sampling rate of 1,000 Hz using a Neuroscan SynAmps data acquisition system and an electrode cap (BNU: Quik-CapEEG; UMD: Electrocap International) mounted with 29 AgCl electrodes at the following sites: *midline*: Fz, FCz, Cz, CPz, Pz, Oz; *lateral*: FP1, F3/4, F7/8 FC3/4, FT7/8, C3/4, T7/8, CP3/4, TP7/8, P4/5, P7/8, and O1/2 (UMD: had FP2, but *no* Oz). Recordings were referenced online to the right mastoid and re-referenced offline to averaged left and right mastoids. The electro-oculogram (EOG) was recorded at four electrode sites: vertical EOG was recorded from electrodes placed above and below the left eye; horizontal EOG was recorded from electrodes situated at the outer canthus of each eye. Electrode impedances were kept below 5kΩ. The EEG and EOG recordings were amplified and digitized online at 1 kHz with a bandpass filter of 0.1–100 Hz.

EEG Data Processing

Consistent with the approach used in the related study reported in Pelzl et al. (2020), data from fifteen central electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4, CP3, CPz, CP4, P3, Pz, P4) were chosen for final analysis. To reduce some mild non-normality in the data, any trial with an absolute value greater than 50 μV was removed prior to final data analysis. Finally, only trials that elicited correct behavioral responses (correct acceptance or correct rejection) were retained for final analysis.

Data from one L1 participant was excluded due to equipment failure. Data from three additional L1 participants and one L2 participant were excluded due to having greater than 40% artifacts on experimental trials (a second L2 participant's data was borderline at 41.67% trials rejected due to artifacts, but was retained due to the difficulty of obtaining advanced L2 data). After excluding these participants, artifact rejection affected 10.55% of experimental trials (L1 8.31%; L2 13.18%). A single average amplitude was obtained for each trial for each electrode for each subject in an early PMN window (200–400 ms) and a later LPC window (400–600 ms). These windows were chosen on the basis of previous research and by visual inspection of grand average waveforms. We recognize the reliance on visual inspection for window selection as a potential limitation, and future work should improve on it in line with advice in Luck and Gaspelin (2017).

After exclusion of incorrect trials, the final PMN dataset contained 42,613 data points (80.0% out of total possible 53,290 data points: L1 = 88.1%; L2 = 70.4%) and the LPC dataset contained 42,610 data points (80.0% out of total possible 53,290 data points. L1 = 88.1%; L2 = 70.4%).

RESULTS

Behavioral Results and Analysis

Reliability for picture-phonology matching data was high (List A: $\alpha = 0.91$; List B: $\alpha = 0.93$; List C: $\alpha = 0.94$). Descriptive results are shown in **Table 2**. Overall, L1 listeners were more accurate than L2 listeners. Whereas L1 listeners were least accurate in judging vowel mismatches, L2 listeners were least accurate in judging tone mismatches.

To further investigate response patterns, we also computed d-prime (d') for each participant, contrasting vowel mismatches with matching real words, and tone mismatches with real words. Laplace smoothing was used to correct for infinite values (Jurafsky and Martin, 2009; Barrios et al., 2016). As with accuracy, d' results suggest overall higher sensitivity to mismatches for L1 listeners, with better scores for tone mismatches compared to vowel mismatches (vowel $d' = 3.49$, sd = 0.49; tone $d' = 3.91$, sd = 0.41). In contrast, L2 had less sensitivity overall, with vowel mismatches detected more readily

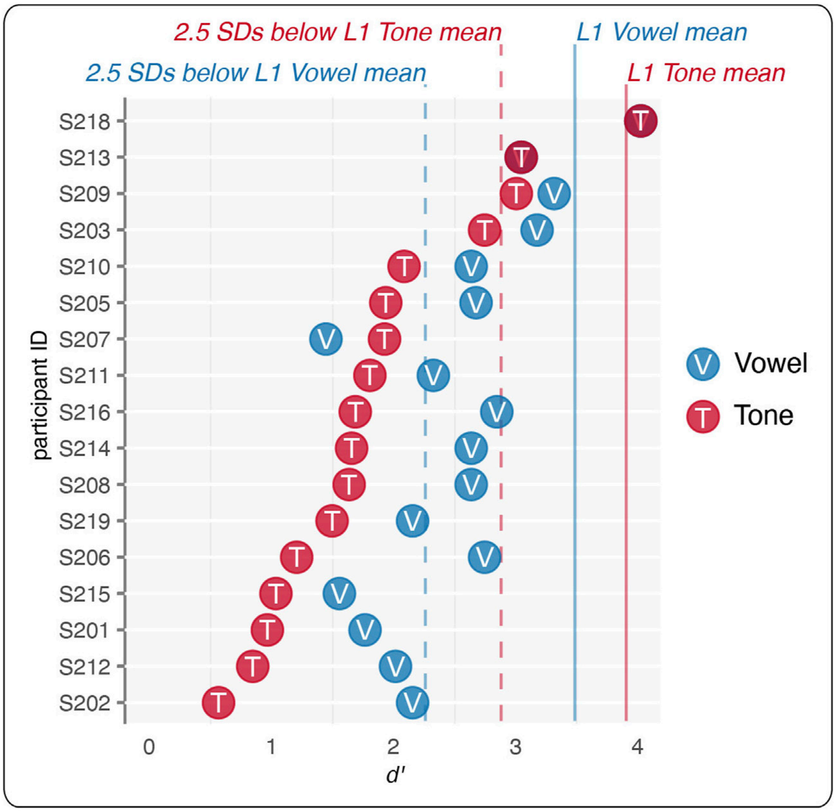


FIGURE 3 | Individual L2 participants' (n = 17) d' results for vowel (V) and tone (T) mismatch conditions in the picture-phonology matching task.

than tone mismatches (vowel d' = 2.54, sd = 0.66; tone d' = 1.87, sd = 0.91).

Figure 3 depicts individual d′ results for each L2 participant. All but three L2 participants had tone d′ values more than 2.5 standard deviations below the L1 mean for tone mismatches, while vowel mismatches were more mixed. More importantly, in all but three cases (S218, S213, and S207), individual L2 participants had lower d′ values for tone mismatches than for vowel mismatches.

All statistical analyses reported below were conducted in *R* (version 4.0.3, R Core Team, 2020). Mixed-effects models were fit using the *lme4* package (version 1.1.21, Bates et al., 2015b). Effects coding was applied using the *mixed* function in *afex* (Singmann et al., 2017). Rather than examining general model outcomes that were not of importance for our research questions (e.g., whether there is a main effect of group or condition), we focus on the specific outcomes of interest (Schad et al., 2020), which we specified using the *multcomp* (Hothorn et al., 2008) and *emmeans* (Lenth, 2020) packages (full model results are reported in the **Supplementary Appendix C**).

Accuracy results were submitted to a mixed-effect logistic regression (using the *bobyqa* optimizer) with crossed random effects for subjects and items. The dependent variable was accuracy (1, 0). Fixed effects included the factors *condition* (real word, tone mismatch, vowel mismatch), and *group* (L1, L2), and their interaction. The maximal random effects model

TABLE 2 | Descriptive accuracy results for picture-phonology matching task.

Group	Condition	Mean acc. % (sd)
L1 (*n* = 20)	Real	97.5 (15.6)
	Vowel	92.7 (26.1)
	tone	98.1 (13.6)
L2 (*n* = 17)	Real	87.5 (33.1)
	Vowel	88.0 (32.5)
	tone	69.1 (46.2)

was fit first (Barr et al., 2013; Bates et al., 2015a). Model convergence difficulties were addressed by suppressing correlations in random effects (using "expand_re = TRUE" in the *mixed* function). The best fitting model was determined by model comparison conducted through likelihood ratio tests, building from the maximal model (which was rejected due to convergence issues) to progressively less complex models. The final model included by-subject random intercepts and slopes for the effect of condition, and by-item random intercepts and slopes for condition and group and their interaction: (*glmer* model formula)*:* accuracy ~ condition * group + (condition | subject) + (condition * group || item).

The critical comparison was whether the L2 group displays a difference between vowel and tone accuracy. To complete the picture, we also examined how this difference compares to the same contrast in the L1 group. Critical comparisons are

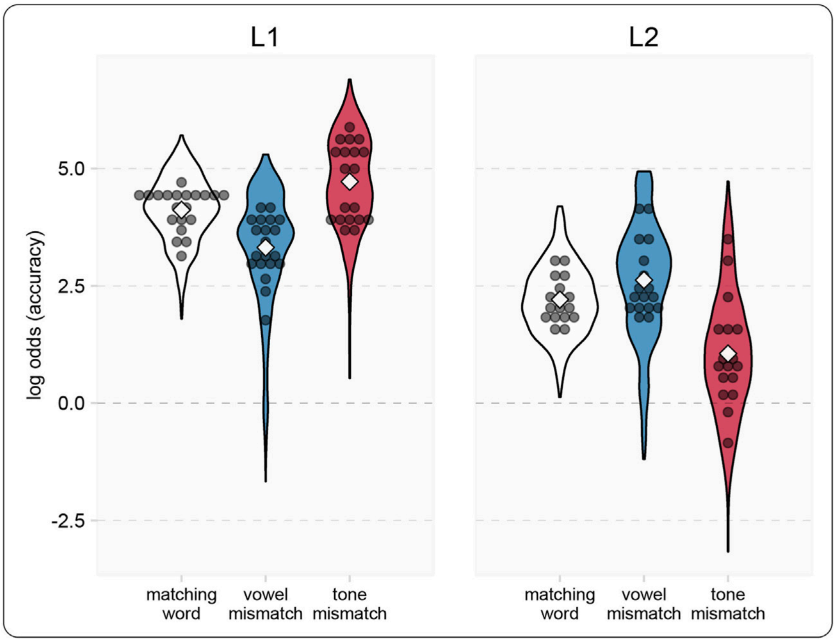


FIGURE 4 | Violin plots of model estimated log odds of a correct response in the picture-phonology matching task. Width of the plot indicates distribution of model estimated responses. White diamonds indicate group means. Gray circles indicate individual participant mean scores. The zero line indicates chance.

TABLE 3 | Critical comparisons for vowel and tone accuracy in the picture-phonology matching task.

Comparison	b	SE	95% CI[a]	Z	p (>/z/)[b]
L1 vowel vs tone	−1.36	0.51	[−2.55, −0.18]	-2.69	0.007
L2 vowel vs tone	1.73	0.36	[0.88, 2.58]	4.75	<0.001
L1 vs L2: Vowel vs tone	−3.09	0.54	[−4.35, −1.83]	-5.72	<0.001

[a]*Asymptotic confidence intervals.*
[b]*Adjusted using Bonferroni-Holm method.*

summarized in **Table 3**, and model results are depicted in **Figure 4**. L2 listeners were significantly more accurate in rejection of vowel mismatches than of tone mismatches. They were about two and a half times more likely to incorrectly accept tone mismatches than vowel mismatches (30.9/12 = 2.6). There was also a statistically significant difference in accuracy between mismatch conditions for the L1 group, with L1 more accurate for tone than vowel mismatches. Compared to L1, the accuracy difference between mismatch conditions for L2 was larger and in the opposite direction.

In summary, whereas L1 listeners had more difficulty detecting vowel mismatches than tone mismatches, L2 listeners had more difficulty detecting tone mismatches than vowel mismatches.

Table 4 displays descriptive results for the offline vocabulary test, along with related accuracy for those items in the picture-phonology matching task. We find that, overall, L2 learners were quite confident of the definitions they provided (mostly ratings of high or mid confidence), and that higher confidence appears to relate strongly to the accuracy of those definitions. In other words, learners know which words they know and which they do not. Learners' confidence ratings for their explicit tone knowledge indicate less certainty for tones than for definitions. Although overall accuracy for tones is lower than for definitions, it still does appear to track with confidence ratings. That is, L2 learners generally know which tones they know and which they do not. However, even for the tones they know most confidently, they are still inaccurate for more than one in ten of those tones (*mean* = 86% when counting nonword conditions together). Whereas accurate knowledge of definitions always appears to impact performance in the picture-phonology matching task, accurate knowledge of tones appears to relate only to tone nonword items. This makes sense, as tone knowledge is largely irrelevant for vowel nonword items.

Descriptively, then, we find that L2 offline knowledge suggests some difficulties in accurate encoding of tones in explicit lexical representations, and that this appears to impact accuracy for correct rejection of tone mismatches.

As in Pelzl et al. (2020), we conducted an exploratory "Best case Scenario" analysis using only the subset of trials that targeted nonwords for which an L2 participant had indicated correct and confident knowledge (confidence rating = 3) of both tones and definitions for the real word counterparts. This comprised 256 tone nonword and 255 vowel nonword trials (511 total, 47% of total mismatch trial data). Mean accuracy for vowel nonwords was 93% (sd = 26%); mean accuracy for tone nonwords was 80% (sd = 40%). The accuracy results were submitted to a generalized linear mixed effects model following procedures outlined for previous analyses. The model included the fixed effect of nonword condition. The maximal model was fit, and included

TABLE 4 | Results of offline vocabulary test requiring L2 participants (n = 16) to supply tones, definitions, and confidence ratings for the real word counterparts of critical mismatching nonwords. Tone accuracy indicates whether supplied tones were correct. Picture-phonology (pic-pho) accuracy indicates whether the related nonwords were correctly rejected in the matching task.

	condition	conf. rating	k (items)	definition acc. %	pic-pho acc. %
confidence ratings and accuracy of L2 supplied definitions	vowel nonword	3 (high)	465	98	90
		2 (mid)	27	81	78
		1 (low)	7	43	57
	tone nonword	3 (high)	470	99	70
		2 (mid)	23	83	61
		1 (low)	5	4	40
	condition	conf. rating	k (items)	tone acc. %	pic-pho acc. %
confidence ratings and accuracy of L2 supplied tones	vowel nonword	3 (high)	300	87	92
		2 (mid)	170	52	84
		1 (low)	29	28	83
	tone nonword	3 (high)	309	84	77
		2 (mid)	163	50	57
		1 (low)	35	37	51

random intercepts for subjects and items, and random slopes for the by-subject and by-item effects of condition. There was a significant difference in accuracy for tone and vowel nonwords ($b = -7.71$, $SE = 3.06$, 95%, $z = -2.51$, $p = 0.012$).

In summary, after accounting for offline L2 word knowledge and subjective confidence of that knowledge, L2 learners still showed a more limited ability to reject tone mismatches than vowel mismatches. At the same time, we should not ignore the observable improvement that occurred when results were limited to items known correctly and with certainty. Accuracy for vowel mismatches rose from 88 to 93%, and for tone mismatches the increase was even greater, from 69 to 80%, indicating that—at least among this group of learners—eliminating fuzzy (incorrect and uncertain) lexical representations appears to partially account for performance deficits for both tones and vowels.

ERP Results and Analyses for Phonological Mismatch Negativity and Late Positive Component Windows

Mean amplitudes for ERP responses in the time windows for the PMN (200–400 ms) and LPC (400–600 ms) are displayed in **Table 5**. Grand average ERP waveforms are depicted in **Figure 5**. The L1 group appears to have strong negativities for vowel mismatches in the PMN window; though L2 responses are more positive overall, over centro-posterior electrodes the same pattern holds, with vowel mismatches showing the most negative amplitude among condition. In the LPC window, responses for real words are most negative (least positive), followed by vowel mismatches, with tone mismatch responses being the most positive. While there are differences in absolute amplitudes between groups, over centro-posterior electrodes the overall ordering of responses (real, vowel, tone) is similar within L1 and L2 groups.

Average amplitudes for correct trials in the PMN and LPC windows were submitted to linear mixed-effects regression model with crossed random effects for subjects and items. Fixed effects were *condition* (match, mismatch) and *group* (L1, L2) and their interaction. Convergence difficulties were addressed by specifying uncorrelated random effects. Effects coding was used, and p-values were obtained using Satterthwaite's method. The maximal model that successfully converged was fit first and was then compared to less complex models to test random effects. The final maximal models for both data sets were parallel, and included random slopes for subjects and items, with electrodes nested under subjects. The models also included by-item random.

Though our primary interest in this study is in L2 sensitivity to vowels and tones, in order to evaluate L2 responses, we need to compare them to an L1 baseline. To this end, we report critical comparisons for three relevant contrasts (matching word vs. vowel mismatch, matching word vs. tone mismatch, vowel mismatch vs. tone mismatch) within and between L1 and L2 groups. Results for the PMN window are shown in **Table 6**, and depicted in violin plots in **Figure 6**. For both the L1 and L2 group, responses to vowel mismatches were significantly more negative than both matching word and tone mismatch responses. Despite the similar overall pattern of their responses, there were interactions between group and condition. For the L1 group, the magnitude of differences for the matching word vs. vowel mismatch and vowel vs. tone mismatch were significantly larger than the same contrasts for L2 participants. In other words, though L1 vowel mismatch responses were stronger overall, the same pattern of responses applied for both groups, with *neither* group showing strong PMN deflections for tone mismatches.

Results for the LPC window are shown in **Table 7**, and depicted in violin plots in **Figure 7**. For both the L1 and L2 group, responses to tone mismatches were significantly more positive than responses to matching words. For the L1, but not the L2 group, tone mismatch responses were more positive than vowel mismatch responses. Vowel mismatch responses did not differ significantly from matching word responses. Interactions between groups and conditions There was a significant interaction between group and condition for the contrast of matching words vs vowel mismatches with the L2 difference being larger than the L1 difference. There was

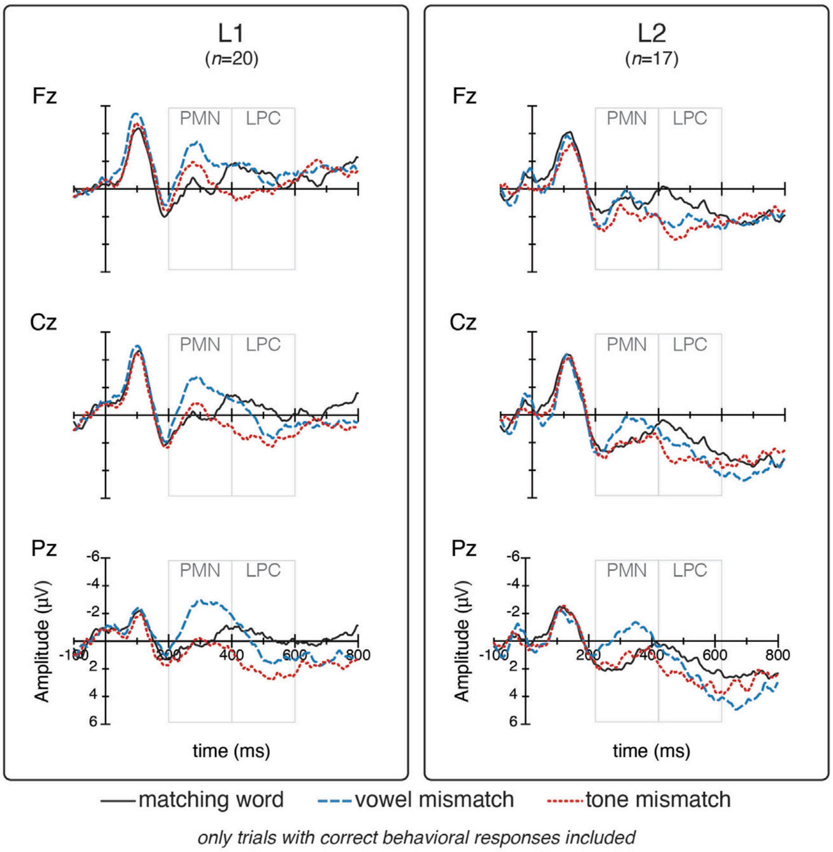


FIGURE 5 | Grand average waveforms for L1 and L2 participants. Time windows highlighted for PMN (200–400) and LPC (400–600). (Waveforms for all 15 electrodes are available in the **Supplementary Appendix D**).

TABLE 5 | Mean amplitude (in µV) and standard error (SE) of PMN and LPC responses (correct trials only).

		PMN		*LPC*	
group	**condition**	**mean amp**	**(SE)**	**mean amp**	**(SE)**
L1	Real	0.24	(0.10)	−0.37	(0.11)
L1	Vowel	−1.66	(0.10)	−0.13	(0.12)
L1	Tone	0.20	(0.10)	1.39	(0.11)
L2	Real	1.35	(0.11)	0.66	(0.12)
L2	Vowel	0.20	(0.11)	1.52	(0.13)
L2	tone	1.69	(0.13)	2.53	(0.15)

also a significant interaction between groups and vowel vs. tone mismatches, the size of the difference being larger for L1 than for L2 responses. Confidence intervals for all comparisons suggest some imprecision and so results should be interpreted with appropriate caution.

DISCUSSION

In order to test advanced L2 Mandarin learners' sensitivity to lexical tones, we conducted a picture-phonology matching task with L1 and L2 speakers of Mandarin. Key results can be summarized as follows. 1) L2 participants were less accurate at rejecting tone mismatches than vowel mismatches—the opposite pattern from L1 participants who were more accurate in all conditions overall, but less accurate for vowel than tone mismatches. 2) After limiting the analysis to trials for words L2 participants knew correctly and confidently, their accuracy for both tone and vowel mismatch trials increased, but tone mismatch trials still remained significantly less accurate than vowel mismatch trials. For ERP results, which targeted only trials with correct behavioral responses, 3) in the early PMN window, both L1 and L2 listeners displayed significantly more negative responses to vowel mismatches, than to either matching words or tone mismatches. Though there were differences in the magnitude of effects between L1 and L2, the overall patterning of responses was similar. 4) In the later LPC window, both groups displayed strong positive responses following tone mismatches, with some differences in the magnitude of responses to vowel mismatches relative to tone mismatches and real words. Below, we discuss these results in more detail, while also connecting them to broader discussions of L2 tone word learning and fuzzy L2 lexical representations.

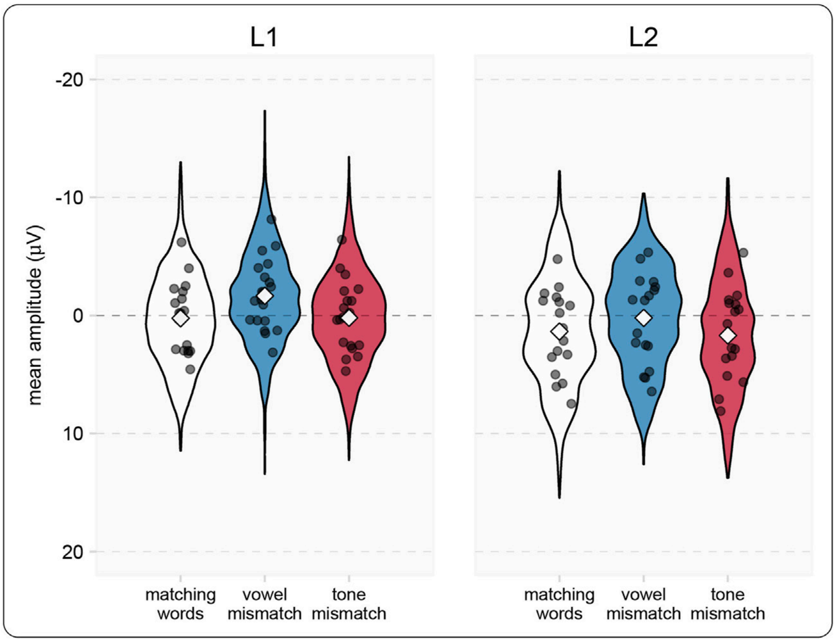


FIGURE 6 | Model estimates for PMN (200–400 ms) amplitude in the picture-phonology matching task (correct trials only). White diamonds indicated model estimated group means for each condition, with shaded areas representing the distribution of estimated responses. Each gray dot indicates an individual participant's mean amplitude in the condition.

TABLE 6 | ERP results and analyses for PMN window (200–400 ms).

Comparison	b	SE	95% CI[a]	Z	p (>/z/)[b]
L1 match vs vowel	2.10	0.38	[1.08, 3.12]	5.52	<0.001
L1 match vs tone	0.05	0.34	[−0.86, 0.96]	0.14	1.000
L1 vowel vs tone	−2.05	0.37	[−3.05, −1.04]	−5.48	<0.001
L2 match vs vowel	1.17	0.39	[0.13, 2.21]	3.03	0.010
L2 match vs tone	−0.14	0.36	[−1.09, 0.82]	−0.39	1.000
L2 vowel vs tone	1.31	0.39	[0.27, 2.35]	3.38	0.004
L1 vs L2: Match vs vowel	0.92	0.20	[0.40, 1.45]	4.74	<0.001
L1 vs L2: Match vs tone	0.19	0.21	[−0.37, 0.74]	0.90	1.000
L1 vs L2: Vowel vs tone	−0.74	0.21	[−1.30, −0.18]	-3.56	0.002

[a]*Asymptotic confidence intervals.*
[b]*Adjusted using Bonferroni-Holm method.*

Tones Are Difficult (Again)

Our results echo those seen in previous studies, indicating that—for nontonal L1 speakers—mastery of tone words is a major L2 learning challenge (Han and Tsukada, 2020; Ling and Grüter, 2020; Pelzl et al., 2019; Pelzl et al., 2020). Given the nature of the picture-phonology matching task, the present results are perhaps the clearest indication yet of how difficult L2 tone word recognition is. As noted above, the picture-phonology matching task was less demanding than previously used lexical decision tasks. As long as a person knew the pictured word and its tones, they could directly judge whether the target matched or not. There was no need to search their mental lexicon to verify the *absence* of a nonword. Nevertheless, we found that the L2 group made errors on 31% of tone mismatch trials overall, compared to 12% of vowel mismatch trials. When we limited consideration to correctly and confidently known words, they still made errors on 20% of tone mismatch trials. In other words, for these L2 participants, explicit knowledge of tone words only accounted for, at most, one-third of their tone errors.

Three General Accounts of Tone Difficulties

As outlined in our introduction (**Figure 1**), there are three broad accounts that could uniquely or jointly explain these outcomes, positing perception, processing, or representation as the locus of L2 tone word breakdowns. Present results do not allow us to determine the relative contribution of these accounts to lexical tone learning difficulties. At the same time, they do suggest directions for future study.

First, though the present experiment did not directly test auditory perception, the overall accuracy for tones after limiting analysis to correctly and confidently known words (mean = 80%) bears a striking resemblance to the same L2 participants' overall accuracy for tone identification in disyllabic contexts (*mean* = 82%, see **Table 1**). In other words, it may well be the case that, once explicit knowledge of tones has been established, L2 listeners' remaining tonal difficulties are due primarily to difficulties perceiving tones faithfully in *multisyllabic strings*. Since our first study (Pelzl et al., 2019), the particular difficulty of disyllabic, as opposed to monosyllabic tone words, has remained an open question. Studies with naïve, novice, and intermediate proficiency L2 participants have reported this pattern in tone category identification tasks (Broselow et al., 1987; Hao, 2018; Sun, 1998; see also; Chan and Leung, 2020).

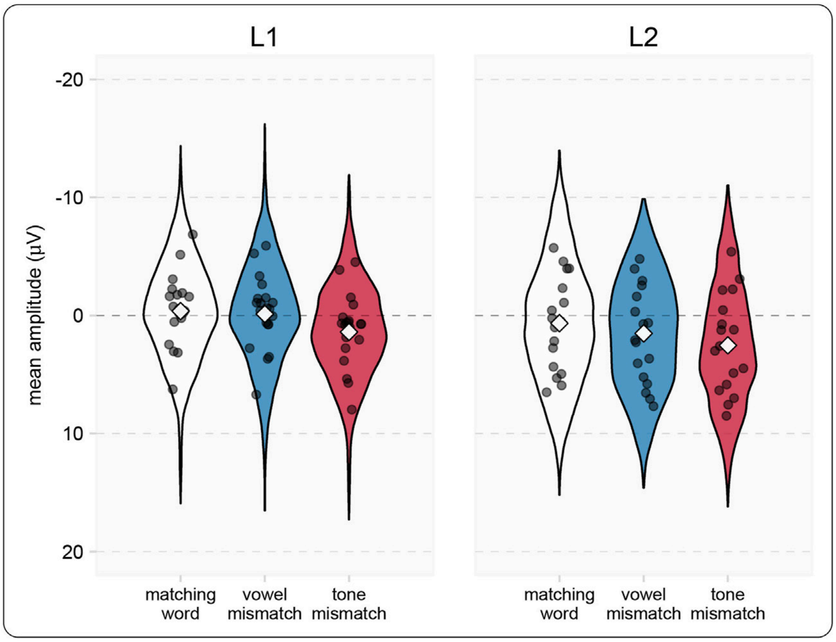


FIGURE 7 | Model estimates for LPC (400–600 ms) amplitude in the picture-phonology matching task (correct trials only). White diamonds indicated model estimated group means for each condition, with shaded areas representing the distribution of estimated responses. Gray dots indicate individual participants mean amplitude in the condition.

TABLE 7 | ERP results and analyses for LPC *window (400–600)*.

Comparison	b	SE	95% CI[a]	Z	p (>/z/)[b]
L1 match vs vowel	0.01	0.44	[−1.17, 1.19]	0.03	1.000
L1 match vs tone	−1.71	0.36	[−2.68, −0.74]	−4.71	<0.001
L1 vowel vs tone	−1.72	0.41	[−2.83, −0.61]	−4.16	<0.001
L2 match vs vowel	−0.95	0.45	[−2.15, 0.25]	−2.13	0.133
L2 match vs tone	−1.57	0.38	[−2.59, −0.54]	−4.09	<0.001
L2 vowel vs tone	0.61	0.43	[−0.54, 1.77]	1.43	0.462
L1vs L2: Match vs vowel	0.96	0.22	[0.38, 1.55]	4.42	<0.001
L1vs L2: Match vs tone	−0.15	0.23	[−0.76, 0.47]	−0.63	1.000
L1vs L2: Vowel vs tone	−1.11	0.23	[−1.73, −0.49]	−4.78	<0.001

[a]*Asymptotic confidence intervals.*
[b]*Adjusted using Bonferroni-Holm method.*

In a tone word training study with naive learners, Chang and Bowles (2015) found disyllabic words to be much more challenging than monosyllabic words. That longer strings of syllables are more difficult is not surprising in and of itself, but the exact cause of the difficulty remains unclear. In Pelzl et al. (2019), our tone identification task used monosyllables clipped from context, thus preserving the coarticulation of the tones, but removing the potentially useful contextual cues provided by neighboring syllables—and L2 participants performed with near-native accuracy on all but Tone 2. In contrast, the tone identification, lexical decision and picture-phonology matching tasks used with the participants in the present study (and in Pelzl et al., 2020) were produced more slowly and clearly, and all maintained the contextual cues, nevertheless, most L2 learners showed some difficulties. Future research will need to examine additional factors that may be impacting multisyllable tone perception, such as memory constraints (e.g., *the phonological loop* Baddeley, 1968), L1 prosodic biases that might operate across multiple syllables (Braun et al., 2014; Braun and Johnson, 2011; Schaefer and Darcy, 2014; So and Best, 2010; So and Best 2014), and potential ordering effects in the perception of co-articulated tones (Xu, 1994; Xu, 1997).

Second, the phonolexical processing account can also naturally explain the incorrect responses on trials where participants reported correct and confident knowledge in the offline task. Despite having explicit knowledge of the pictured words, L2 listeners may have occasionally allowed their native processing biases to take over, ignoring tonal cues as they accessed words.

We have argued elsewhere (Pelzl et al., 2020) that tones are often redundant with other available cues. Most Mandarin words are longer than a single syllable, making the likelihood of a plausible minimal tone pair competitor low. Perhaps more importantly, the broader context will usually guide interpretation of what is heard. SLA scholars have long noted the difficulties associated with redundant cues in an L2 (e.g., VanPatten, 1996; DeKeyser, 2005). Insofar as tones are redundant with other available cues, recent discussion of the phenomena of *unlearning* and *blocking* may provide insights to the source of L2 failures to learn them (see, especially, Nixon, 2020; but also Ellis, 2006; MacWhinney and Bates, 1989). First, through long experience with a non-tonal L1, L2 Mandarin listeners have unlearned tone cues—that is, they have learned through negative evidence that F0 height and shape on vowels/syllables is not informative for speech comprehension, and thus have down-weighted such cues. When confronted with new F0 cues in the L2, they need to re-weight these cues appropriately, but because tones typically co-occur with other disambiguating cues, there is little opportunity for prediction error to guide this re-weighting process. This leaves primarily statistical learning mechanisms to guide the development of L2 tone processing. Indeed, statistical learning mechanisms have been shown in L2 tone learning for *highly frequent* syllable + tone co-occurrence probabilities (Wiener et al., 2018).

It has also been proposed that as vocabulary size increases, minimal pairs might push learners toward more sensitivity for difficult L2 contrasts (cf. discussion of L2LP model in Wiener et al., 2019; see also Bundgaard-Nielsen et al., 2011; Llompart and Reinisch, 2020). While not denying that minimal pairs *could* play a role in improved L2 outcomes, our own work so far has given us a rather pessimistic view of the strength of minimal pairs in typical L2 tone learning. Though it is not difficult to find tonal minimal pairs in Mandarin if one goes looking for them, for L2 learners these pairs accrue very gradually over time, and it is likely that many other developing L2 abilities will allow learners to further capitalize on contextual cues to the detriment of tones. Thus, it may be that, for most L2 learners, only the most frequent tone words will ever be processed phonolexically.

Returning to present results, with respect to the representational account, if the explicit knowledge of words directly captures how those words are encoded in phonolexical representations, the representational account cannot explain persistent L2 difficulties for correctly and confidently known words in the present data (note, this would be true for vowel mismatches as well). Still, as we will consider in more detail below, the representational account cannot be fully rejected as a contributor to the current pattern of results, as it could be that explicit knowledge of tone words was not the main source L2 listeners drew upon when judging whether a tone word matched a picture.

ERPs

Results from our PMN analysis suggest that, for *correctly* judged trials, both L1 and L2 listeners had the strongest (most negative) response to vowel mismatches. This suggests that L2 listeners are able to generate phonological expectations based on context, at least when there is plenty of time available to do so after the context appears, as in the current study. At the same time, PMN results failed to show significant differences between tone mismatches and real words, suggesting that mismatching vowels may affect ERPs earlier than tones for *all* listeners. This pattern of results is consistent with several previous studies which provided contextual cues for lexical expectation in phrases or sentences and found reduced N400s for tone mismatches relative to rhyme mismatches (Hu et al., 2012; Pelzl et al., 2021; Zou et al., 2020), though such differences do not always appear (Brown-Schmidt and Canseco-Gonzalez, 2004; Schirmer et al., 2005; Pelzl et al., 2019). On the other hand, as we expected for these correctly judged trials, in the later LPC time window both groups displayed strong positive deflections for tone mismatches relative to the matched word condition. In fact, at the LPC the tone mismatch response was significantly larger than the vowel mismatch response for the L1 speakers.

The seemingly similar delayed response to tone mismatches relative to vowel mismatches across groups might be tied to the nature of tone contrasts, especially as they occur in contextualized syllables. For many trials, it may be that in order to identify the F0 contour as it unfolds over time, more of the syllable needs to be available than in the case of vowels; sometimes listeners may even need the contextual information of the following syllable in order to make the identity of the tone unambiguously wrong (cf. J. Huang and Holt, 2009). This extended perceptual analysis for tones could be too late to impact the early phonological perception computations that may be driving the phonological mismatch negativity. In contrast, mismatching vowels reveal themselves almost immediately, which could drive a stronger negativity across the PMN and the N400 time-windows.

It is also worth noting, however, that visual examination of anterior electrodes indicates a numerical trend toward a PMN for tone mismatches in the L1 group that is not visible in the L2 group (**Figure 5**; see also **Supplementary Appendix D** for waveforms of all 15 tested electrodes). Therefore, it could be that we did not have enough power to reliably detect differences between real word and tone mismatch responses. Perhaps the nesting of electrodes in our models (rather than testing electrode locations as a fixed effect) washed out effects that were more prominent at some sites than others. That is, had we targeted only frontal electrodes, an L1 PMN for tones would have been observed. Perhaps the large LPC observed for tones in the L1 group actually began early enough to wash out PMN and N400 effects at posterior electrodes. In contrast, significant L2 PMNs for tones seem less much weaker, regardless of electrode site. If this were the case, it would constitute some evidence in favor of a different processing timeline for L1 and L2 listeners. However, further targeted investigation with new datasets would be needed to draw any such conclusions[4].

Regardless of how we understand the group differences in the PMN window, the later LPCs indicate that for both groups, correctly rejected trials ultimately lead to the same process of repair or reanalysis. The subtle differences between the patterns observed at the LPC, however, are intriguing. As in several prior ERP studies (Pelzl, 2018; Pelzl et al., 2021), in the L1 group we observed a slightly larger LPC for tone mismatches than vowel mismatches. In the L2 group there was no such tendency. Although we had no predictions for group differences for correct trials in this later time window, it is tempting to speculate that the slightly larger LPC for tone

[4]Another explanation for between group differences might be the greater number of correctly judged trials available for analysis. Since L1 responded correctly more often, the magnitude of PMN responses was greater. However, this doesn't fit with the response patterns of the L1 group itself, as they responded incorrectly to more vowel trials than tone trials.

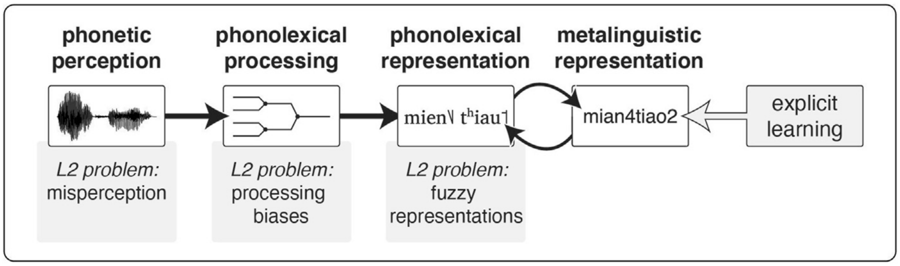


FIGURE 8 | Expanded problem space for L2 tone word recognition. Explicit learning of tones in relation to words may result in separate metalinguistic representations that interact with, but do not necessarily directly reflect the information encoded in phonolexical representations.

mismatches in L1 is a reflection of stronger sensitivity to them—perhaps increased attempts by L1 listeners to reanalyze the input or consider alternative explanations for the unexpected tone. This may be an interesting avenue for future investigation.

Fuzzy Tone Word Representations and Metalinguistic Tone Word Knowledge

Present results once again demonstrate that L2 tone words fit well under the umbrella of fuzzy L2 lexical representations. As with some other L2 instances of fuzzy lexical representations, the fuzziness can be directly linked to the difficulty of novel L2 speech sounds (Broersma and Cutler, 2008; Broersma and Cutler, 2011; Broersma, 2012; Díaz et al., 2012). However, L2 tone word difficulties may also be somewhat unique. Rather than competing with existing L1 phonological contrasts, tones may exist outside the native language phonological space, requiring learners to use F0 cues in a new way. For this reason, some of the fuzzy lexical effects found in L2 tone studies may be qualitatively different from those documented in other L2 contexts. In particular, there is a possibility for metalinguistic tone knowledge to play a very strong role in L2 tone word recognition.

As in other areas of L2 learning, the contrast between implicit and explicit knowledge might be a key for understanding L2 tone word outcomes (DeKeyser, 2003; Suzuki and DeKeyser, 2017). Whereas L2 learners spend great effort to establish metalinguistic tone word representations (encoded in writing via Pinyin romanization), these metalinguistic representations may be a separate form of knowledge that is not automatically drawn upon during word recognition. This is depicted in **Figure 8**. While implicit (fuzzy) tone word lexical representations guide L2 word recognition in the earliest, automatic stages, the metalinguistic tone word representations can serve to identify words (with effort) in later stages. While most often the implicit and metalinguistic representations will be aligned, occasionally, it may happen that despite correct explicit word knowledge, L2 speakers might still have weakly developed implicit phonolexical representations. As these fuzzy representations take the lead during word recognition, they can lead to occasional behavioral errors, even in tasks that allow learners to draw heavily on their explicit knowledge.

LIMITATIONS

While results of the present study are consistent with previous work showing weaknesses in advanced L2 tone perception, we acknowledge some clear limitations. First, the sample of participants, especially L2 participants, was relatively small. Advanced L2 Mandarin learners are difficult to find, but this practical consideration does not affect the statistical facts: it certainly could be the case that we had insufficient power to detect smaller differences between groups and/or conditions, especially for ERP outcomes. Though difficult, it is worth striving to improve in this regard in future work (Brysbaert, 2020).

Second, present results may have been impacted by an ordering effect. As part of a larger set of experiments, the picture-phonology matching task always followed a lexical decision task (see **Supplementary Appendix A**). No stimuli were shared between the lexical decision and picture-phonology matching experiments, but it is possible that L2 participants were more aware of tones in the picture-phonology experiment as they had already experienced the lexical decision task. We did not consider this a problem for the current study, where we aimed to give L2 learners the best chance possible to succeed at the task.

CONCLUSION

This study provides converging evidence of weaknesses in tone word recognition by advanced L2 learners. Learners have clear difficulty in encoding tones in explicit long-term memory, and "Best Case Scenario" results suggest that, even when they do succeed in encoding tones, they do not always succeed at utilizing tones during online Mandarin word recognition. ERP results suggested L1 listeners use early sensitivity to phonological cues to successfully reject mismatching vowels, but there was no clear evidence of other ERP effects in either the L1 or L2 group.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/3aue9/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board at the University of Maryland, College Park. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

EP, EL, and RD contributed to conception and design of the study. EP and TG conducted the research in Beijing. EP conducted the research at Maryland. EP processed the data, performed the statistical analysis, and wrote the first draft of the manuscript. EL contributed sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This research was supported in part by NSF-IGERT grant 0801465 and Ling-DDRI grant 1728851 to pay for data collection. Open access publication fee provided through the NSF SBE Postdoctoral Research Fellowship under Grant No. 2004279. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.689423/full#supplementary-material

REFERENCES

Alexander, J. A., Wong, P. C., and Bradlow, A. R. (2005). Lexical Tone Perception in Musicians and Non-musicians. *Interspeech*, 397–400. Available at: http://groups.linguistics.northwestern.edu/speech_comm_group/publications/2005/Alexander-Wong-Bradlow-2005.pdf (Accessed May 13, 2015).

Amengual, M. (2016). The Perception of Language-specific Phonetic Categories Does Not Guarantee Accurate Phonological Representations in the Lexicon of Early Bilinguals. *Appl. Psycholinguistics* 37 (05), 1221–1251. doi:10.1017/S0142716415000557

Baddeley, A. D. (1968). How Does Acoustic Similarity Influence Short-Term Memory? *Q. J. Exp. Psychol.* 20 (3), 249–264. doi:10.1080/14640746808400159

Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *J. Mem. Lang.* 68 (3), 255–278. doi:10.1016/j.jml.2012.11.001

Barrios, S. L., Namyst, A. M., Lau, E. F., Feldman, N. H., and Idsardi, W. J. (2016). Establishing New Mappings between Familiar Phones: Neural and Behavioral Evidence for Early Automatic Processing of Nonnative Contrasts. *Front. Psychol.* 7, 1–16. doi:10.3389/fpsyg.2016.00995

Bates, D., Kliegl, R., Vasishth, S., and Baayen, H. (2015a). Parsimonious Mixed Models. ArXiv:1506.04967 [Stat]. Available at: http://arxiv.org/abs/1506.04967 (Accessed May 4, 2018). doi:10.5821/palimpsesto.14.4719

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015b). Fitting Linear Mixed-Effects Models Usinglme4. *J. Stat. Soft.* 67 (1), 1–48. doi:10.18637/jss.v067.i01

Bent, T., Bradlow, A. R., and Wright, B. A. (2006). The Influence of Linguistic Experience on the Cognitive Processing of Pitch in Speech and Nonspeech Sounds. *J. Exp. Psychol. Hum. Perception Perform.* 32 (1), 97–103. doi:10.1037/0096-1523.32.1.97

Best, C. T., and Tyler, M. D. (2007). "Nonnative and Second-Language Speech Perception," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*. Editors O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 13–34. doi:10.1075/lllt.17.07bes

Braun, B., Galts, T., and Kabak, B. (2014). Lexical Encoding of L2 Tones: The Role of L1 Stress, Pitch Accent and Intonation. *Second Lang. Res.* 30 (3), 323–350. doi:10.1177/0267658313510926

Braun, B., and Johnson, E. K. (2011). Question or Tone 2? How Language Experience and Linguistic Function Guide Pitch Processing. *J. Phonetics* 39, 585–594. doi:10.1016/j.wocn.2011.06.002

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., and Lepage, M. (2010). The Bank of Standardized Stimuli (BOSS), a New Set of 480 Normative Photos of Objects to Be Used as Visual Stimuli in Cognitive Research. *PLoS ONE* 5 (5), e10773. doi:10.1371/journal.pone.0010773

Broersma, M., and Cutler, A. (2011). Competition Dynamics of Second-Language Listening. *Q. J. Exp. Psychol.* 64 (1), 74–95. doi:10.1080/17470218.2010.499174

Broersma, M., and Cutler, A. (2008). Phantom Word Activation in L2. *System* 36, 22–34. doi:10.1016/j.system.2007.11.003

Broersma, M. (2012). Increased Lexical Activation and Reduced Competition in Second-Language Listening. *Lang. Cogn. Process.* 27 (7–8), 1205–1224. doi:10.1080/01690965.2012.660170

Broselow, E., Hurtig, R. R., and Ringen, C. (1987). "The Perception of Second Language Prosody," in *Interlanguage Phonology: The Acquisition of a Second Language Sound System*. Editors G. Ioup and S. H. Weinberger (New York, NY: Newbury House Publishers), 350–364.

Brown-Schmidt, S., and Canseco-Gonzalez, E. (2004). Who Do You Love, Your Mother or Your Horse? An Event-Related Brain Potential Analysis of Tone Processing in Mandarin Chinese. *J. Psycholinguist Res.* 33 (2), 103–135. doi:10.1023/b:jopr.0000017223.98667.10

Brunellière, A., and Soto-Faraco, S. (2015). The Interplay between Semantic and Phonological Constraints during Spoken-word Comprehension. *Psychophysiol.* 52 (1), 46–58. doi:10.1111/psyp.12285

Brysbaert, M. (2020). Power Considerations in Bilingualism Research: Time to Step up Our Game. *Bilingualism* 1, 1–6. doi:10.1017/S1366728920000437

Bundgaard-Nielsen, R. L., Best, C. T., and Tyler, M. D. (2011). Vocabulary Size Matters: The Assimilation of Second-Language Australian English Vowels to First-Language Japanese Vowel Categories. *Appl. Psycholinguistics* 32 (1), 51–67. doi:10.1017/S0142716410000287

Cai, Q., and Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLoS One* 5 (6), e10729. doi:10.1371/journal.pone.0010729

Chan, R. K., and Leung, J. H. (2020). Why Are Lexical Tones Difficult to Learn? *Stud. Second Lang. Acquis.* 42 (1), 33–59. doi:10.1017/S0272263119000482

Chandrasekaran, B., Sampath, P. D., and Wong, P. C. M. (2010). Individual Variability in Cue-Weighting and Lexical Tone Learning. *J. Acoust. Soc. America* 128 (1), 456–465. doi:10.1121/1.3445785

Chang, C. B., and Bowles, A. R. (2015). Context Effects on Second-Language Learning of Tonal Contrasts. *J. Acoust. Soc. America* 138 (6), 3703–3716. doi:10.1121/1.4937612

Chang, C. B. (2018). Perceptual Attention as the Locus of Transfer to Nonnative Speech Perception. *J. Phonetics* 68, 85–102. doi:10.1016/j.wocn.2018.03.003

Chrabaszcz, A., and Gor, K. (2014). Context Effects in the Processing of Phonolexical Ambiguity in L2. *Lang. Learn.* 64 (3), 415–455. doi:10.1111/lang.12063

Connolly, J. F., and Phillips, N. A. (1994). Event-related Potential Components Reflect Phonological and Semantic Processing of the Terminal Word of Spoken Sentences. *J. Cogn. Neurosci.* 6 (3), 256–266. doi:10.1162/jocn.1994.6.3.256

Cook, S. V., and Gor, K. (2015). Lexical Access in L2. *Ml* 10 (2), 247–270. doi:10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-To-Meaning Mappings. *Front. Psychol.* 7, 1–17. doi:10.3389/fpsyg.2016.01345

Coulson, S., King, J. W., and Kutas, M. (1998). ERPs and Domain Specificity: Beating a Straw Horse. *Lang. Cogn. Process.* 13 (6), 653–672. doi:10.1080/016909698386410

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric Lexical Access and Fuzzy Lexical Representations in Second Language Learners. *Ml* 8 (3), 372–420. doi:10.1075/ml.8.3.06dar

DeKeyser, R. M. (2003). "Implicit and Explicit Learning," in *The Handbook of Second Language Acquisition*. Editors C. J. Doughty and M. H. Long (Malden, MA: Blackwell Publishing), 313–348.

DeKeyser, R. M. (2005). What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Lang. Learn.* 55 (S1), 1–25. doi:10.1111/j.0023-8333.2005.00294.x

Desroches, A. S., Newman, R. L., and Joanisse, M. F. (2009). Investigating the Time Course of Spoken Word Recognition: Electrophysiological Evidence for the Influences of Phonological Similarity. *J. Cogn. Neurosci.* 21 (10), 1893–1906. doi:10.1162/jocn.2008.21142

Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). Individual Differences in Late Bilinguals' L2 Phonological Processes: From Acoustic-Phonetic Analysis to Lexical Access. *Learn. Individual Differences* 22 (6), 680–689. doi:10.1016/j.lindif.2012.05.005

Diependaele, K., Lemhöfer, K., and Brysbaert, M. (2013). The Word Frequency Effect in First- and Second-Language Word Recognition: A Lexical Entrenchment Account. *Q. J. Exp. Psychol.* 66 (5), 843–863. doi:10.1080/17470218.2012.720994

Ellis, N. C. (2006). Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Appl. Linguistics* 27 (2), 164–194. doi:10.1093/applin/aml015

Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 Speech Perception Research and Phonological Theory. *Stud. Sec. Lang. Acq.* 26 (04). doi:10.1017/S0272263104040021

Flege, J. E. (1995). "Second Language Speech Learning: Theory, Findings, and Problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Editor W. Strange (Timonium, MD: York), 233–277.

Gor, K., and Cook, S. V. (2020). A Mare in a Pub? Nonnative Facilitation in Phonological Priming. *Second Lang. Res.* 36 (1), 123–140. doi:10.1177/0267658318769962

Gouvea, A. C., Phillips, C., Kazanina, N., and Poeppel, D. (2010). The Linguistic Processes Underlying the P600. *Lang. Cogn. Process.* 25 (2), 149–188. doi:10.1080/01690960902965951

Han, J.-I., and Tsukada, K. (2020). Lexical Representation of Mandarin Tones by Non-tonal Second-Language Learners. *J. Acoust. Soc. America* 148 (1), EL46–EL50. doi:10.1121/10.0001586

Hao, Y.-C. (2018). Contextual Effect in Second Language Perception and Production of Mandarin Tones. *Speech Commun.* 97, 32–42. doi:10.1016/j.specom.2017.12.015

Hao, Y.-C. (2012). Second Language Acquisition of Mandarin Chinese Tones by Tonal and Non-tonal Language Speakers. *J. Phonetics* 40 (2), 269–279.

Ho, A. T. (1976). The Acoustic Variation of Mandarin Tones. *Phonetica* 33, 353–367. doi:10.1159/000259792

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biom. J.* 50 (3), 346–363. doi:10.1002/bimj.200810425

Howie, J. M. (1974). On the Domain of Tone in Mandarin. *Phonetica* 30 (3), 129–148. doi:10.1159/000259484

Hu, J., Gao, S., Ma, W., and Yao, D. (2012). Dissociation of Tone and Vowel Processing in Mandarin Idioms. *Psychophysiol.* 49 (9), 1179–1190. doi:10.1111/j.1469-8986.2012.01406.x

Huang, J., and Holt, L. L. (2009). General Perceptual Contributions to Lexical Tone Normalization. *J. Acoust. Soc. America* 125 (6), 3983–3994. doi:10.1121/1.3125342

Huang, T., and Johnson, K. (2011). Language Specificity in Speech Perception: Perception of Mandarin Tones by Native and Nonnative Listeners. *Phonetica* 67 (4), 243–267. doi:10.1159/000327392

Jurafsky, D., and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Upper Saddle River, NJ: Prentice-Hall.

Kaan, E., and Swaab, T. Y. (2003). Repair, Revision, and Complexity in Syntactic Analysis: An Electrophysiological Differentiation. *J. Cogn. Neurosci.* 15 (1), 98–110. doi:10.1162/089892903321107855

Kujala, A., Alho, K., Service, E., Ilmoniemi, R. J., and Connolly, J. F. (2004). Activation in the Anterior Left Auditory Cortex Associated with Phonological Analysis of Speech Input: Localization of the Phonological Mismatch Negativity Response with MEG. *Cogn. Brain Res.* 21 (1), 106–113. doi:10.1016/j.cogbrainres.2004.05.011

Leckey, M., and Federmeier, K. D. (2019). The P3b and P600(s): Positive Contributions to Language Comprehension. *Psychophysiology* 57, e13351. doi:10.1111/psyp.13351

Lee, C.-Y., Tao, L., and Bond, Z. S. (2010). Identification of Multi-Speaker Mandarin Tones in Noise by Native and Non-native Listeners. *Speech Commun.* 52 (11–12), 900–910. doi:10.1016/j.specom.2010.01.004

Lenth, R. (2020). Emmeans: Estimated Marginal Means, Aka Least-Squares Means. (R package version 1.5.3.) [Computer software]. Available at: https://CRAN.R-project.org/package=emmeans.

Ling, W., and Grüter, T. (2020). From Sounds to Words: The Relation between Phonological and Lexical Processing of Tone in L2 Mandarin. *Second Lang. Res.*, 026765832094154, doi:10.1177/0267658320941546

Liu, Y., Shu, H., and Wei, J. (2006). Spoken Word Recognition in Context: Evidence from Chinese ERP Analyses. *Brain Lang.* 96 (1), 37–48. doi:10.1016/j.bandl.2005.08.007

Llompart, M., and Reinisch, E. (2020). The Phonological Form of Lexical Items Modulates the Encoding of Challenging Second-Language Sound Contrasts. *J. Exp. Psychol. Learn. Mem. Cogn.* 46 (8), 1590–1610. doi:10.1037/xlm0000832

Luck, S. J., and Gaspelin, N. (2017). How to Get Statistically Significant Effects in Any ERP experiment (And Why You Shouldn't). *Psychophysiol.* 54 (1), 146–157. doi:10.1111/psyp.12639

MacWhinney, B., and Bates, E. (1989). *The Cross-Linguistic Study of Sentence Processing* (New York, NY: Cambridge University Press).

Malins, J. G., and Joanisse, M. F. (2012). Setting the Tone: An ERP Investigation of the Influences of Phonological Similarity on Spoken Word Recognition in Mandarin Chinese. *Neuropsychologia* 50 (8), 2032–2043. doi:10.1016/j.neuropsychologia.2012.05.002

Matusevych, Y., Kamper, H., Schatz, T., Feldman, N. H., and Goldwater, S. (2021). A Phonetic Model of Non-native Spoken Word Processing. ArXiv:2101.11332 [Cs]. Available at: http://arxiv.org/abs/2101.11332 (Accessed March 12, 2021).

Moreno-Martínez, F. J., and Montoro, P. R. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PLoS ONE* 7 (5), e37527. doi:10.1371/journal.pone.0037527

Newman, R. L., and Connolly, J. F. (2009). Electrophysiological Markers of Pre-lexical Speech Processing: Evidence for Bottom-Up and Top-Down Effects on Spoken Word Processing. *Biol. Psychol.* 80 (1), 114–121. doi:10.1016/j.biopsycho.2008.04.008

Nixon, J. S. (2020). Of Mice and Men: Speech Sound Acquisition as Discriminative Learning from Prediction Error, Not Just Statistical Tracking. *Cognition* 197, 104081. doi:10.1016/j.cognition.2019.104081

Osterhout, L., and Holcomb, P. J. (1992). Event-related Brain Potentials Elicited by Syntactic Anomaly. *J. Mem. Lang.* 31, 785–806. doi:10.1016/0749-596x(92)90039-z

Pelzl, E., Lau, E. F., Guo, T., and DeKeyser, R. (2019). Advanced Second Language Learners' Perception of Lexical Tone Contrasts. *Stud. Second Lang. Acquis* 41 (1), 59–86. doi:10.1017/S0272263117000444

Pelzl, E., Lau, E. F., Guo, T., and DeKeyser, R. (2020). Even in the Best-Case Scenario L2 Learners Have Persistent Difficulty Perceiving and Utilizing Tones in Mandarin. *Stud. Second Lang. Acquis*, 1–29. doi:10.1017/S027226312000039X

Pelzl, E., Lau, E. F., Jackson, S. R., Guo, T., and Gor, K. (2021). Behavioral and Neural Responses to Tone Errors in Foreign-Accented Mandarin. *Lang. Learn.* 71, 414–452. doi:10.1111/lang.12438

Pelzl, E. (2018). *Second Language Lexical Representation and Processing of Mandarin Chinese Tones*. College Park: University of Maryland.

Pelzl, E. (2019). What Makes Second Language Perception of Mandarin Tones Hard? *Csl* 54 (1), 51–78. doi:10.1075/csl.18009.pel

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org/.

Romero-Rivas, C., Martin, C. D., and Costa, A. (2015). Processing Changes when Listening to Foreign-Accented Speech. *Front. Hum. Neurosci.* 9, 1–15. doi:10.3389/fnhum.2015.00167

Sassenhagen, J., and Bornkessel-Schlesewsky, I. (2015). The P600 as a Correlate of Ventral Attention Network Reorientation. *Cortex* 66, A3–A20. doi:10.1016/j.cortex.2014.12.019

Sassenhagen, J., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2014). The P600-As-P3 Hypothesis Revisited: Single-Trial Analyses Reveal that the Late EEG Positivity Following Linguistically Deviant Material Is Reaction Time Aligned. *Brain Lang.* 137, 29–39. doi:10.1016/j.bandl.2014.07.010

Schad, D. J., Vasishth, S., Hohenstein, S., and Kliegl, R. (2020). How to Capitalize on A Priori Contrasts in Linear (Mixed) Models: A Tutorial. *J. Mem. Lang.* 110, 104038. doi:10.1016/j.jml.2019.104038

Schaefer, V., and Darcy, I. (2014). Lexical Function of Pitch in the First Language Shapes Cross-Linguistic Perception of Thai Tones. *Lab. Phonology* 5 (4), 489–522. doi:10.1515/lp-2014-0016

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., and Chen, H.-C. (2005). Brain Responses to Segmentally and Tonally Induced Semantic Violations in Cantonese. *J. Cogn. Neurosci.* 17 (1), 1–12. doi:10.1162/0898929052880057

Schmidt-Kassow, M., and Kotz, S. A. (2009). Event-related Brain Potentials Suggest a Late Interaction of Meter and Syntax in the P600. *J. Cogn. Neurosci.* 21 (9), 1693–1708. doi:10.1162/jocn.2008.21153

Shen, G., and Froud, K. (2016). Categorical Perception of Lexical Tones by English Learners of Mandarin Chinese. *J. Acoust. Soc. America* 140 (6), 4396–4403. doi:10.1121/1.4971765

Shen, G., and Froud, K. (2019). Electrophysiological Correlates of Categorical Perception of Lexical Tones by English Learners of Mandarin Chinese: An ERP Study. *Bilingualism* 22 (2), 253–265. doi:10.1017/S136672891800038X

Singmann, H., Bolker, B., Westfall, J., and Aust, F. (2017). Afex: Analysis of Factorial Experiments. R package version 0.17-8. Available at: http://cran.r-project.org/package=afex [Computer software].

So, C. K., and Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Lang. Speech* 53 (2), 273–293. doi:10.1177/0023830909357156

So, C. K., and Best, C. T. (2014). Phonetic Influences on English and French Listeners' Assimilation of Mandarin Tones to Native Prosodic Categories. *Stud. Second Lang. Acquis* 36 (02), 195–221. doi:10.1017/S0272263114000047

Strange, W. (2011). Automatic Selective Perception (ASP) of First and Second Language Speech: A Working Model. *J. Phonetics* 39 (4), 456–466. doi:10.1016/j.wocn.2010.09.001

Sun, S. H. (1998). *The Development of a Lexical Tone Phonology in American Adult Learners of Standard Mandarin Chinese*. Honolulu, HI: Second Language Teaching & Curriculum Center.

Suzuki, Y., and DeKeyser, R. (2017). The Interface of Explicit and Implicit Knowledge in a Second Language: Insights From Individual Differences in Cognitive Aptitudes. *Lang. Learn.* 67 (4), 747–790. doi:10.1111/lang.12241

Tsukada, K., and Han, J.-I. (2019). The Perception of Mandarin Lexical Tones by Native Korean Speakers Differing in Their Experience with Mandarin. *Second Lang. Res.* 35 (3), 305–318. doi:10.1177/0267658318775155

Van Den Brink, D., Brown, C. M., and Hagoort, P. (2001). Electrophysiological Evidence for Early Contextual Influences during Spoken-word Recognition: N200 versus N400 Effects. *J. Cogn. Neurosci.* 13 (7), 967–985. doi:10.1162/089892901753165872

VanPatten, B. (1996). *Input Processing and Grammar Instruction in Second Language Acquisition*. New York, NY: Ablex Publishing.

Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American Listeners to Perceive Mandarin Tones. *J. Acoust. Soc. America* 106, 3649–3658. doi:10.1121/1.428217

Wiener, S., Ito, K., and Speer, S. R. (2018). Early L2 Spoken Word Recognition Combines Input-Based and Knowledge-Based Processing. *Lang. Speech* 61 (4), 632–656. doi:10.1177/0023830918761762

Wiener, S., Ito, K., and Speer, S. R. (2020). Effects of Multitalker Input and Instructional Method on the Dimension-Based Statistical Learning of Syllable-Tone Combinations. *Stud. Second Lang. Acquis* 43, 155–180. doi:10.1017/S0272263120000418

Wiener, S., Lee, C. Y., and Tao, L. (2019). Statistical Regularities Affect the Perception of Second Language Speech: Evidence From Adult Classroom Learners of Mandarin Chinese. *Lang. Learn.* 69 (3), 527–558. doi:10.1111/lang.12342

Wong, P. C. M., and Perrachione, T. K. (2007). Learning Pitch Patterns in Lexical Identification by Native English-speaking Adults. *Appl. Psycholinguistics* 28 (04), 565–585. doi:10.1017/s0142716407070312

Xu, Y. (1997). Contextual Tonal Variations in Mandarin. *J. Phonetics* 25, 61–83. doi:10.1006/jpho.1996.0034

Xu, Y. (1994). Production and Perception of Coarticulated Tones. *J. Acoust. Soc. America* 95 (4), 2240–2253. doi:10.1121/1.408684

Zhao, J., Guo, J., Zhou, F., and Shu, H. (2011). Time Course of Chinese Monosyllabic Spoken Word Recognition: Evidence from ERP Analyses. *Neuropsychologia* 49 (7), 1761–1770. doi:10.1016/j.neuropsychologia.2011.02.054

Zou, T., Chen, Y., and Caspers, J. (2017). The Developmental Trajectories of Attention Distribution and Segment-Tone Integration in Dutch Learners of Mandarin Tones. *Bilingualism* 20 (5), 1017–1029. doi:10.1017/S1366728916000791

Zou, Y., Lui, M., and Tsang, Y.-K. (2020). The Roles of Lexical Tone and Rime during Mandarin Sentence Comprehension: An Event-Related Potential Study. *Neuropsychologia* 147, 107578. doi:10.1016/j.neuropsychologia.2020.107578

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 06 July 2021
doi: 10.3389/fpsyg.2021.688160

Contrasting Similar Words Facilitates Second Language Vocabulary Learning in Children by Sharpening Lexical Representations

Peta Baxter[1]*, Mienke Droop[2], Marianne van den Hurk[2], Harold Bekkering[1], Ton Dijkstra[1] and Frank Leoné[1]

[1] Donders Centre for Cognition, Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands, [2] Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

OPEN ACCESS

Edited by:
Denisa Bordag,
Leipzig University, Germany

Reviewed by:
Arturo Hernandez,
University of Houston, United States
Maria M. Arredondo,
University of Texas at Austin,
United States

***Correspondence:**
Peta Baxter
peta.baxter@donders.ru.nl

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: 30 March 2021
Accepted: 08 June 2021
Published: 06 July 2021

Citation:
Baxter P, Droop M, van den Hurk M, Bekkering H, Dijkstra T and Leoné F (2021) Contrasting Similar Words Facilitates Second Language Vocabulary Learning in Children by Sharpening Lexical Representations. Front. Psychol. 12:688160. doi: 10.3389/fpsyg.2021.688160

This study considers one of the cognitive mechanisms underlying the development of second language (L2) vocabulary in children: The differentiation and sharpening of lexical representations. We propose that sharpening is triggered by an implicit comparison of similar representations, a process we call contrasting. We investigate whether integrating contrasting in a learning method in which children contrast orthographically and semantically similar L2 words facilitates learning of those words by sharpening their new lexical representations. In our study, 48 Dutch-speaking children learned unfamiliar orthographically and semantically similar English words in a multiple-choice learning task. One half of the group learned the similar words by contrasting them, while the other half did not contrast them. Their word knowledge was measured immediately after learning as well as 1 week later. Contrasting was found to facilitate learning by leading to more precise lexical representations. However, only highly skilled readers benefitted from contrasting. Our findings offer novel insights into the development of L2 lexical representations from fuzzy to more precise, and have potential implications for education.

Keywords: second language learning, vocabulary, lexical representations, representational specificity, language instruction, contrasting

INTRODUCTION

Toward the end of primary school education, most children will have developed a vocabulary of considerable size in their first language (L1). They will be able to read and pronounce a large number of words and know their meanings. At this time, many will also start learning a second language (L2), for which they must acquire new orthographic and phonological word forms and map them onto mostly familiar meanings. This learning process necessarily entails the differentiation and refinement of those foreign lexical representations from less to more precise, not only with respect to phonology, but also orthography and semantics. We propose that a process that triggers the sharpening of representations is implicitly comparing representations to similar ones. We call this representational refinement *contrasting*. In this study, we investigate the process of contrasting and demonstrate that it can effectively be exploited to facilitate L2 word learning by explicitly integrating it in a multiple-choice teaching method.

We will set the stage for our study by first considering to what extent the differentiation process of new words applies to the dimensions of phonology, orthography, and semantics. This will be followed by a discussion on how a foreign-language teaching method that integrates contrasting of similar foreign words may benefit L2 word learning in children. We will then present experimental evidence to show that it can indeed be beneficial under specific circumstances.

Our starting point is the *fuzzy lexicon hypothesis*, according to which the phonolexical and/or phonological representations of newly learned L2 words are initially underspecified, or *fuzzy* (Cook, 2012; Cook and Gor, 2015; Cook et al., 2016; Lancaster and Gor, 2016). This fuzziness leads to inaccuracies in auditory speech perception. One source of errors is that learners may not accurately perceive phonemes that do not exist in their native language. For example, a Dutch-speaking learner of English may in this way confuse *thin* ['θɪn] with *fin* ['fɪn], because the phoneme /θ/ does not exist in Dutch. Another source of errors is that L2 learners often rely on phonological information to decide on the meaning of a word. In the absence of semantic or orthographic knowledge about the word *spider*, they might for example use phonological similarity with other known words to erroneously conclude that the word is semantically related to *spy* or *spiral*. As proficiency increases, and more similar words are encountered, representations gradually become more specified.

The fuzzy lexicon hypothesis has predominantly been investigated for L2 phonological processing in adults. Given the overall prominent role of phonology in the development of linguistic skills (see van Goch, 2016; Janssen, 2017), and the fact that phonological representations are crucial to the development of stable representations in memory (Baddeley, 1992; Baddeley et al., 1998), it is unsurprising that it has been the natural starting point for much of the research on the nature of novel L2 representations. However, the notion of a gradual specification process from "holistic" to detailed knowledge and representations also pervades native language acquisition research in children and beyond the phonological dimension. Several theories here posit a gradual sharpening process across linguistic dimensions, such as the lexical tuning hypothesis (Castles et al., 2007), the lexical restructuring hypothesis (Metsala, 1997; Metsala and Walley, 1998), the psycholinguistic grain size theory (Ziegler and Goswami, 2005), and the lexical quality hypothesis (Perfetti and Hart, 2002; Perfetti, 2007).

Many of these theories are concerned with the process of learning to read (e.g., the lexical quality hypothesis), which differs from acquiring phonology or semantics because reading has to be explicitly taught. Nevertheless, similar gradual processes are at play in that learners may initially rely on a fuzzy perceptual representation to recognize a printed word such as *salt*, but their orthographic representations must become more precise when they encounter the word *slat*. By comparing *salt* and *slat*, learners can obtain more precise letter position information necessary for accurate visual word recognition (Grainger and Van Heuven, 2004). Similar considerations apply not only to reading in the L1, but also in the L2. Since reading, including the processing of orthographic input, is a guided process, it is a particularly interesting dimension for considering the effect of specific instruction methods, as in the present study.

This gradual specification process can also be seen during the development of L1 lexical-semantic representations. As for L1 phonological representations, it is an implicit process in native language acquisition. During early language development, cross-situational co-occurrences of semantically related words enable children to associate a word form with the correct concept (Smith and Yu, 2008; Suanda et al., 2014). When they first hear the word *dog* while seeing a dog and a cat, they might generalize the word *dog* to either of these concepts. As the word *dog* is encountered in more contexts, the association between *dog* and the correct animal is updated, becoming more detailed and specific (Clark, 2004).

The development of both orthographic and semantic representations is undoubtedly linked to phonology. Orthographic processing is interrelated with phonological (but also semantic) processing, as models like the Dual Route Cascaded model (Coltheart et al., 2001) attest. Learning how to read involves mapping known sounds onto graphemes, during which phonological information is automatically activated (Frost, 1998). Learning the meaning of words requires children to distinguish speech sounds. In addition, phonology is sometimes required to disambiguate word meaning, for instance to determine whether "read" is in the present or past tense (pronounced ['ri:d] or ['rɛd], respectively).

Nevertheless, there is also evidence of cases where other lexical dimensions can overrule phonology. For example, children have been shown to be able to use sublexical orthography to infer the meaning of words without phonology mediating the process (Nation and Cocksey, 2009). Similarly, when encountering new words in the L2, learners may not always have complete or even accurate information about the associated L2 phonology, for instance, when they are learning from word lists or are reading. Even when they do, special items such as homophones require consultation of orthography or semantics to learn the word correctly. For example, because *bawl* ['bɔl] and *ball*['bɔl] are pronounced the same; their difference in meaning is signaled by their spelling, necessitating precise orthographic knowledge. Even when L2 learners are acquiring non-special words, they need to acquire specific knowledge of the meaning, and build precise links between form and meaning in order to not confuse words with related concepts. Therefore, drawing on the lexical quality hypothesis (Perfetti and Hart, 2002; Perfetti, 2007), L2 word competency depends on the specificity of not only phonological representations, but also orthographic and semantic representations. A poor sharpening of semantic and orthographic representations may thus also be a contributing factor to why even advanced L2 learners still confuse similar words (Llach, 2015).

In sum, a development from fuzzy to more specified lexical representations during L2 learning is crucial not only when developing novel phonological representations, but for orthographic and semantic representations as well. We therefore propose and test an extension of the fuzzy lexicon hypothesis to L2 orthographic and semantic dimensions. In addition, we propose an extension of the lexical quality hypothesis to a foreign language. Indeed, as indicated above, successful word retrieval depends on specific and tightly bound representations with respect to all three dimensions of orthography, phonology, and semantics (Perfetti and Hart, 2002; Perfetti, 2007). Before we test this extended theoretical view in an experiment, we consider the processes involved in lexical specification in more detail.

The fuzzy lexicon hypothesis implies that one of the forces that drives representations to become more specific during L2 word learning is *similarity*. Similarity already plays a large role in early native language development: As children encounter more similar words, they detect the statistical regularities in how these words sound, clustering them in similar-sounding "competitor sets" (Ziegler and Goswami, 2005). Doing so enables efficient retrieval, as representations must become increasingly detailed to distinguish them from neighboring similar words. This notion of competitor set is also important in the visual domain for L1 and L2 word reading, where sets of similar words make up neighborhoods of the presented target (van Heuven et al., 1998). This notion underlines the role of similarity as a driving force in sharpening representations. As long as no similar words are encountered, a fuzzy representation is sufficient. Only when similar words enter the lexicon, the representations are driven to become more detailed.

In line with this, we propose that an important process involved in lexical specification is what we describe as contrasting: In order to trigger the sharpening process, learners must (implicitly or, as in our study, explicitly) carefully compare similar representations. Drawing on studies of perceptual learning, contrasting likely involves mechanisms of selective attention that guide attention toward the relevant (i.e., distinctive) information (Goldstone, 1998; Francis et al., 2000; Francis and Nusbaum, 2002), thereby decreasing the perceived level of similarity between items (Adini et al., 2002) and resulting in more specific representations. We propose that contrasting is a learning process essential for theories such as the fuzzy lexicon hypothesis and lexical quality hypothesis, and posit that contrasting and gradual specification processes also take part in L2 learning by children, even though they differ from adults in that they are still developing their native representations.

Assuming that contrasting is indeed a process involved during learning, it becomes important from both a theoretical and practical perspective to investigate whether it can be integrated in a learning method, and thereby influence the efficiency with which representations evolve from fuzzy to more specific.

At first sight, using similar L2 words for testing our theoretical position may seem to be counterintuitive. The reason is that researchers have often observed that the presence of orthographic and semantic similarity between L2 words is problematic for learning. Several studies have shown that semantic similarity between words negatively impacts learning (e.g., Tinkham, 1993; Waring, 1997; Papathanasiou, 2009; Ishii, 2015). Orthographic similarity, though less widely researched, can also negatively impact learning (Laufer, 1988; Llach, 2015). As a consequence, some have warned against teaching words in lexical sets (Nation, 2000). However, a commonality of several of these studies is that they manipulate similarity in terms of lists of words, in which the words are similar or dissimilar. A disadvantage of doing so is that one cannot easily disentangle to what extent differences found are due to list effects. Moreover, learners will also need to learn the similar words eventually, so rather than circumventing the similarity issue by presenting dissimilar lists, as sometimes proposed (Nation, 2000), ideally instruction tools should be able to cope with it.

However, none of these studies take into account the characteristics of the *learning method*. Potentially, using a different method may lead to less confusion between similar words, or even facilitation. For instance, learning similar words in a traditional way such as from word lists typically does not encourage learners to focus their attention on challenging lexical elements. Possibly, if these studies had used methods that encourage learners to carefully compare and contrast similar words, focusing on differences between them, similarity might have turned out to not be a hindrance, perhaps even helpful.

In fact, there is some evidence showing that such tasks may help specify representations. In the phonological dimension, lexical specificity training (Logan et al., 1991; Bradlow et al., 1999) is an example of such a contrasting task. In this task, learners repeatedly contrast difficult or undifferentiated phonological contrasts, which sharpens the boundaries between these contrasts. This method has, for example, proven to be helpful for Japanese learners of English, who conflate the phonemes "r" and "l" into a single phonemic category. Using a similar learning task and tracking learners' eye movements, Llompart and Reinisch (2020) determined that, as posited previously, the benefits of contrasting words containing similar phonemes is due to attention being guided toward the relevant, i.e., distinctive, lexical information. This process enables them to encode this information more successfully. These findings match those in the field of perceptual learning, where studies have shown that contrasting similar visual stimuli subsequently made them easier to differentiate (Adini et al., 2002). Recent studies have also shown that lexical specificity training also facilitates L2 vocabulary learning, both in children and (young) adults (van Goch et al., 2014; Janssen et al., 2015; van de Ven et al., 2018). The training leads to more specific, higher quality phonological representations that make new words easier for learners to retrieve.

Though we are not aware of any studies that consider the effects of contrasting the orthographic or semantic dimension in L2 vocabulary learning, the notion of contrasting as a learning method does appear to be generalizable to other learning domains. For instance, in general learning, researchers looked at the effects of learning with multiple-choice questions that were manipulated in such a way that the distractor answers were all plausible, i.e., similar on a certain level (Little et al., 2012; Little and Bjork, 2015). Compared to conditions where the answers were not all equally plausible, this manipulation led to better learning. These findings suggest that contrasting is not only a cognitive mechanism essential to learning, but can be used as an instrument for learning in various fields.

Contrasting methods may be generally beneficial to learning, but it is also likely that their effects are modulated by certain factors, especially in children. In this study, we consider some of the linguistic factors that may impact learning with contrasting.

In general, aspects such as prior L2 knowledge (Elgort et al., 2015) and verbal working memory (Kormos and Sáfár, 2008; Linck et al., 2014) are likely to play a role in how well-children are able to learn new words. Because children are beginning readers with still developing L1 representations, other, more specific, characteristics may also come into play during learning by contrasting.

First, there is vocabulary size, which is a robust predictor of language ability (Lee, 2011) and is intrinsically linked to the quality of representations. Verhoeven and Perfetti (2011) note that vocabulary growth can be seen as the combination of *quantity* and *quality* of word representations. A larger vocabulary implies that many of these representations will be more specific, since more similar words will be known. Therefore, given that children have a smaller vocabulary than adults, they might be particularly good candidates to benefit from contrasting. However, we might observe less of an advantage in children with a relatively large vocabulary.

A second child characteristic is reading skill level. Better readers are both more proficient at decoding word orthography and extracting word meaning (Gough and Tunmer, 1986; Verhoeven and Perfetti, 2011). According to the lexical quality hypothesis, the ability to recognize a word depends on the specificity of the form representation, and the ability to extract the meaning depends on the specificity of the meaning representation and link to form (Perfetti and Hart, 2002; Perfetti, 2007). Consequently, less skilled readers might experience too much confusion from a contrasting task to display a learning advantage. In comparison, skilled readers might be particularly good at detecting the differences between the similar words, thereby benefiting from contrasting most.

The primary goal of our study is to determine whether a learning task in which children contrast orthographically and semantically similar L2 words triggers the specification process, consequently facilitating learning. We conducted a visual multiple-choice L2 word learning experiment in which Dutch children learned orthographically and semantically similar English words by either contrasting them or not. The relationship between our task and the specification process is illustrated in **Figure 1**. Multiple-choice has been shown to be an effective L2 vocabulary learning method (Nakata and Webb, 2016), as it benefits learners by requiring them to practice retrieving the correct answer (see Roediger and Butler, 2011). In the learning task at hand, children saw a Dutch word with three possible English translations, selected an answer, and learned from the feedback. When the answer options were similar, contrasting occurs, as closer comparisons must be made to select the correct options. Word knowledge was then tested with a L2 to L1 translation task at two different points in time.

We hypothesized that contrasting would facilitate L2 word learning by directing learners' attention to relevant lexical information, allowing them to encode this information more precisely, thereby forging more specific representations. In addition, we considered the possibility that the aforementioned linguistic characteristics would moderate the effect of contrasting. In particular, children with a larger vocabulary might benefit from contrasting to a lesser extent than those with a smaller vocabulary; and children with better reading skills might experience a larger contrasting advantage than less skilled readers.

In sum, from a theoretical perspective, determining whether a contrasting learning method is beneficial will further our understanding of how L2 learning results in refined representations for different lexical dimensions and in different participant populations. From a practical perspective, our results may have implications for L2 vocabulary instructions using contrasting as a teaching device.

METHOD

Participants

Fifty-one children in five primary schools in the Netherlands (US grade 4, age range 9–10; 31 boys and 20 girls) participated in the study. They had minimal prior formal English instruction (prior to grade 4 they received a maximum of 30 min per week of informal exposure to English, by listening to songs for example; in grade 4 they received 1 h per week of formal lessons). The study was approved by the Ethics Committee Social Sciences (ECSS) of the Radboud University Nijmegen, and informed consent was obtained from the parents. The pretest revealed that two children knew more than 20% of the to-be-learned words and their data were therefore excluded. In addition, the reading skills scores for one child were not available, therefore they were also excluded from further analysis, resulting in the data of 48 children being analyzed in total.

Stimuli

The children all learned the same 27 words, namely nine orthographically similar, nine semantically similar, and nine fully dissimilar English words. The Dutch translations were concrete nouns, selected to match in length and frequency across conditions in English. Semantically similar words fell into a common category (e.g., bicycle parts), their semantic relatedness was checked using Snaut (Mandera et al., 2017), an empirically validated online software that calculates the semantic distance between items. Orthographically similar words were selected to be at least 50% similar using normalized Levenshtein distance (Levenshtein, 1966). The fully dissimilar words had little to no orthographic or semantic overlap. The full list of stimuli and their orthographic and semantic similarity can be seen in **Appendix 1** (**Table A1** for the contrasted condition, **Table A2** for the not contrasted condition). To ensure that the children would be familiar with the meaning of the words, we selected Dutch words that are typically acquired earlier than age 9 using (Brysbaert and Biemiller, 2017) age of acquisition (AoA) database. For some words we were unable to find an AoA, to make sure that the meaning would be known we added a picture to each word. Pictures were either retrieved from the Multipic database (Duñabeitia et al., 2018), or were copyright-free images altered to resemble the style of the Multipic pictures.

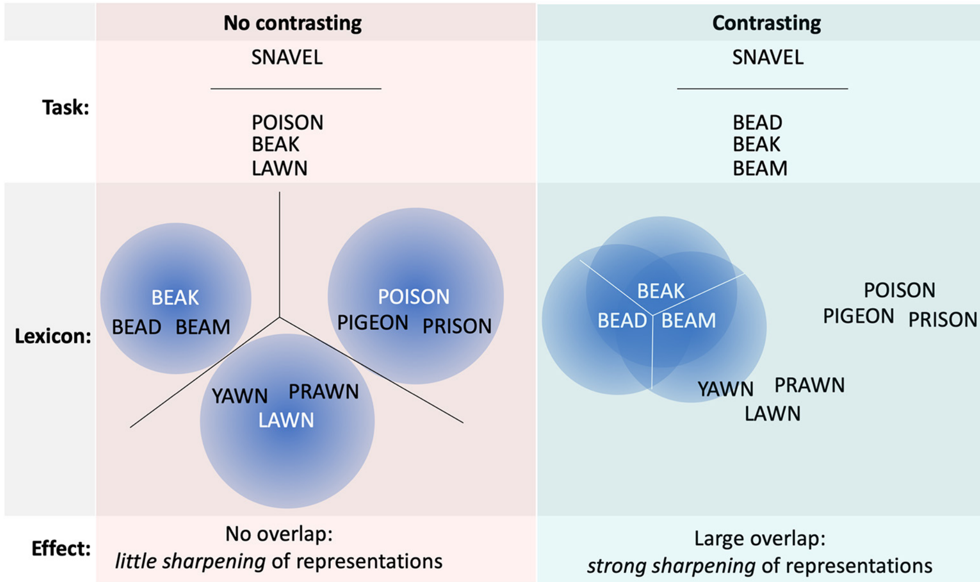


FIGURE 1 | Schematic overview of the expected difference between the no contrasting and contrasting condition. Task shows the used multiple-choice task; Lexicon contains the to-be-learned words, sorted by orthographic similarity in this example; Effect highlights the expected effect on the representations. When not contrasting (left; red), the fuzzy representations (blue gaussian circles) activated for the presented words are not overlapping, hence little sharpening of the representations ensues. When contrasting (right; cyan), the words to be distinguished are largely overlapping, hence the representations need to be sharpened to differentiate the words (white lines).

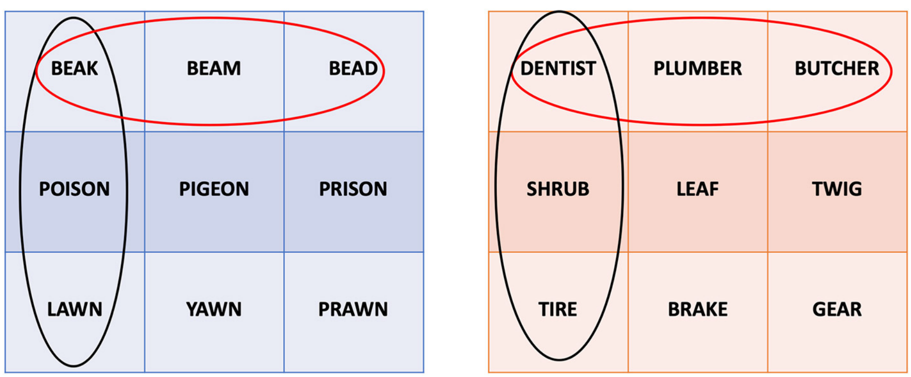


FIGURE 2 | Presentation system allowing for the between-subjects contrasting condition. The blue table on the left contains the orthographically similar words used in the experiment. The orange table on the right contains the semantically similar words used in the experiment. Red circles illustrate example triplets for the contrasted condition, black circles illustrate example triplets for the not contrasted condition. The same principle is applied to all rows and columns.

Design

Word List Structure

In the multiple-choice task, the word list was presented in two distinct ways. One half of the children saw the orthographically and semantically similar words sorted in such a way that they would be contrasted, and the other half not. The fully dissimilar words were presented the same way across conditions, serving as a baseline. This paradigm allowed the exact same words to be learned either contrasted or not, thereby eliminating possible list effects.

Each of the sets of nine words per condition consisted of three groups of three words, which could be combined into three unique triplets of answer options for a single target word. Each triplet was presented once during learning, meaning that each word was seen three times as a target word, and six times as a distractor.

For the similar words, in the contrasting condition, each group contained three orthographically or semantically similar words, resulting in similar triplets (**Figure 2**, rows). In the no contrasting condition, one word from each group was picked to create three

FIGURE 3 | (Left) Example of a trial. At the top, the word in the native language (Dutch), at the bottom three possible English translations. In this example, the words are semantically similar, and in the contrasted condition. (Right) Example of the feedback after children have clicked on an answer. The correct translation pair becomes green, the incorrect answers are red.

new groups, resulting in dissimilar triplets with the same words (**Figure 2**, columns).

For the dissimilar condition, one set of nine fully dissimilar words was selected, resulting in three dissimilar triplets presented once each. For this condition, the children in both the contrasting and the no contrasting condition saw exactly the same triplets during learning.

General

The experiment was carried out at five different primary schools in the Netherlands, in relatively quiet rooms. The experimental design consisted of three sessions containing a pretest, learning phase, and posttest, which was repeated a week later as a retention test. In the first session, we administered the pretest. In the second session, which took place at least 1 day later, children carried out the learning phase and immediate post-test. In the third session, exactly 1 week after the learning session, they completed the posttest again as a retention test.

Procedure

The pretest, learning task, and posttest were all programmed in Expyriment (Krause and Lindemann, 2014), and carried out on a Dell laptop (screen size 14 inches, resolution 1920*1080 pixels).

Pre- and Post-tests

The pre- and post-tests had an identical format. Children saw all of the English words in the list on screen, in random order. The task was to type the Dutch translation. They did not receive any feedback. Spelling mistakes in Dutch were counted as correct. After the experiment was over, after the retention post-test, the experimenters distributed an answer sheet with the correct translations, and shortly debriefed with the children to discuss how the task went and which words had been learned correctly or incorrectly. The pre- and post-tests lasted 5–10 min each on average.

TABLE 1 | Summary of the generalized linear mixed-effects model for the learning data, including estimates, standard errors (SE), *z* values, and significance level.

Fixed Effects				
	Est/Beta	**SE**	***z***	***p***
Intercept	0.43	0.44	0.98	N.s.
Ortho vs. semantics	0.17	0.25	0.67	N.s.
Similar vs. dissimilar	0.01	0.25	0.02	N.s.
Contrasting	−0.36	0.19	−1.88	N.s.
Block	**0.61**	**0.05**	**13.18**	**<0.001**
AVI	**0.14**	**0.03**	**4.01**	**<0.001**
PPVT	0.03	0.09	0.35	N.s.
15-WT	−0.08	0.09	−0.94	N.s.
English knowledge	**−0.37**	**0.15**	**−2.44**	**<0.05**
Contrasting * ortho vs. semantics	0.01	0.18	0.03	N.s.
Contrasting * similar vs. dissimilar	**0.48**	**0.18**	**2.67**	**<0.005**
Random effects				
	Variance	**S.D.**		
School: subject (Intercept)	0.25	0.50		
Word (Intercept)	0.20	0.45		
Model fit				
R^2	**Marginal**	**Conditional**		
	0.10	0.21		

*Model equation: accuracy ∼ similarity*contrasting + AVI + PPVT + 15-WT + English Knowledge + (1|school: subject) + (1|word).*
Bolded values are significant.
N.s., Non-significant; Ortho, Orthography; AVI, Reading skills; PPVT, Dutch vocabulary; 15-WT, Verbal memory.

Learning Phase

The learning phase started with an instruction round in which children were familiarized with the task. The instructions were given on screen through a practice round, and the experimenters also provided oral explanations. Before starting the experiment, the experimenters thoroughly checked whether the task was clear. The children were informed that their classmates would not see the words in the same order as them, and were asked to focus on their own task. In the learning task, children completed three blocks in which each word was presented once. This thus resulted in 81 trials. On each trial, they saw a Dutch word and corresponding picture, with three possible translations aligned vertically below (**Figure 3**, left). They were instructed to carefully read all answer options, and then click on the translation they thought to be correct. There was no time limit to click on an answer.

Once a word had been clicked, visual feedback was presented for 5 s before the next trial appeared. Feedback was given by highlighting the correct translation and target word in green, while the incorrect answers were highlighted in red (**Figure 3**, right). This ensured that the feedback was visually identical regardless of whether a mistake had been made or not.

Between each block they saw how many trials they had answered correctly, and were encouraged on screen to try to improve their score. This was included to increase motivation to perform the task. In total, the learning task took ∼20 min.

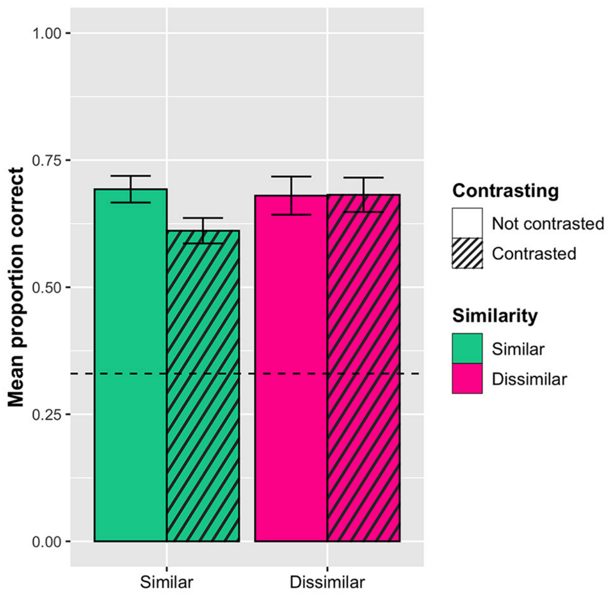


FIGURE 4 | Interaction effect between sorting and similarity (orthographic and semantic similarity collapsed) with 95% confidence interval bars. Dashed line represents chance level (0.33). The similar words are significantly less accurate in the contrasted condition compared to the no contrasting condition.

Additional Measures

In addition to the main experimental components, we also measured covariates that were likely to affect learning or interact with contrasting, namely Dutch vocabulary size, verbal working memory, reading skills, and amount of contact with English.

Dutch Vocabulary Size

We measured the children's Dutch vocabulary size using a computerized version of the Dutch Peabody Picture Vocabulary Test (PPVT) (Dunn et al., 2005). In this task, children hear a Dutch word and must indicate to which of four pictures they see on screen it corresponds. Words are clustered in sets of 12 of increasing difficulty. The start set is selected according to the children's age. These sets are then used to calculate the children's raw scores (number of words heard-number of mistakes), which gives an indication of how many words they know. The PPVT has good reliability (0.94; Dunn et al., 2005). We administered the PPVT during the first session, before the pretest, and it took ∼20 min.

Verbal Working Memory

Verbal working memory was assessed using an adaptation of the Dutch 15-Woordentest (15-WT) (Saan and Deelman, 1986), in which children had to remember a series of 15 auditory words. They heard each series five times, after each time they had to recall all the words they remembered and their score was calculated. We only measured immediate recall to limit the testing time. Their raw scores were the number of words remembered in total. The 15-WT has good reliability (0.80–0.83; Saan and Deelman, 1986). We administered the 15-WT during the first session, after the PPVT, and it took ∼10 min.

Reading Skills

In addition, we also measured the children's reading skills using the Dutch standardized school test "Analyse van Individualiseringsvormen" (AVI) (Krom et al., 2010). This test measures whether children's reading skills by measuring their speed and accuracy while reading short texts of increasing difficulty. Their score consists of the grade their reading skills correspond to. This can be on par with their current grade, or grades above or below their current grade. In our study, this resulted in 9 possible scores (1 = middle of US grade 2; 9 = beyond the end of US grade 5; with a score of 5 corresponding to a level equivalent to their current grade). The reliability of this test is good (0.94–0.97; Krom et al., 2010). This test was administered by the teachers at the end of the previous school year.

Contact With English

Finally, we also measured how much contact children had with written and spoken English outside of school by means of a questionnaire. This questionnaire consisted of 7 questions, which included questions such as "How often do you watch films in English?" The outcome of the contact with English questionnaire was a value between one and four (1 = no contact with English, 4 = a lot of contact with English). The full questionnaire translated to English can be found in **Appendix 2**. We administered this questionnaire at the beginning of the third session, and children took ∼5 min to complete it.

RESULTS

Data Analysis

To determine whether contrasting facilitated learning in the different similarity conditions, as well as the effect of the covariates, we conducted generalized linear mixed-effects regression models using the lme4 package in R (Bates et al., 2014). Following the recommendations of Meteyard and Davies (2020), the models' parameters selection was driven by the research question, and we thus only included relevant interactions, which also helps avoid overfitting issues.

The learning phase and post-tests data were analyzed separately. In the learning model, the fixed factors were block (as a continuous factor), word similarity (orthographically similar / semantically similar / fully dissimilar), contrasting (contrasted / not contrasted), and the interaction between contrasting and similarity. In the post-tests model, the fixed factors were time of testing (immediate/retention), word similarity, contrasting, and the interactions between contrasting and time of testing, and contrasting and word similarity. Accuracy both in the learning phase and the post-tests was measured binarily.

For the similarity variable, we used Helmert contrasts to compare the effect of semantic similarity vs. orthographic similarity, and similar words (orthographically + semantically) vs. dissimilar words.

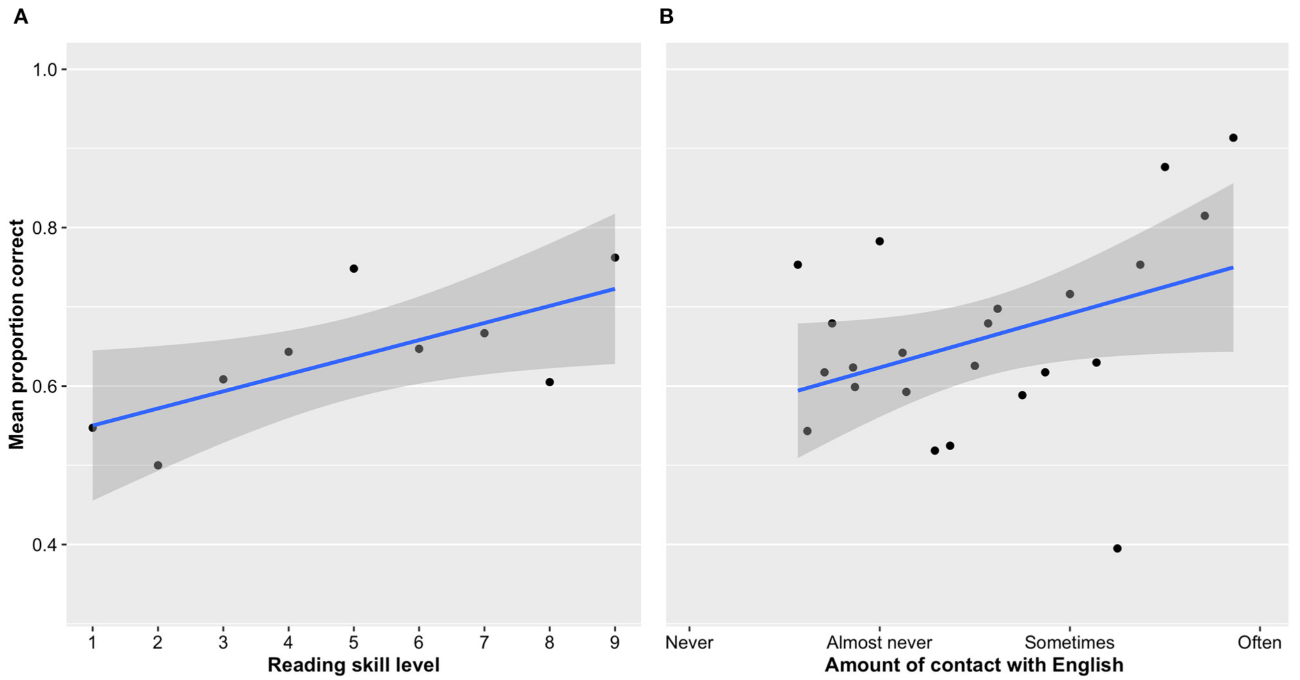


FIGURE 5 | The shaded areas represent the 95% confidence interval. **(A)** Effect of reading skill level on accuracy during learning. Skilled readers learned more words in the multiple-choice task. The dots represent the average proportion of words remembered for all the children within a given level of reading skill **(B)** Effect of knowledge of English on accuracy during learning. Children who had more previous knowledge of English learned more words during the multiple-choice learning task. The dots represent the average proportion of words remembered for all the children within a certain level of English knowledge.

All additional measures (i.e., PPVT scores, AVI scores, 15-WT scores, and contact with English) were added as covariates to the models. The PPVT and 15-WT data were rescaled by z-score normalization and centered around the mean.

In all models, we added random intercepts for words and subjects nested within school. Nesting the subjects within the schools provides an indirect way of controlling for potential differences in socio-economic status, which can impact educational achievement (Strenze, 2007). We did not add any random slopes, as we had no theoretical reason to believe these would affect the results and would result in an unnecessarily complex set of models.

Descriptive Statistics

On average, children knew 3% ($SD = 1.8\%$) of the words, or approximately one word, before doing the learning task. This word was not consistent across children, but the words "prison," "smoke," and "pants" were often known to the children prior to learning. At the end of the last block of the learning phase, they accurately selected 76.9% ($SD = 42.2\%$) of the words on average, or ~21 of the 27 words. Immediately after learning, the children were able to recall over one-third of the words (37.6% [$SD = 48.4\%$], or ~10 words). One week later, this performance decreased to slightly less than one third (24.6% [$SD = 43.1\%$], or ~7 words). As could be expected, there was a large amount of individual variation between the children's performances. The lowest accuracy across posttests was one word ($M_{acc} = 1.9\%$), while the highest was 19 words ($M_{acc} = 72.2\%$).

Learning Phase

The analysis results, in which we considered the effects of word similarity, contrasting, and their interaction, can be seen in **Table 1**.

Main Variables

The analysis showed a significant main effect of block, indicating that children learned from the multiple-choice task during the learning phase ($M_{prop\_corr} = 0.54$, $SD = 0.50$; $M_{prop\_corr} = 0.67$, $SD = 0.47$; $M_{prop\_corr} = 0.77$, $SD = 0.42$ for blocks 1, 2, and 3 respectively). There were no other significant main effects.

In addition, the analysis revealed a significant interaction effect between contrasting and similarity, with the similar words (orthographically + semantically) in the contrasted condition being less accurate ($M_{prop\_corr} = 0.61$, $SD = 0.48$) than the similar words in the not contrasted condition ($M_{prop\_corr} = 0.69$, $SD = 0.46$). In contrast, there was no difference for the dissimilar words ($M_{prop\_corr} = 0.68$, $SD = 0.47$ in the contrasted condition; $M_{prop\_corr} = 0.68$, $SD = 0.47$ in the not contrasted condition). In other words, seeing the similar words presented together as distractors in the multiple-choice task made learning them more difficult. This effect can be seen in **Figure 4**.

Additional Measures

The analysis revealed a strong effect of reading skills as measured by the AVI, with the children with higher reading skills performing better on the learning task than those with lower reading skills (**Figure 5A**). In addition, children with better

TABLE 2 | Summary of the generalized linear mixed-effects model for the post-tests data, including estimates, standard errors (SE), *z* values, and significance level.

Fixed Effects				
	Est/Beta	**SE**	***z***	***p***
Intercept	−1.08	0.71	−1.52	N.s.
Ortho vs. semantics	0.14	0.26	0.53	N.s.
Similar vs. dissimilar	0.13	0.15	0.89	N.s.
Contrasting	0.22	0.29	0.75	N.s.
Time_testing	**−1.00**	**0.15**	**−6.61**	**<0.001**
AVI	**0.23**	**0.06**	**4.06**	**<0.001**
PPVT	−0.04	0.14	−0.30	N.s.
15-WT	−0.04	0.15	−0.30	N.s.
English knowledge	−0.40	0.25	−1.58	N.s.
Contrasting * Ortho vs. semantics	0.03	0.12	0.24	N.s.
Contrasting * Similar vs. dissimilar	**−0.21**	**0.07**	**−2.97**	**<0.005**
Contrasting*Time_testing	0.28	0.20	1.40	N.s.
Random effects				
	Variance	**S.D.**		
School: subject (Intercept)	0.77	0.88		
Word (Intercept)	1.04	1.02		
Model fit				
R^2	**Marginal**	**Conditional**		
	0.11	0.42		

*Model equation: accuracy ∼ similarity*contrasting + block + AVI + PPVT + 15-WT + English Knowledge + (1|school: subject) + (1|word).*
Bolded values are significant. N.s., Non-significant; Ortho, Orthography; AVI, Reading skills; PPVT, Dutch vocabulary; 15-WT, Verbal memory.

knowledge of English also learned better (**Figure 5B**). The other covariates had no effect on learning.

Post-tests

The analysis results, which consider the effect of word similarity and contrasting during both the posttest and the retention test can be seen in **Table 2**.

Main Variables

Unsurprisingly, analysis of the post-test data revealed a main effect of time, with children knowing more words immediately after learning ($M_{prop\_corr} = 0.38$, $SD = 0.48$) than 1 week later ($M_{prop\_corr} = 0.25$, $SD = 0.43$). There were no other significant main effects.

In addition, the analysis did reveal a significant interaction between word similarity and contrasting: The similar words were remembered better when they had been contrasted ($M_{prop\_corr} = 0.33$, $SD = 0.47$) than when they had not ($M_{prop\_corr} = 0.27$, $SD = 0.44$), whereas there was no difference for the dissimilar words ($M_{prop\_corr} = 0.31$, $SD = 0.46$ in the contrasted condition; $M_{prop\_corr} = 0.34$, $SD = 0.48$ in the not contrasted condition). This effect can be seen in **Figure 6**.

Additional Measures

The analysis showed a significant effect of reading skills as measured by the AVI, with children with higher reading skills performing better on the post-tests than those with lower reading skills (**Figure 7**). The other measures did not explain any of the variance.

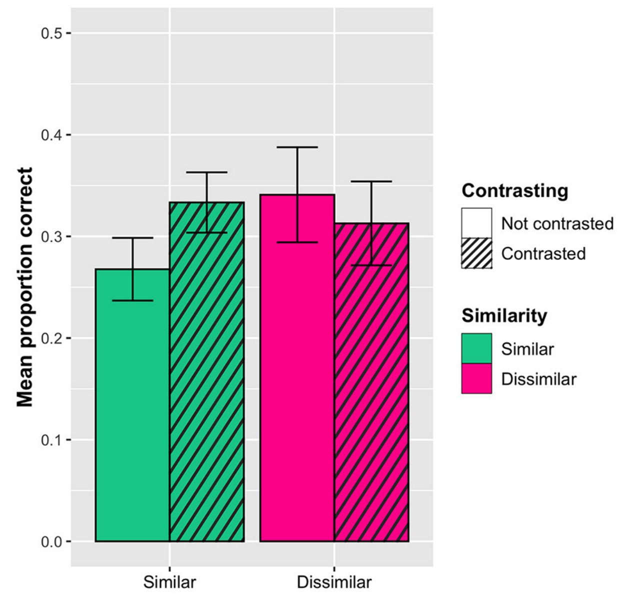


FIGURE 6 | Interaction effect on the post-tests between sorting and similarity (orthographic and semantic similarity collapsed) with 95% confidence interval bars. The similar words are remembered significantly better when they have been contrasted than when they have not.

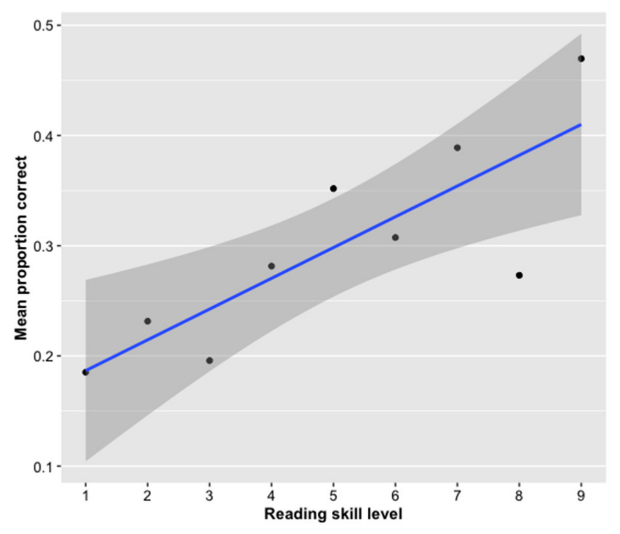


FIGURE 7 | Effect of reading skill level on accuracy in the post-tests. Skilled readers remembered more words than less skilled readers. The dots represent the average proportion of words remembered for all the children within a given level of reading skill. The shaded area represents the 95% confidence interval.

Post-hoc Analyses

Moderating Role of Reading Skills

Because reading skills appeared to explain a highly significant part of the variance in both the learning phase and the post-tests, we investigated this further in *post hoc* tests. In particular,

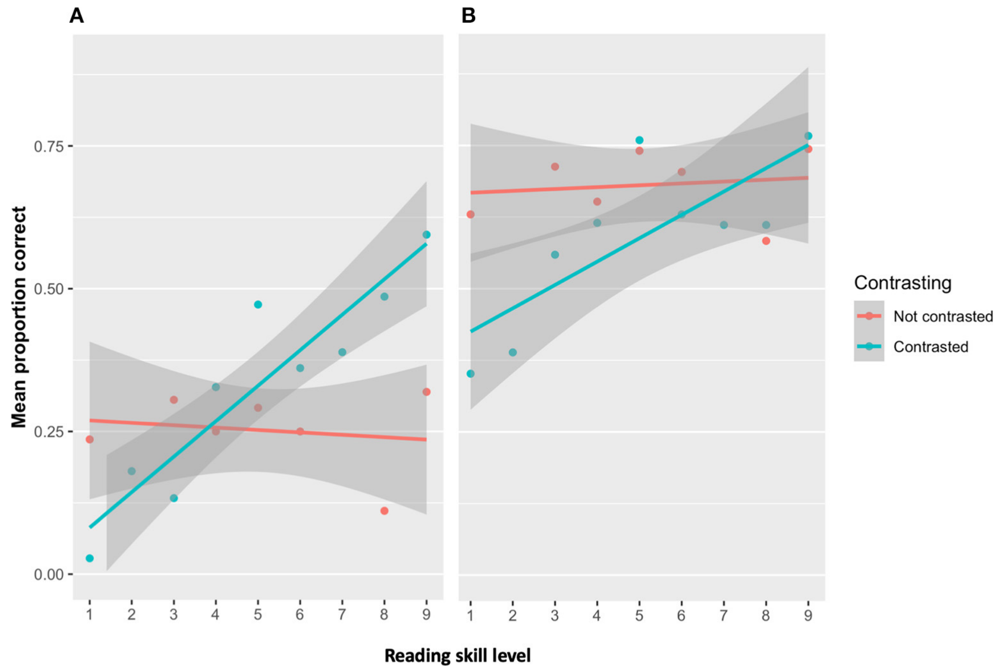


FIGURE 8 | The dots represent the average proportion of words remembered for all the children within a given level of reading skill (red = all children in the not contrasted condition, blue = all children in the contrasted condition). The shaded area represents the 95% confidence interval. **(A)** Effect of contrasting on the similar words during learning depends on reading skill level. Less skilled readers were negatively impacted by contrasting than skilled readers. **(B)** Effect of contrasting on the similar words on the post-tests depends on reading skill level. Contrasting leads to more remembering of the words, but only for skilled readers.

we investigated whether reading skills moderated the effect of contrasting with multiple regression models. We only modeled the orthographically and semantically similar conditions, since in the dissimilar condition no contrasting occurs.

The regression on the learning phase data revealed that reading skills significantly moderated the effects of contrasting [$b = 0.02$, $t_{(3884)} = 3.49$, $p < 0.001$]. Relative to the not contrasted condition, the performance of children in the contrasted condition more strongly depended on reading skills. Specifically, children with lower reading skills were less accurate during learning than those with higher reading skills (**Figure 8A**). This indicates that the children with lower reading skills were negatively affected by contrasting during learning, while those with higher reading skills were not.

In the post-tests, the same analysis revealed that reading skills level also significantly moderated the effects of contrasting [$b = 0.05$, $t_{(2,588)} = 7.32$, $p < 0.001$]. The children in the contrasted condition with higher reading skills remembered more words than those with lower reading skills. In addition, the confidence intervals (see **Figure 8B**) indicate that the children in the contrasting condition that had average to above average reading skills (i.e., scores 5–9) remembered more words than all children in the not contrasted condition.

Error Characteristics

In order to gain more insight into the data, we also analyzed the children's types of errors on the post-tests, comparing the contrasted to the not contrasted condition. Again, we only considered the similar words. Specifically, we considered cases in which children confused a word for a similar one in the list. We conducted a Wilcoxon signed rank test, which revealed that children in the contrasted condition made significantly more misselections of similar words in the list (e.g., answering the Dutch word for *prison* instead of *poison*) than children in the not contrasted condition ($N = 63$ and $N = 36$, respectively, $W = 14{,}430$, $p < 0.05$). This suggests that children's representations for these words were more specific than for those for which they provided a dissimilar answer or no answer.

DISCUSSION

The main goal of this study was to determine whether contrasting orthographically and semantically similar L2 words in a multiple-choice learning task would facilitate children's learning of these words. We proposed that theories such as the fuzzy lexicon hypothesis (Cook et al., 2016) and the lexical quality hypothesis (Perfetti and Hart, 2002; Perfetti, 2007) apply to L2 orthographic and semantic dimensions, and that these theories can be further differentiated by incorporating contrasting as an underlying process involved in specification. In this process, lexical representations evolve from fuzzy to specific, because a sharpening process is triggered when they are contrasted with other, similar, representations. We therefore hypothesized that integrating contrasting in a learning method would facilitate L2 word learning, but that children's linguistic characteristics, such as native vocabulary size and reading skills may mediate this effect.

Our study provides promising initial evidence that contrasting is an effective learning method that has the potential to facilitate L2 word learning in children. After learning orthographically and semantically similar words by contrasting them, children made fewer mistakes when trying to recall these words than children who did not contrast them. These effects persisted a week after learning, which is remarkable given that the learning session only lasted 15 min. Because no differences between contrasting conditions arose for fully dissimilar words, this indicates that contrasting is indeed a process involved in specification, and that our results were not due to a general task effect between groups. As expected, the findings show that the effects of contrasting partly depended on the children's language skills. Specifically, in our study children with lower reading skills were negatively affected by contrasting during learning, but experienced no learning (dis)advantage on later post-tests. Children with higher reading skills were not hindered by contrasting during learning, and recalled significantly more words than all other children on the post-tests.

In the next section, we consider the underlying learning mechanisms in more detail.

Underlying Mechanisms

As mentioned previously, the sharpening of lexical representations by contrasting most likely involves mechanisms of selective attention (Nosofsky, 1986; Goldstone, 1998; Francis et al., 2000; Adini et al., 2002; Francis and Nusbaum, 2002). In the case of language learning, selectively attending to the lexical dimension in which similarity occurs allows more precise representations to be built, which in turn facilitates L2 word learning because the representations are of higher quality (Perfetti and Hart, 2002; Perfetti, 2007; van de Ven et al., 2018; Llompart and Reinisch, 2020).

In our study, contrasting the similar words explicitly required children to focus on the distinctive lexical information in order to discriminate the correct answer from the distractors. In this way, the task required them to sharpen the boundaries between the novel word representations, thereby boosting how efficiently novel representations are sharpened. Children who did not contrast the similar words, were not stimulated to create as precise representations on a trial-by-trial basis, since the clearly distinct orthography and meaning could be used to identify the correct translation. This is reflected by the finding that children who contrasted were less accurate during learning, but more accurate on the post-tests. The children who did not contrast thus had an easier task during learning, but did not undergo the sharpening process to the same extent. Their initial, fuzzier representations were therefore sufficient to complete the learning task successfully, but when they had to retrieve the correct translation during the post-tests, the impreciseness of the representation led to confusion. This is further supported by the kind of errors children who did not contrast made: They came up with a larger number of translations that were not similar to the correct translation (e.g., translating *beak* with the Dutch word for *poison*). By comparison, children who did contrast the similar words not only recalled more words after learning, but the mistakes they made were orthographically or semantically closer to the correct answer (e.g., translating *prison* with the Dutch word for *poison*). This supports the idea that the children who contrasted developed more precise lexical representations.

Given these findings, it is also possible that contrasting benefited learning because errors were more numerous during learning in this condition. Research has extensively shown that making errors largely benefits (word) learning when feedback is provided, particularly in instances where learners are not confident in their answers (Pashler et al., 2005; Metcalfe and Kornell, 2007; Butler et al., 2008). In addition, making related errors during learning has been shown to lead to better retention of materials (Huelser and Metcalfe, 2012). Since incorrect answers in the contrasted condition were always related (i.e., similar) to the target, this might also have played a role during learning.

Our study also revealed that only skilled readers benefitted from contrasting in the multiple-choice method. Since Dutch (L1) vocabulary size did not impact the effects of contrasting, it is likely that contrasting effects in English (L2) were affected not so much by the initial precision of the Dutch representations, but by fundamental aspects that are required for word reading, such as word decoding and word meaning extraction (Gough and Tunmer, 1986; Verhoeven and Perfetti, 2011). Said differently, for less skilled readers, contrasting the three similar alternatives in the multiple-choice task would have been particularly confusing, making them less efficient at visually teasing the orthographically similar words apart or to differentiate the similar meanings during learning. If reading skill level affects children's efficiency of contrasting, adapting the learning task in certain respects for children with lower reading skills might actually be beneficial. For example, simply making the learning phase longer might affect the outcome of contrasting. In addition, providing explicit cues to attend to certain lexical characteristics could also contribute to an increase of contrasting effectiveness. This latter method has been successfully applied in earlier studies to the orthographic dimension, where children with poor reading skills benefited from a method drawing their attention to the relevant grapheme position in minimal pairs of words (McCandliss et al., 2003).

Limitations

In our study, we focused on orthographic and semantic representations. However, as we discussed in the introduction, phonology plays an important role in acquiring new words visually. This is especially true for children, who have less well-developed orthographic and semantic representations (Perfetti, 2007; van Goch, 2016; Janssen, 2017; Meade, 2020). While we did not explicitly offer phonological information in our learning task, we assume that children automatically activated phonology when reading the words (Frost, 1998). Therefore, it is likely that a similar contrasting process implicitly occurred on the phonological dimension during the learning task. Given previous findings (e.g., Janssen et al., 2015; van Goch et al., 2017; van de Ven et al., 2018), contrasting phonological information during learning most likely did contribute to the sharpening of the children's lexical representations and possibly created an additive effect. Future research should explore this issue by adding a condition in which the phonetic sounds of the words are made available during learning.

In addition, our results must be interpreted with some caution, as there is a degree of uncertainty with regards to the magnitude of the effects. Because the overall number of learned words was relatively low (befitting the short learning duration), the differences between conditions in terms of words learned is also only a one-word difference (from 5 to 6 words). Percentage-wise, however, the difference translates to a 20% increase relative to all similar words learned. It remains to be determined, though, whether the effects we observed are absolute, i.e., whether when the item sample size is increased the one-word difference would remain the same or scale up correspondingly. If the latter is the case, this would represent significant learning gains potential. In order to determine the actual magnitude of the effects, follow-up studies with increased power in terms of participants and items are required. Additionally, research has shown that a larger number of repetitions positively impact vocabulary learning outcomes (Webb, 2007). Therefore, it is possible that a longer learning phase leads to greater learning gains, particularly for less skilled readers. A longer learning task would allow testing vocabulary knowledge in a more challenging manner, for instance by asking learners to type out the L2 translations of the words. This could offer additional insight into the nature of their lexical representations.

Finally, on a related and practical note, there are some drawbacks to conducting research in ecologically valid settings such as schools, for instance the possibility of distraction during the experiment. However, in our study, we nevertheless obtained significant results in such an environment. While laboratory studies may offer more controlled insights into the mechanisms of contrasting as a learning method, our study highlights its effectiveness in instructional settings.

Future Directions

Our study has shed light on an important process involved in lexical specification. Contrasting offers several opportunities for future research aimed to gain a more comprehensive understanding of L2 word learning. In particular, it would be valuable to replicate this research with different measuring methods, in order to obtain more detailed insight into the attentional mechanisms underlying contrasting. For instance, to find out more about the lexical aspects learners attend to while they are learning by contrasting, tracking their eye movements would offer valuable additional information (cf. Llompart and Reinisch, 2020).

Furthermore, more work is needed to determine the precise circumstances under which contrasting facilitates learning, in particular in relation to the linguistic contexts and individual characteristics. To gain a fuller understanding of the effects of contrasting in different linguistic settings, future research could explore its effects on different L1-L2 pairings (such as languages with less cross-linguistic similarity), non-alphabetic languages, or even consider the effects of contrasting several languages simultaneously. Moreover, particularly for target learners with linguistic difficulties (e.g., lower reading skills), testing how the learning method can be optimized is useful. We suggest that such research should be teacher-led (Churches and Dommett, 2016), because of teachers' practical insights on how learning methods could be fine-tuned to the learner group at hand. This would also help bridge the gap between L2 language learning research and instructional practices. In turn, this would enable researchers to build even more comprehensive models of L2 vocabulary learning.

Finally, our work offers novel insight into how L2 vocabulary instruction can be optimized by boosting lexical specification. Multiple-choice is already used as a digital learning tool (see Nakata, 2011). This makes the step of adding a contrasting feature easy. The next step would be to make such a tool adaptive to the learner. For instance, features could be included such as the ascertainment of an optimal difficulty level depending on the learners' language skills, or the addition of certain lexical cues for learners who require them. Such a learning tool could also be adapted differently for early and late learners, given that contrasting may show differential effects depending on the linguistic development stage of the learners (cf. Baxter et al., under review).

CONCLUSION

We provide initial evidence that contrasting, a process essential for the differentiation of L2 lexical representations can, under specific circumstances, effectively be exploited in a teaching method to facilitate L2 word learning in children by sharpening their representations. Our study extends existing theories that propose a gradual lexical specification process, such as the fuzzy lexicon hypothesis (Cook et al., 2016), as well as those positing a causal link between specificity of representations and retrieval efficiency, such as the lexical quality hypothesis (Perfetti and Hart, 2002). In particular, our study takes a step further by offering insights into how the specification of L2 representations can be made more efficient for different dimensions and participant populations. As we see it, our findings offer a starting point to contribute to the successful development of more comprehensive theoretical models of L2 vocabulary learning, and to direct applications for instructional practice.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee Social Sciences (ECSS) of the Radboud University Nijmegen. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

HB, MD, MH, TD, and FL acquired the funding for the project. FL and PB designed the study. PB designed the stimuli, programmed the experiment, acquired and analyzed the data, and wrote the draft of the manuscript. PB, MD, MH, and TD developed the theoretical framework for the manuscript. MD, MH, and TD revised the early stages of the manuscript. All authors contributed to reviewing the final stages of the manuscript, approved the final manuscript, and conceptualized the study.

FUNDING

This project was funded by the Nationaal Regieorgaan Onderwijsonderzoek, project number 405-17-304.

ACKNOWLEDGMENTS

The authors would like to thank all of the children who took part in this study. They also extend a warm thank you to all of the schools, teachers, and parents who allowed for this study to take place. Finally, the authors also thank V. van den Berg, W. Harmsen, M. van de Kraats, and M. Prins for their help in collecting the data, as well as L. Billen, M. Levenig, and N. Sommer for helping pre-process the data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.688160/full#supplementary-material

REFERENCES

Adini, Y., Sagi, D., and Tsodyks, M. (2002). Context-enabled learning in the human visual system. *Nature* 415, 790–793. doi: 10.1038/415790a

Baddeley, A. (1992). Working memory. *Science* 255, 556–559. doi: 10.1126/science.1736359

Baddeley, A., Gathercole, S., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev.* 105, 158–173. doi: 10.1037/0033-295X.105.1.158

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., and Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: long-term retention of learning in perception and production. *Perception Psychophys.* 61, 977–985. doi: 10.3758/BF03206911

Brysbaert, M., and Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Res. Methods* 49, 1520–1523. doi: 10.3758/s13428-016-0811-4

Butler, A. C., Karpicke, J. D., and Roediger, H. L. (2008). Correcting a metacognitive error: feedback increases retention of low-confidence correct responses. *J. Exp. Psychol. Learn. Memory Cognition* 34, 918–928. doi: 10.1037/0278-7393.34.4.918

Castles, A., Davis, C., Cavalot, P., and Forster, K. (2007). Tracking the acquisition of orthographic skills in developing readers: masked priming effects. *J. Exp. Child Psychol.* 97, 165–182. doi: 10.1016/j.jecp.2007.01.006

Churches, R., and Dommett, E. (2016). *Teacher-Led Research: Designing and Implementing Randomised Controlled Trials and Other Forms of Experimental Research*. Camarthen: Crown House Publishing.

Clark, E. V. (2004). How language acquisition builds on cognitive development. *Trends Cognitive Sci.* 8, 472–478. doi: 10.1016/j.tics.2004.08.012

Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychol. Rev.* 108, 204–256. doi: 10.1037/0033-295X.108.1.204

Cook, S. V. (2012). *Phonological Form in L2 Lexical Access: Friend or Foe?* Unpublished doctoral dissertation, University of Maryland.

Cook, S. V., and Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *Mental Lexicon* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., PandŽa, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., et al. (2018). MultiPic: a standardized set of 750 drawings with norms for six European languages. *Quarterly J. Exp. Psychol.* 71, 808–816. doi: 10.1080/17470218.2017.1310261

Dunn, L. M., Dunn, L. M., and Schlichting, J. E. P. T. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam: Netherlands: Pearson.

Elgort, I., Perfetti, C. A., Rickles, B., and Stafura, J. Z. (2015). Contextual learning of L2 word meanings: second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language Cogn. Neurosci.* 30, 506–528. doi: 10.1080/23273798.2014.942673

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception Psychophys.* 62, 1668–1680. doi: 10.3758/BF03212164

Francis, A. L., and Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *J. Exp. Psychol. Human Percep. Perform.* 28, 349–366. doi: 10.1037/0096-1523.28.2.349

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychol. Bull.* 123, 71–99. doi: 10.1037/0033-2909.123.1.71

Goldstone, R. L. (1998). Perceptual learning. *Annual Rev. Psychol.* 49, 585–612. doi: 10.1146/annurev.psych.49.1.585

Gough, P. B., and Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial Special Educ.* 7, 6–10. doi: 10.1177/074193258600700104

Grainger, J., and Van Heuven, W. J. B. (2004). "Modeling letter position coding in printed word perception," in *The Mental Lexicon*, ed Bonin P. (New-York, NY: Nova Science), 1–23.

Huelser, B. J., and Metcalfe, J. (2012). Making related errors facilitates learning, but learners do not know it. *Memory Cogn.* 40, 514–527. doi: 10.3758/s13421-011-0167-z

Ishii, T. (2015). Semantic connection or visual connection: Investigating the true source of confusion. *Language Teach. Res.* 19, 712–722. doi: 10.1177/1362168814559799

Janssen, C. (2017). *Phonological Foundations of Early Literacy in First and Second Language Learners*. Doctoral dissertation, Radboud University Nijmegen.

Janssen, C., Segers, E., McQueen, J. M., and Verhoeven, L. (2015). Lexical specificity training effects in second language learners. *Language Learn.* 65, 358–389. doi: 10.1111/lang.12102

Kormos, J., and Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism* 11:261. doi: 10.1017/S1366728908003416

Krause, F., and Lindemann, O. (2014). Expyriment: a Python library for cognitive and neuroscientific experiments. *Behav. Res. Methods* 46, 416–428. doi: 10.3758/s13428-013-0390-6

Krom, R., Jongen, I., Verhelst, N., Kamphuis, F., and Kleintjes, F. (2010). *Wetenschappelijke Verantwoording DMT en AVI*. Arnhem, Netherlands: Cito.

Lancaster, A., and Gor, K. (2016). Abstraction of phonological representations in adult nonnative speakers. *Proc. Linguistic Soc. Am.* 24:1–15. doi: 10.3765/plsa.v1i0.3725

Laufer, B. (1988). The concept of "synforms" (similar lexical forms) in vocabulary acquisition. *Language Educ.* 2, 113–132. doi: 10.1080/09500788809541228

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Appl. Psycholinguist.* 32:69. doi: 10.1017/S0142716410000299

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Phys. Doklady* 10, 707–710.

Linck, J. A., Osthus, P., Koeth, J. T., and Bunting, M. F. (2014). Working memory and second language comprehension and production: a meta-analysis. *Psychonomic Bull. Rev.* 21, 861–883. doi: 10.3758/s13423-013-0565-2

Little, J. L., and Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory Cogn.* 43, 14–26. doi: 10.3758/s13421-014-0452-8

Little, J. L., Bjork, E. L., Bjork, R. A., and Angello, G. (2012). Multiple-choice tests exonerated, at least of some charges: fostering test-induced learning and avoiding test-induced forgetting. *Psychol. Sci.* 23, 1337–1344. doi: 10.1177/0956797612443370

Llach M. P. A. (2015). Lexical errors in writing at the end of primary and secondary education: description and pedagogical implications. *Porta Linguarum* 23, 109–124. doi: 10.30827/Digibug.53758

Llompart, M., and Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *J. Exp. Psychol. Learn. Memory Cogn.* 46, 1590–1610. doi: 10.1037/xlm0000832

Logan, J. S., Lively, S. E., and Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: a first report. *J. Acoustical Soc. Am.* 89, 874–886. doi: 10.1121/1.1894649

Mandera, P., Keuleers, E., and Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: a review and empirical validation. *J. Memory Language* 92, 57–78. doi: 10.1016/j.jml.2016.04.001

McCandliss, B., Beck, I. L., Sandak, R., and Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: design and preliminary tests of the word building intervention. *Sci. Stud. Read.* 7, 75–104. doi: 10.1207/S1532799XSSR0701_05

Meade, G. (2020). The role of phonology during visual word learning in adults: an integrative review. *Psychonomic Bull. Rev.* 27, 15–23. doi: 10.3758/s13423-019-01647-0

Metcalfe, J., and Kornell, N. (2007). Principles of cognitive science in education: the effects of generation, errors, and feedback. *Psychon. Bull. Rev.* 14, 225–229. doi: 10.3758/BF03194056

Meteyard, L., and Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *J. Memory Language* 112:104092. doi: 10.1016/j.jml.2020.104092

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory Cogn.* 25, 47–56. doi: 10.3758/BF03197284

Metsala, J. L., and Walley, A. C. (1998). "Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability," in *Word Recognition in Beginning Literacy*, eds J. L. Metsala and L.C. Ehri (Mahwah, NJ: Erlbaum), 89–120.

Nakata, T. (2011). Computer-assisted second language vocabulary learning in a paired-associate paradigm: a critical investigation of flashcard software. *Computer Assisted Language Learn.* 24, 17–38. doi: 10.1080/09588221.2010.520675

Nakata, T., and Webb, S. (2016). Does studying vocabulary in smaller sets increase learning. *Stud. Sec. Lang. Acquisit*. 38, 523–552. doi: 10.1017/S0272263115000236

Nation, K., and Cocksey, J. (2009). Beginning readers activate semantics from sub-word orthography. *Cognition* 110, 273–278. doi: 10.1016/j.cognition.2008.11.004

Nation, P. (2000). Learning vocabulary in lexical sets: dangers and guidelines. *TESOL J.* 9, 6–10. doi: 10.1002/j.1949-3533.2000.tb00239.x

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *J. Exp. Psychol.* 115, 39–57. doi: 10.1037/0096-3445.115.1.39

Papathanasiou, E. (2009). An investigation of two ways of presenting vocabulary. *ELT J.* 63, 313–322. doi: 10.1093/elt/ccp014

Pashler, H., Cepeda, N. J., Wixted, J. T., and Rohrer, D. (2005). When does feedback facilitate learning of words? *J. Exp. Psychol. Learn. Memory Cogn.* 31, 3–8. doi: 10.1037/0278-7393.31.1.3

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730

Perfetti, C. A., and Hart, L. (2002). The lexical quality hypothesis. *Precursors Functional Literacy* 11, 67–86. doi: 10.1075/swll.11.14per

Roediger, H. L. 3rd, and Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends Cogn. Sci.* 15, 20–27. doi: 10.1016/j.tics.2010.09.003

Saan, R. J., and Deelman, B. G. (1986). *De 15-woordentest A en B (een voorlopige handleiding)*. Groningen: Department of Neuropsychology, Academic Hospital Groningen.

Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010

Strenze, T. (2007). Intelligence and socioeconomic success: a meta-analytic review of longitudinal research. *Intelligence* 35, 401–426. doi: 10.1016/j.intell.2006.09.004

Suanda, S. H., Mugwanya, N., and Namy, L. L. (2014). Cross-situational statistical word learning in young children. *J. Exp. Child Psychol.* 126, 395–411. doi: 10.1016/j.jecp.2014.06.003

Tinkham, T. (1993). The effect of semantic clustering on the learning of second language vocabulary. *System* 21, 371–380. doi: 10.1016/0346-251X(93)90027-E

van de Ven, M., Segers, E., and Verhoeven, L. (2018). Enhanced second language vocabulary learning through phonological specificity training in adolescents: vocabulary learning through phonological training. *Language Learn.* 52:323. doi: 10.1111/lang.12330

van Goch, M. M. (2016). *Lexical Specificity: The Importance of Rich Phonological Representations When Learning to Read*. Doctoral dissertation, Radboud University Nijmegen.

van Goch, M. M., McQueen, J. M., and Verhoeven, L. (2014). Learning phonologically specific new words fosters rhyme awareness in dutch preliterate children. *Sci. Stud. Read.* 18, 155–172. doi: 10.1080/10888438.2013.827199

van Goch, M. M., Verhoeven, L., and McQueen, J. M. (2017). Trainability in lexical specificity mediates between short-term memory and both vocabulary and rhyme awareness. *Learn. Ind. Diff*. 57, 163–169. doi: 10.1016/j.lindif.2017.05.008

van Heuven, W. J. B., Dijkstra, T., and Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *J. Memory Language* 39, 458–483. doi: 10.1006/jmla.1998.2584

Verhoeven, L., and Perfetti, C. A. (2011). Introduction to this special issue: vocabulary growth and reading skill. *Sci. Stud. Read.* 15, 1–7. doi: 10.1080/10888438.2011.536124

Waring, R. (1997). The negative effects of learning words in semantic sets: a replication. *System* 25, 261–274. doi: 10.1016/S0346-251X(97)00013-4

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Appl. Linguistics* 28, 46–65. doi: 10.1093/applin/aml048

Ziegler, J. C., and Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychol. Bull.* 131, 3–29. doi: 10.1037/0033-2909.131.1.3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 08 July 2021
doi: 10.3389/fpsyg.2021.689052

On the Locus of L2 Lexical Fuzziness: Insights From L1 Spoken Word Recognition and Novel Word Learning

*Efthymia C. Kapnoula**

Basque Center on Cognition, Brain and Language, San Sebastián, Spain

The examination of how words are learned can offer valuable insights into the nature of lexical representations. For example, a common assessment of novel word learning is based on its ability to interfere with other words; given that words are known to compete with each other (Luce and Pisoni, 1998; Dahan et al., 2001), we can use the capacity of a novel word to interfere with the activation of other lexical representations as a measure of the degree to which it is integrated into the mental lexicon (Leach and Samuel, 2007). This measure allows us to assess novel word learning in L1 or L2, but also the degree to which representations from the two lexica interact with each other (Marian and Spivey, 2003). Despite the somewhat independent lines of research on L1 and L2 word learning, common patterns emerge across the two literatures (Lindsay and Gaskell, 2010; Palma and Titone, 2020). In both cases, lexicalization appears to follow a similar trajectory. In L1, newly encoded words often fail at first to engage in competition with known words, but they do so later, after they have been better integrated into the mental lexicon (Gaskell and Dumay, 2003; Dumay and Gaskell, 2012; Bakker et al., 2014). Similarly, L2 words generally have a facilitatory effect, which can, however, become inhibitory in the case of more robust (high-frequency) lexical representations. Despite the similar pattern, L1 lexicalization is described in terms of inter-lexical connections (Leach and Samuel, 2007), leading to more automatic processing (McMurray et al., 2016); whereas in L2 word learning, lack of lexical inhibition is attributed to less robust (i.e., fuzzy) L2 lexical representations. Here, I point to these similarities and I use them to argue that a common mechanism may underlie similar patterns across the two literatures.

Keywords: word learning, fuzzy lexicon, mental lexicon, lexical representation, lexical representation and processing

OPEN ACCESS

Edited by:
Kira Gor,
University of Maryland, College Park, United States

Reviewed by:
Miquel Simonet,
University of Arizona, United States

Correspondence:
Efthymia C. Kapnoula
kapnoula@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *31 March 2021*
Accepted: *15 June 2021*
Published: *08 July 2021*

Citation:
Kapnoula EC (2021) On the Locus of L2 Lexical Fuzziness: Insights From L1 Spoken Word Recognition and Novel Word Learning. Front. Psychol. 12:689052. doi: 10.3389/fpsyg.2021.689052

A THEORETICAL FRAMEWORK FOR EVALUATING LEXICALIZATION

Knowing a word means it is part of one's mental lexicon. Thus, learning a new word requires integrating its representation in the mental lexicon in a way that allows it to be accessed (recognized and produced) in real time. According to Leach and Samuel (2007), this integration can be described as the acquisition of two lexical properties. *Lexical configuration* refers to the minimum amount of information required to "know" a word-form, which allows listeners to recognize it. This property consists of bottom-up pathways that map acoustic or phonological information to words (upward arrows in **Figure 1**). Then, *lexical engagement* refers to how a word interacts with other words (links between words in **Figure 1**), or lower level representations (top-down connections in **Figure 1**).

Indeed, there is robust evidence for lexical engagement, both in the form of words inhibiting each other (Luce and Pisoni, 1998; Vitevitch and Luce, 1998; Dahan et al., 2001), and in the form of top-down flow of information affecting perception of speech sounds in real-time (Magnuson et al., 2003; Samuel and Pitt, 2003; Luthra et al., 2021) and over the course of learning (Norris et al., 2003; Kraljic and Samuel, 2006).

Within this framework, we can use these two lexical properties to assess novel word learning. That is, we know that real words can affect the perception of speech sounds (Samuel and Kraljic, 2009; Luthra et al., 2021). For example, Warren (1970) showed that if we take a word (e.g., "legislature") and replace one speech sound (e.g., the /s/) with a cough sound, listeners report that the original sound is there. This is known as *phonemic restoration* (see also Samuel, 1996). Another example is *perceptual learning* (Norris et al., 2003). Here we replace a speech sound with an ambiguous sound (e.g., we replace the /s/ in "personal" with a sound in-between /s/ and /ʃ/). If participants are exposed to many words like this, they learn to perceive the ambiguous sound as an /s/. In both cases, the effect can only be driven by real words. This means that, if a novel word can drive such top-down effects, this can be taken as evidence for lexicalization. Indeed, Leach and Samuel (2007) used this assessment to examine how several factors affect word learning. Participants learned a number of novel words and then it was assessed how well those items were integrated into the lexicon by measuring their ability to affect the perception of speech sounds (by driving phonemic restoration and perceptual learning). New words acted as real words in driving these effects, but only in some cases, depending on the details of the training procedure. Thus, this kind of lexicality test can help us assess which training works better and offer insights into the process of lexicalization.

Following a similar rationale, since known words compete with each other, we can use the capacity of a novel word to interfere with other lexical representations as a measure of the degree to which it is integrated into the mental lexicon. For example, Gaskell and Dumay (2003) examined the conditions under which newly learned words form inhibitory links with known words. Participants learned new words that overlapped with real L1 words (e.g., novel word: "cathudruke" overlapping with known word: "cathedral"). The results showed that newly learned words did not interfere with the recognition of their known-word competitors immediately after learning, but they did so after 3 days of training (see also Dumay and Gaskell, 2007; Bakker et al., 2014; Kapnoula et al., 2015; Kapnoula and McMurray, 2016a, for similar use of lexical competition as evidence for lexicalization). In addition, a reversal of the effect has been observed at the earliest stages of learning, with new words facilitating the recognition of similar-sounding words (Dumay and Gaskell, 2012). Thus, a shift from facilitation to inhibition is thought to reflect lexicalization.

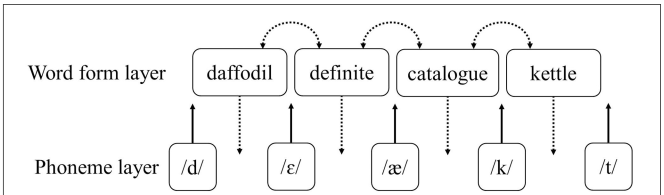


FIGURE 1 | Visualization of lexical configuration versus lexical engagement properties in the context of spoken word recognition. Note that representations are depicted as localist nodes for ease of visualization, but no theoretical commitment is made regarding their nature.

These results demonstrate how different training parameters can lead to different outcomes in terms of how well a new word is integrated into the mental lexicon. In turn, the degree of lexical integration has implications for real-time recognition; well-integrated words are better (i.e., more automatically) recognized (for a review on the relationship between lexical integration and recognition automaticity, see McMurray et al., 2016). Critically, differences in how word recognition unfolds in real time are observed well beyond the initial stages of learning. For example, divergence from typical L1 spoken word recognition has been reported for individuals with specific language impairment (McMurray et al., 2010), developmental language disorder (McMurray et al., 2019b), and cochlear-implant users (McMurray et al., 2019a), while even within typically developing/hearing individuals, the way in which spoken words are recognized in real time changes over development (Rigler et al., 2015). These results, suggest that automaticity of word recognition can vary even amongst well-known, familiar words. In line with this idea, a study by Kapnoula and McMurray (2016b) found that the real-time dynamics of L1 word recognition are malleable. Participants were exposed to familiar words and each one was assigned to one of two experimental groups; in the high-competition group, pairs of similar-sounding familiar words (e.g., "net" and "neck") were presented close together (temporally and/or spatially) in a manner that required participants to resolve the competition between them. In contrast, in the low-competition group, co-activation of words in each pair was minimized. After a 40-min exposure phase, the authors used a visual world paradigm task to track the time-course of lexical competition between words in each pair. They found that only participants in the high-competition group were able to fully suppress the activation of the competitor word. Moreover, computer simulations (using jTRACE; Strauss et al., 2007) pointed to increased inter-lexical inhibition as the parameter that helped participants in the high-competition group better suppress competitors.

Based on the studies presented above, we can conclude the following: First, it is broadly accepted that lexicality (i.e., lexical status) can be defined on the basis of how well a word is interlinked with other representations (e.g., other words) and that assessing the formation of these links can help us evaluate the degree to which a novel word has been learned. Second, such links are malleable, even for well-known L1 words, in the sense that they can be fine-tuned, possibly to accommodate short- and long-term demands of the language comprehension system. How is this framework relevant to L2 word learning?

EVALUATING LEXICALIZATION IN L2

To address this question, one must take into account the additional factor of phonological differences between L1 and L2; that is, non-native listeners often have to learn to distinguish between words based on L2 phonological contrasts that do not exist in their native language (Cutler and Otake, 2004; Weber and Cutler, 2004). For example, Dutch listeners find it difficult to differentiate between the English phonemes /æ/ and /ε/, which means they likely activate both "definite" and "daffodil" when hearing /daef/. Indeed, using a cross-modal priming paradigm, Broersma and Cutler (2011) found that hearing /daef / facilitated visual recognition of the word "definite" for Dutch, but not for native English listeners. This pattern of results is taken as evidence for *phantom activation* in L2 word recognition, which refers to the activation of irrelevant words that are treated by the system as lexical competitors due to phonological confusability. Interestingly though, this increased competitor activation does not necessarily lead to increased inhibition of the target word. Specifically, Broersma (2012) showed that for native English speakers, hearing "deficit" inhibited subsequent visual recognition of the word "daffodil," but for Dutch speakers, its effect was *facilitatory* to the same degree as hearing the target word ("daffodil").

This seemingly paradoxical pattern of results has been explained in terms of *fuzzy lexical representations* (Darcy et al., 2013; Cook and Gor, 2015; Gor, 2018; closely linked to the Lexical Quality hypothesis, Perfetti, 2007). According to this hypothesis, some L2 words are encoded in the mental lexicon in a phonolexically underdifferentiated (i.e., fuzzy) way (Gor, 2018). This happens when words include phonemes that belong to non-native contrasts, which makes them easily confusable for L2 listeners (e.g., the /æ/ and /ε/ contrast that does not exist in Dutch). In those cases, L2 listeners activate similar-sounding words – as is the case with "daffodil" being activated when L2 listeners hear "deficit." Despite the increased number of competitors, their fuzziness makes them poor inhibitors. At the same time, the cumulative sublexical activation is facilitatory, leading to a facilitatory net effect. This pattern is also in line with work on L1 word recognition showing independent and opposite effects at the lexical and sublexical levels (Vitevitch and Luce, 1998, 1999).

Indeed, there is growing support for the idea that L2 lexical representations can be fuzzily encoded due to perceptual confusability at the phoneme level, with the key finding consisting in non-native facilitation in priming tasks (Ota et al., 2009; Gor et al., 2010; Cook and Gor, 2015; Cook et al., 2016; Gor, 2018; Gor and Cook, 2020). Moreover, L1 phonology appears to be relevant, even when processing takes place in the visual modality (Ota et al., 2009) – a finding that offers support for the idea that L2 lexical representations are shaped by L1 phonology. Lastly, this effect is more robust for less familiar/low-frequency words. In contrast, when L2 prime words are well known (i.e., highly familiar and/or frequent), they seem to drive an inhibitory effect, similar to that observed in native speakers, a modulation that has been attributed to decreased lexical fuzziness of high-frequency primes (Cook and Gor, 2015; Gor and Cook, 2020).

FUZZY REPRESENTATIONS AND FUZZY CONNECTIONS

Bringing the two lines of work together, we can think of how they fit together and how L2 word learning effects such as phantom activation can be explained within the theoretical framework described earlier.

First, there is a striking similarity between the two literatures; in both cases, robust lexicalization is manifested as an inhibitory effect (see also Marian and Spivey, 2003; Qiao and Forster, 2017). However, in L1 word learning, inhibition is attributed to robust inter-lexical connections (i.e., lexical engagement); whereas in L2 word learning, inhibition is thought to reflect higher-resolution/less fuzzy encoding. The two accounts differ in perhaps subtle, but theoretically important ways. In the first case, the quality of lexical representations is not solely defined by how well encoded they are (which would fit under the lexical configuration property); rather lexical quality is also determined by the links between a word and other representations and, thus may be better described as an emergent property of lexical processing. In that respect, the two accounts are not theoretically incompatible; indeed, a word could be both fuzzily encoded and weakly interconnected with other words. In fact, it makes sense that fuzzy lexical encoding would lead to weak inter-lexical connections (both for L2, but also less familiar L1 words); however, the reverse is not guaranteed–that is, weak inter-lexical connections are not necessarily due to fuzzy encoding.

Second, within a framework such as the one described for L1 word learning, words with ambiguous phonemes (as is the case with difficult, non-native contrasts) are expected to have connections of similar strength with both speech categories, because the categories themselves are not well separated. In that sense, phantom activation effects could again be attributed to less robust lexical engagement in the form of weak links between a word and its phonemes. That is, assuming a system such as the one shown in **Figure 1**, in which there is interactive activation between the lexical and sublexical layers (McClelland and Elman, 1986; Luthra et al., 2021), activation of "daffodil" should spread to both /æ/ and /ε/ categories for Dutch speakers, which in turn would strengthen activation of "definite." Moreover, this sequence is expected to take place independent of the modality in which lexical activation is originally triggered (auditory or visual), making this account also compatible with cross-modal effects (Ota et al., 2009).

In sum, phantom activation, priming facilitation, and modulation of the priming effect by word frequency are all well-established effects in L2 word recognition and they are commonly attributed to the fuzzy encoding of L2 lexical representations. However, I argue that these effects can also be explained in terms of processing automaticity (McMurray et al., 2016) and lexical engagement (Leach and Samuel, 2007).

CONCLUSION

My goal was to highlight similar patterns across the literatures on L1 and L2 word learning and contribute to the effort of drawing connections between them (Lindsay and Gaskell, 2010; Palma and Titone, 2020). In doing so, I focused on a set of behavioral effects that are commonly attributed to fuzzy L2 lexical representations and I briefly described how these effects could be explained within a different theoretical framework, taken from the L1 word learning literature. It is important to note that the two accounts are not mutually exclusive and that it would be difficult to experimentally disentangle between the two. Rather than arguing for one mechanism over another, the purpose of this piece is to urge both sides to work closer together, considering that a common mechanism may (at least partly) underlie similar patterns across the two literatures.

AUTHOR CONTRIBUTIONS

ECK confirms being solely responsible for the conception and drafting of this work and has approved the final manuscript for publication.

FUNDING

Support for this project was provided by the Spanish Ministry of Economy and Competitiveness, through the Juan de la Cierva-Formación fellowship, # FJCI-2016- 28019, awarded to ECK. This work was partially supported by the Basque Government through the BERC 2018-2021 program, and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490. This project has received funding from the European Union's Horizon 2020 Research and Innovation Program, under the Marie Skłodowska-Curie grant agreement No 793919, awarded to ECK.

REFERENCES

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., and McQueen, J. M. (2014). Competition from unseen or unheard novel words: lexical consolidation across modalities. *J. Mem. Lang.* 73, 116–130. doi: 10.1016/j.jml.2014.03.002

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Lang. Cogn. Process.* 27, 1205–1224. doi: 10.1080/01690965.2012.660170

Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174

Cook, S. V., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *Ment. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Cutler, A., and Otake, T. (2004). Pseudo-homophony in non-native listening. *J. Acoust. Soc. Am.* 115, 2392–2392. doi: 10.1121/1.4780547

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Lang. Cogn. Process.* 16, 507–534. doi: 10.1080/01690960143000074

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Ment. Lex.* 8, 372–420. doi: 10.1075/ml.8.3.06dar

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

Dumay, N., and Gaskell, M. G. (2012). Overnight lexical consolidation revealed by speech segmentation. *Cognition* 123, 119–132. doi: 10.1016/j.cognition.2011.12.009

Gaskell, M. G., and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition* 89, 105–132. doi: 10.1016/S0010-0277(03)00070-2

Gor, K. (2018). Phonological priming and the role of phonology in nonnative word recognition. *Bilingualism* 21, 437–442. doi: 10.1017/S1366728918000056

Gor, K., Cook, S., and Jackson, S. (2010). "Lexical access in highly proficient late L2 learners: evidence from semantic and phonological auditory priming," in *Paper Presented at theSecond Language Research Forum (SLRF)*, University of Maryland.

Gor, K., and Cook, S. V. (2020). A mare in a pub? Nonnative facilitation in phonological priming. *Second Lang. Res.* 36, 123–140. doi: 10.1177/0267658318769962

Kapnoula, E. C., and McMurray, B. (2016a). Newly learned word-forms are abstract and integrated immediately after acquisition. *Psychon. Bull. Rev.* 23, 491–499. doi: 10.3758/s13423-015-0897-1

Kapnoula, E. C., and McMurray, B. (2016b). Training alters the resolution of lexical interference: evidence for plasticity of competition and inhibition. *J. Exp. Psychol. Gen.* 145, 8–30. doi: 10.1037/xge0000123

Kapnoula, E. C., Packard, S., Gupta, P., and McMurray, B. (2015). Immediate lexical integration of novel word forms. *Cognition* 134, 85–99. doi: 10.1016/j.cognition.2014.09.007

Kraljic, T., and Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychon. Bull. Rev.* 13, 262–268. doi: 10.3758/bf03193841

Leach, L., and Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cogn. Psychol.* 55, 306–353. doi: 10.1016/j.cogpsych.2007.01.001

Lindsay, S., and Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Lang. Learn.* 60(Suppl. 2), 45–63. doi: 10.1111/j.1467-9922.2010.00600.x

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

Luthra, S., Peraza-Santiago, G., Beeson, K., Saltzman, D., Crinnion, A. M., and Magnuson, J. S. (2021). Robust lexically-mediated compensation for coarticulation: christmash time is here again. *Cogn. Sci.* 45, 1–20. doi: 10.1007/978-1-84457-843-6_1

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., and Dahan, D. (2003). The time course of spoken word learning and recognition: studies with artificial lexicons. *J. Exp. Psychol. Gen.* 132, 202–227. doi: 10.1037/0096-3445.132.2.202

Marian, V., and Spivey, M. (2003). Competing activation in bilingual language processing: within-and between-language competition. *Biling. Lang. Cogn.* 60, 97–115. doi: 10.1017/s1366728903001068

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McMurray, B., Ellis, T. P., and Apfelbaum, K. S. (2019a). How do you deal with uncertainty? Cochlear implant users differ in the dynamics of lexical processing of noncanonical inputs. *Ear Hear.* 40, 961–980. doi: 10.1097/AUD.0000000000000681

McMurray, B., Kapnoula, E. C., and Gaskell, M. G. (2016). "Learning and integration of new word-forms: consolidation, pruning, and the emergence of automaticity," in *Speech Perception and Spoken Word Recognition*, eds M. G. Gaskell and J. Mirkoviæ (London: Psychology Press), doi: 10.4324/9781315772110

McMurray, B., Klein-Packard, J., and Tomblin, J. B. (2019b). A real-time mechanism underlying lexical deficits in developmental language disorder: between-word inhibition. *Cognition* 191:104000. doi: 10.1016/j.cognition.2019.06.012

McMurray, B., Samelson, V. M., Lee, S. H., and Tomblin, J. B. (2010). Individual differences in online spoken word recognition: implications for SLI. *Cogn. Psychol.* 60, 1–39. doi: 10.1016/j.cogpsych.2009.06.003

Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cogn. Psychol.* 47, 204–238. doi: 10.1016/S0010-0285(03)00006-9

Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition* 111, 263–269. doi: 10.1016/j.cognition.2008.12.007

Palma, P., and Titone, D. (2020). Something old, something new: a review of the literature on sleep-related lexicalization of novel words in adults. *Psychon. Bull. Rev.* 28, 96–121. doi: 10.3758/s13423-020-01809-5

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730

Qiao, X., and Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition* 158, 147–152. doi: 10.1016/j.cognition.2016.10.026

Rigler, H., Farris-Trimble, A., Greiner, L., Walker, J., Tomblin, J. B., and McMurray, B. (2015). The slow developmental timecourse of real-time spoken word recognition. *Dev. Psychol.* 51, 1690–1703. doi: 10.1037/dev0000044

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *J. Exp. Psychol. Gen.* 125, 28–51. doi: 10.1037/0096-3445.125.1.28

Samuel, A. G., and Kraljic, T. (2009). Perceptual learning for speech. *Atten. Percept. Psychophys.* 71, 1207–1218. doi: 10.3758/APP.71.6.1207

Samuel, A. G., and Pitt, M. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *J. Mem. Lang.* 48, 416–434. doi: 10.1016/s0749-596x(02)00514-4

Strauss, T. J., Harris, H. D., and Magnuson, J. S. (2007). jTRACE: a reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behav. Res. Methods* 39, 19–30. doi: 10.3758/bf03192840

Vitevitch, M. S., and Luce, P. A. (1998). When words compete: levels of processing in perception of spoken words. *Psychol. Sci.* 9, 325–329. doi: 10.1111/1467-9280.00064

Vitevitch, M. S., and Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Mem. Lang.* 40, 374–408. doi: 10.1006/jmla.1998.2618

Warren, R. (1970). Perceptual restoration of missing speech sounds. *Science* 167, 392–393. doi: 10.1126/science.167.3917.392

Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/s0749-596x(03)00105-0

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 20 July 2021
doi: 10.3389/fcomm.2021.689470

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

Shannon Barrios and Rachel Hayes-Harb*

Department of Linguistics, University of Utah, Salt Lake City, UT, United States

Second language (L2) learners often exhibit difficulty perceiving novel phonological contrasts and/or using them to distinguish similar-sounding words. The auditory lexical decision (LD) task has emerged as a promising method to elicit the asymmetries in lexical processing performance that help to identify the locus of learners' difficulty. However, LD tasks have been implemented and interpreted variably in the literature, complicating their utility in distinguishing between cases where learners' difficulty lies at the level of perceptual and/or lexical coding. Building on previous work, we elaborate a set of LD ordinal accuracy predictions associated with various logically possible scenarios concerning the locus of learner difficulty, and provide new LD data involving multiple contrasts and native language (L1) groups. The inclusion of a native speaker control group allows us to isolate which patterns are unique to L2 learners, and the combination of multiple contrasts and L1 groups allows us to elicit evidence of various scenarios. We present findings of an experiment where native English, Korean, and Mandarin speakers completed an LD task that probed the robustness of listeners' phonological representations of the English /æ/-/ɛ/ and /l/-/ɹ/ contrasts. Words contained the target phonemes, and nonwords were created by replacing the target phoneme with its counterpart (e.g., *lecture/*[ɹ]ecture*, *battle/*b[ɛ]ttle*). For the /æ/-/ɛ/ contrast, all three groups exhibited the same pattern of accuracy: near-ceiling acceptance of words and an asymmetric pattern of responses to nonwords, with higher accuracy for nonwords containing [æ] than [ɛ]. For the /l/-/ɹ/ contrast, we found three distinct accuracy patterns: native English speakers' performance was highly accurate and symmetric for words and nonwords, native Mandarin speakers exhibited asymmetries favoring [l] items for words and nonwords (interpreted as evidence that they experienced difficulty at the perceptual coding level), and native Korean speakers exhibited asymmetries in opposite directions for words (favoring [l]) and nonwords (favoring [ɹ]; evidence of difficulty at the lexical coding level). Our findings suggest that the auditory LD task holds promise for determining the locus of learners' difficulty with L2 contrasts; however, we raise several issues requiring attention to maximize its utility in investigating L2 phonolexical processing.

Keywords: second language learning, second language phonology, phonolexical representation, speech perception, Korean, Mandarin, English, lexical decision

Edited by:
Svetlana V. Cook, University of Maryland, College Park, United States

Reviewed by:
Jeong-Im Han, Konkuk University, South Korea
Hanyong Park, University of Wisconsin–Milwaukee, United States
Linda Polka, McGill University, Canada

***Correspondence:**
Shannon Barrios
s.barrios@utah.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *31 March 2021*
Accepted: *28 June 2021*
Published: *20 July 2021*

Citation:
Barrios S and 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. Front. Commun. 6:689470. doi: 10.3389/fcomm.2021.689470

INTRODUCTION

Second language (L2) learners are typically faced with the challenge of learning to perceive and produce novel phonemic contrasts, as well as to build a lexicon that effectively encodes the phonetic and phonological information associated with these contrasts. A growing body of research has highlighted the role that representation at the phonolexical level may play in the perseverance of learners' difficulty with novel phonological contrasts, independent of the contributions of perceptual and/or production difficulty alone (e.g., Pallier et al., 2001; Weber and Cutler, 2004; Sebastián-Gallés et al., 2005; Escudero et al., 2008; Hayes-Harb and Masuda, 2008; Broersma, 2012; Amengual, 2016b).

Perceptual and/or lexical encoding difficulty can lead to the activation of inappropriate candidates during spoken word recognition (often referred to as spurious lexical activation), as well as to less efficient competition among competitors (see Broersma and Cutler, 2011 for review). Some evidence of spurious lexical activation comes from auditory lexical decision (LD) tasks in which a participant is required to judge whether an auditory stimulus is a word or not. A ubiquitous finding is that nonnative listeners have difficulty rejecting nonwords derived from real words (e.g., deaf/*d[æ]f and lamp/*l[ɛ]mp) when these involve confusable L2 phonemes (e.g., Sebastián-Gallés and Baus, 2005; Sebastián-Gallés et al., 2006; Broersma and Cutler, 2011; Díaz et al., 2012; Darcy et al., 2013; Darcy and Thomas, 2019; Melnik and Peperkamp, 2019). Spurious lexical activation is also known to produce priming or facilitation effects for minimal pairs (Pallier et al., 2001), near-words (Broersma and Cutler, 2011; Broersma, 2012), phonologically-related primes (Cook and Gor, 2015), and semantic associates of phonological neighbors (Cook et al., 2016). Indeed, it remains of much interest how and to what extent L2 listeners utilize various sources of contextual information for coping with phonolexical ambiguity (Chrabaszcz and Gor, 2014; Chrabaszcz and Gor, 2017).

Existing neural evidence from ERP corroborates the behavioral findings reviewed above, with nonnative listeners failing to show typical N400 effects (larger N400 responses to nonwords than for words) for nonwords involving confusable phonemes (Sebastián-Gallés et al., 2006; White et al., 2017). Moreover, incorrect lexical decisions by nonnative speakers may not result in error-related negativity which has been observed in native speakers (Sebastián-Gallés et al., 2006). These studies have been important in documenting the difficulties that even highly proficient bilinguals experience with L2 lexical processing. However, such findings typically remain ambiguous as to the locus of the effects. Indeed, such patterns of spurious lexical activation may result from challenges at the perceptual (phonetic coding) and/or phonolexical (lexical coding) levels.

Asymmetries abound in L2 speech perception and lexical processing and have been helpful in shedding light on these issues. They are often associated with situations where an L2 contrast involves two target language phonemes that map to a single L1 category but with differing degrees of "goodness" (category goodness assimilation according to Best's Perceptual Assimilation Model; Best, 1995). The better fitting category is often referred to as the dominant category and the other as the non-dominant (unfamiliar or new) category. The latter is typically thought to be less robustly encoded than the former. Studies employing the visual world paradigm have provided evidence of perceptual representations that are neutralized in favor of the dominant category contacting differentiated phonolexical representations (e.g., Weber and Cutler, 2004; Cutler et al., 2006; Escudero et al., 2008). For example, Weber and Cutler (2004) demonstrated that Dutch-English bilinguals experienced spurious activation of English words containing underlying /ɛ/ (e.g., looks to a picture of a 'pencil') in response to auditory forms containing [æ] (e.g., "panda") but not the reverse, suggesting that these bilinguals had established differentiated lexical representations for /ɛ/ and /æ/ words, but that their ability to differentially contact these representations was undermined by neutralization of [ɛ] and [æ] to [ɛ] at the level of speech perception. Escudero et al. (2008) replicated this finding with an artificial lexicon study, demonstrating that learners infer the lexical contrast from the written forms of newly-learned words, and Cutler et al. (2006) similarly provide evidence for differentiated lexical representations for English /l/ and /ɹ/ in native Japanese speakers who perceptually neutralize the contrast. In a similar study involving native German learners of English, Llompart and Reinisch (2017) showed that the English /æ/-/ɛ/ lexical contrast can be inferred from seeing the words articulated even though they are perceptually neutralized in favor of [ɛ]. Llompart and Reinisch (2020) demonstrated that German learners of English can establish distinct lexical representations for /æ/ and /ɛ/ following exposure to minimal pairs during word learning, but that in this case, neutralization at the level of perception unexpectedly favored [æ] (rather than [ɛ]).

In other cases, studies employing auditory LD tasks have uncovered asymmetries in performance that are suggestive of the reverse scenario: differentiated perceptual representations contacting imprecise (i.e., fuzzy) lexical representations of the new category (e.g., Darcy et al., 2013; Melnik and Peperkamp, 2019). In these studies, adult learners are presented with L2 words and nonwords where the nonwords are identical to the words except that one phoneme is replaced with a confusible phoneme. In one experiment, Darcy et al. (2013) presented native English speakers at two levels of L2 German language experience German words (e.g., [honiç] "honey" containing the dominant (i.e., familiar) vowel /o/ and [køniç] "king" containing the non-dominant (i.e., new) vowel /ø/), as well as nonwords created by replacing [o] with [ø] and vice-versa (e.g., *[høniç] and *[koniç]). Darcy et al. (2013) predicted that if participants neutralized the contrast at the level of perception (while maintaining a contrast in the lexicon[1]), they would show the following ordinal accuracy pattern:

[1]Darcy et al. (2013) remain agnostic as to whether differentiated phonolexical representations are target-like or simply robust enough to distinguish the contrast (i.e., the non-dominant /ø/ category might be stored as "not /o/").

1. **Word [Dominant]:** Words containing the dominant phoneme will be easy to accept because, e.g., the input [o] is perceived as [o], which matches the underlying phonolexical representation /o/. (Input [honiç] is perceived as [honiç] which matches /honiç/).
2. **Nonword [Dominant]:** Nonwords containing the dominant phoneme will be easy to reject because, e.g., the input [o] is perceived as [o], which does not match the underlying phonolexical representation containing the new phoneme. (Input *[koniç] is perceived as [koniç] which does not match /køniç/).
3. **Word [Non-dominant]:** Words containing the non-dominant phoneme will be difficult to accept because, e.g., the input [ø] is perceived as [o], which does not match the underlying phonolexical representation containing the new phoneme. (Input [køniç] is perceived as [koniç] which does not match /køniç/).
4. **Nonword [Non-dominant]:** Nonwords containing the non-dominant phoneme will be difficult to reject because, e.g., the input [ø] is perceived as [o], which matches the underlying phonolexical representation /o/. (Input *[høniç] is perceived as [honiç] which matches /honiç/).

In this scenario, where the locus of the difficulty is at the level of perceptual coding, both words and nonwords containing the dominant category should be easy to accept and reject, respectively, because learners use accurate perceptual representations of the dominant category to contact lexical representations that encode the contrast. On the other hand, words and nonwords containing the non-dominant category should be more difficult to accept and reject, respectively, due to perceptual neutralization in favor of the dominant category. Darcy et al. (2013) further assume that it will generally be easier to accept words than to reject nonwords (thus 1>2 and 3>4 above). As a result, this scenario, which we will call the perceptual coding scenario "is not expected to yield an interaction between lexical status (word vs. non-word) and category type (old vs. new)" (pp. 379–380). They proposed a second possible scenario, where learners' perceptual coding of the input preserves the contrast, as does the lexicon; however, the phonolexical representation of the non-dominant category is imprecise, or fuzzy, such that it is activated by inputs containing either member of the contrast. Following Darcy et al. (2013) we use /?/ to indicate that a category is represented imprecisely in the phonolexical representation. According to Darcy et al. (2013), such a scenario should result in the following ordinal accuracy pattern:

1. **Word [Dominant]:** Easy to accept because, e.g., the input [o] is perceived as [o], which matches the phonolexical representation /o/. (Input [honiç] is perceived as [honiç] which matches /honiç/).
2. **Word [Non-dominant]:** Less easy to accept because, e.g., the input [ø] is perceived as [ø], and does not perfectly match the fuzzy phonolexical representation containing the new phoneme. (Input [køniç] is perceived as [køniç] which matches /k?niç/).
3. **Nonword [Non-dominant]:** Easy to reject because, e.g., the input and percept [ø] does not match the phonolexical representation /o/. (Input *[høniç] is perceived as [høniç] which does not match /honiç/).
4. **Nonword [Dominant]:** Difficult to reject because, e.g., the input and percept [o] does not mismatch the fuzzy phonolexical representation containing the new phoneme. (Input *[koniç] is perceived as [koniç] which does not mismatch /k?niç/).

In this scenario, where the learner exhibits difficulty at the level of lexical coding, an interaction is expected between lexical status (word, nonword) and segment (dominant, non-dominant), with more accurate performance on words containing the dominant category than words containing the non-dominant category, but more accurate performance on nonwords containing the non-dominant category than nonwords containing the dominant one.

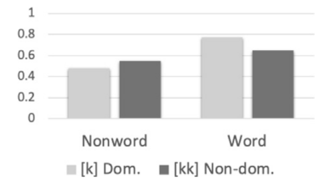
Darcy et al.'s (2013) LD results are summarized **Table 1**, together with the results of several additional studies which are reviewed below. They found that the intermediate-level L1 English learners of German exhibited the interaction of lexical status and segment associated with the lexical coding scenario, with more accurate rejection of nonwords containing [ø] than [o] but more accurate acceptance of words containing [o] than [ø]. The advanced-level learners exhibited a non-significant but descriptively similar pattern. In a separate LD experiment, L1 English learners of Japanese (also at two levels of experience) responded to words containing either singleton (e.g., /k/) or geminate (e.g., /kk/) consonants, in addition to nonwords that were created by replacing singleton consonants with geminates or vice-versa (e.g., [akeru] "to open" / *[akkeru] and [kippu] "ticket" / *[kipu]). Both groups of learners exhibited an interaction of lexical status and segment, with a descriptive pattern of more accurate rejection of nonwords containing [kk] than [k] but more accurate acceptance of words containing [k] than [kk]. Darcy et al. (2013) interpreted this response pattern as evidence that differentiated perception of the Japanese singleton-geminate contrast contacted fuzzy phonolexical representations of the non-dominant geminate consonants. As expected, the native Japanese-speaking control group exhibited high accuracy in all conditions, and no interaction of lexical status and segment. Curiously, however, the native German group exhibited a marginally significant interaction of lexical status and segment that was in the opposite direction of that exhibited by the German learners. This latter finding will be taken up below when we discuss asymmetries in native LD performance.

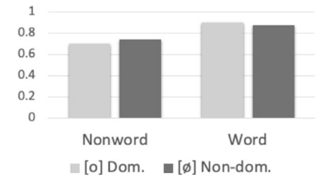
Additional scenarios beyond those presented by Darcy et al. (2013) are logically possible, depending on the degree of precision associated with lexical encoding. We spell out the full set of predictions for the English /æ/-/ε/ contrast in **Table 2**. For this purpose, we treat /ε/ as "dominant" (old/familiar) category and /æ/ as "non-dominant" (new/unfamiliar). Native speakers' phonolexical representations are typically (and often implicitly) assumed to be both distinctive and precise (referred to henceforth as "precise") in that both members of a given contrast will be encoded such that differentiated perceptual representations will clearly match (e.g., [æ] = /æ/ and [ε] = /ε/) or mismatch (e.g., [æ] ≠ /ε/ and [ε] ≠ /æ/). On the other hand, L2 phonolexical representations of non-dominant categories are sometimes characterized as fuzzy or imprecise, though these terms have been used somewhat variably in the literature.

TABLE 1 | Summary of LD findings of previous studies (vertical axis represents mean proportion correct).

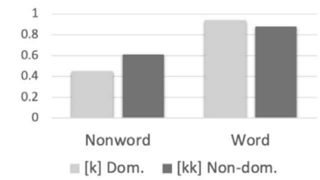
Darcy et al. (2013) Experiment 2

English (Intermediate Learners)



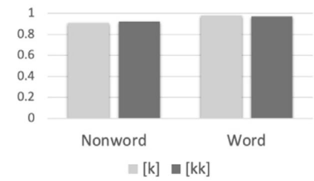


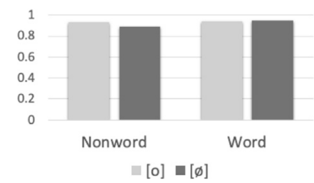
Japanese (Native)[a]



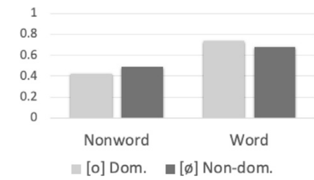
Darcy et al. (2013) Experiment 4

English (Intermediate Learners)



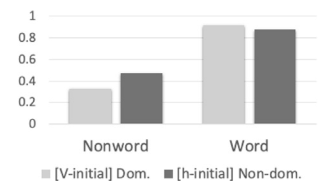


German (Native)[a]

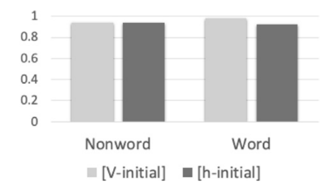


Melnik and Peperkamp (2019)

French (Learners)

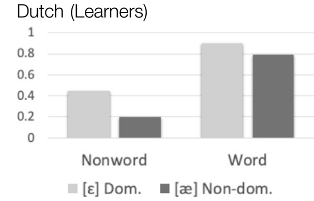


English (Native)

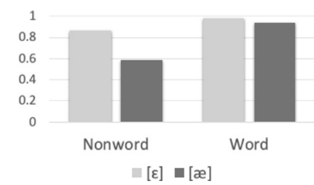


Díaz et al. (2012)

Dutch (Learners)

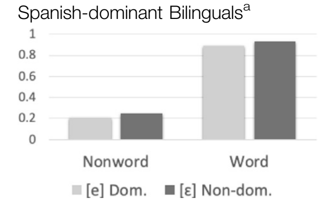


English (Native)

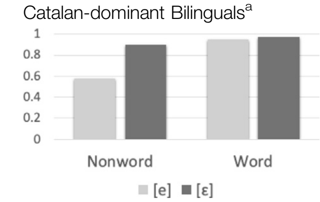


Sebastián-Gallés et al. (2005)

Spanish-dominant Bilinguals[a]

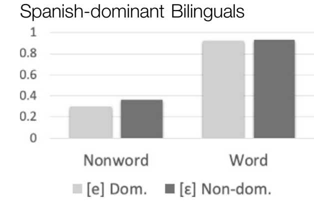


Catalan-dominant Bilinguals[a]

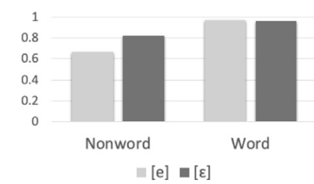


Sebastián-Gallés et al. (2006)

Spanish-dominant Bilinguals

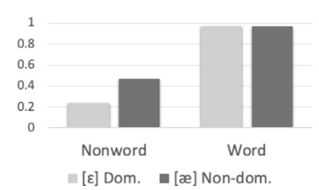


Catalan-dominant Bilinguals



Llompart and Reinisch (2019)

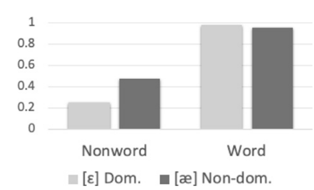
German (Learners)

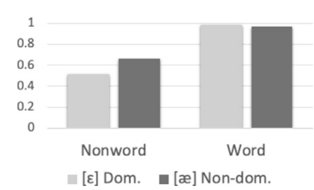


No native speaker group

Llompart (2020)

German (Intermediate Learners)





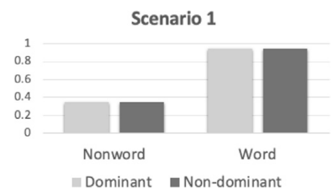
No native speaker group

[a]*Data estimated from figures in the article; otherwise, means are specified in the article or were provided by the authors (Díaz, 2021; Llompart, 2021).*

TABLE 2 | Lexical Decision Scenarios and Ordinal Accuracy Predictions. Summary of the eight scenarios created by crossing neutralized and precise perceptual representations with neutralized, ambiguous, "not X," and precise phonolexical representations. Each condition (Dom = Dominant; Non-dom = Non-dominant) is followed by the corresponding input, percept, phonolexical representation (LexRep) and ordinal accuracy prediction; 1 = highest/high accuracy, 4 = lowest/low accuracy. Each scenario is accompanied by a stylized figure illustrating the ordinal accuracy predictions (note that accuracy predictions are relative, not absolute). Relative predicted accuracy in the following scenarios is computed using the assumption that words are always easier to accept than nonwords are to reject. To create the stylized figures representing the predictions, we imposed a range of 0.75–1.0 for performance on words, and a range of 0.0–0.75 for performance on nonwords (based on trends in the literature). We assigned the following accuracy proportions for words: easy to accept = 0.95, easy-ish to accept = 0.90, difficult to accept = 0.80, and nonwords: easy to reject = 0.70, difficult-ish to reject = 0.45, difficult to reject = 0.35.

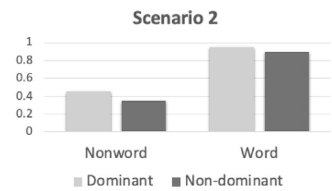
Scenario 1: Neutralized perceptual representation, neutralized phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[lɛmp]	/lɛmp/	Easy to accept (1)
Nonword (Dom)	[lɛmp]	[lɛmp]	/lɛmp/	Difficult to reject (2)
Nonword (Non-dom)	[dæsk]	[dɛsk]	/dɛsk/	Difficult to reject (2)



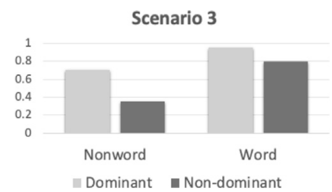
Scenario 2: Neutralized perceptual representation, ambiguous phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[lɛmp]	/l?mp/	Easy-ish to accept (2)
Nonword (Dom)	[lɛmp]	[lɛmp]	/l?mp/	Difficult-ish to reject (3)
Nonword (Non-dom)	[dæsk]	[dɛsk]	/dɛsk/	Difficult to reject (4)



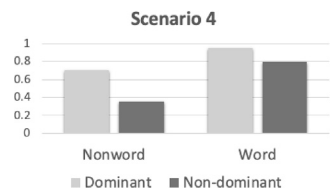
Scenario 3: Neutralized perceptual representation, "not X" phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[lɛmp]	/l{not ɛ}mp/	Difficult to accept (2)
Nonword (Dom)	[lɛmp]	[lɛmp]	/l{not ɛ}mp/	Easy to reject (3)
Nonword (Non-dom)	[dæsk]	[dɛsk]	/dɛsk/	Difficult to reject (4)



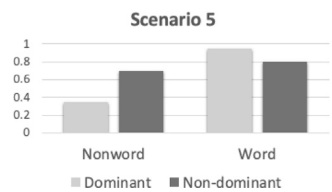
Scenario 4: Neutralized perceptual representation, precise phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[lɛmp]	/læmp/	Difficult to accept (2)
Nonword (Dom)	[lɛmp]	[lɛmp]	/læmp/	Easy to reject (3)
Nonword (Non-dom)	[dæsk]	[dɛsk]	/dɛsk/	Difficult to reject (4)



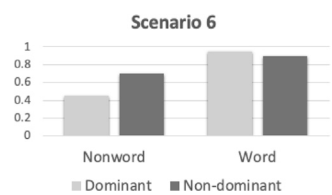
Scenario 5: Precise perceptual representation, neutralized phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[læmp]	/lɛmp/	Difficult to accept (2)
Nonword (Dom)	[lɛmp]	[lɛmp]	/lɛmp/	Difficult to reject (4)
Nonword (Non-dom)	[dæsk]	[dæsk]	/dɛsk/	Easy to reject (3)



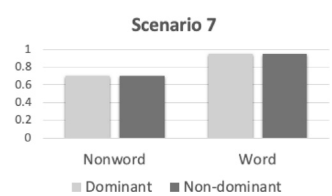
Scenario 6: Precise perceptual representation, ambiguous phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[læmp]	l?mp/	Easy-ish to accept (2)
Nonword (Dom)	[lɛmp]	[lɛmp]	/l?mp/	Difficult-ish to reject (4)
Nonword (Non-dom)	[dæsk]	[dæsk]	/dɛsk/	Easy to reject (3)



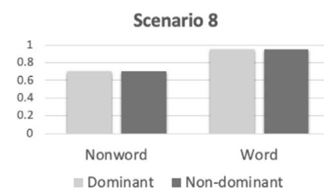
Scenario 7: Precise perceptual representation, "not X" phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[læmp]	/l{not ɛ}mp/	Easy to accept (1)
Nonword (Dom)	[lɛmp]	[lɛmp]	/l{not ɛ}mp/	Easy to reject (2)
Nonword (Non-dom)	[dæsk]	[dæsk]	/dɛsk/	Easy to reject (2)



Scenario 8: Precise perceptual representation, precise phonolexical representation

Condition	Input	Percept	LexRep	Ordinal Prediction
Word (Dom)	[dɛsk]	[dɛsk]	/dɛsk/	Easy to accept (1)
Word (Non-dom)	[læmp]	[læmp]	/læmp/	Easy to accept (1)
Nonword (Dom)	[lɛmp]	[lɛmp]	/læmp/	Easy to reject (2)
Nonword (Non-dom)	[dæsk]	[dæsk]	/dɛsk/	Easy to reject (2)



Indeed, there may be multiple types of phonolexical imprecision: the representation of non-dominant categories might be neutralized to the dominant category (e.g., /æ/ encoded as /ɛ/; “neutralized”), ambiguous (/æ/ encoded as /?/, which neither matches nor mismatches [æ] or [ɛ] (“ambiguous”), or differentiated but imprecise (e.g., /æ/ encoded as /not ɛ/; see Hayes-Harb and Masuda, 2008; “not X”). Importantly, these various types of imprecision might produce different predictions regarding LD accuracy patterns. The eight scenarios (along with ordinal accuracy predictions) that result from crossing the four types of phonolexical encoding of non-dominant categories just described with either neutralized (e.g., [æ] perceived as [ɛ]) or precise (e.g., [æ] perceived as [æ]) perceptual representations are elaborated in **Table 2**. In these scenarios, the dominant phoneme is always assumed to be perceived and phonolexically encoded in a distinctive and precise manner. Given a particular set of assumptions about how relative accuracy is computed (detailed in **Table 2**), each scenario produces a prediction regarding the ordinal accuracy associated with words/nonwords and dominant/non-dominant segments. At one extreme (scenario 1), where an individual’s perceptual and phonolexical representations are neutralized to the dominant category, the learner will exhibit a bias towards YES responses symmetrically for both words and nonwords, resulting in highly accurate performance on words and inaccurate performance on nonwords. At the other extreme (scenario 8), where both perceptual and phonolexical representations are distinctive and precise, individuals will accept words and reject nonwords accurately and symmetrically. It is interesting to note this same pattern is observed when the perceptual representations are precise but phonolexical representations are “not X” (scenario 7). Scenarios 2, 3 and 4, where perceptual representations are neutralized and phonolexical representations are ambiguous, “not X,” or precise, respectively, listeners’ performance will be asymmetric with more accurate performance on stimuli containing the dominant phoneme in both nonword and word conditions (we will refer to these as perceptual coding scenarios).[2] Scenarios 5 and 6, where perceptual representations distinguish the two phonemes and phonolexical representations are neutralized or ambiguous, respectively, listeners’ performance will be asymmetric but with an opposite directional pattern of the asymmetry for nonwords and words (these will be collectively referred to as lexical coding scenarios). It is instructive to compute this full set of scenarios, as doing so reinforces the essential difference between response patterns attributable to perceptual processing difficulty (scenarios 2, 3 and 4/ perceptual coding), and those which may be uniquely attributed to challenges at the phonolexical level (scenarios 5 and 6/lexical coding). Doing so further demonstrates that, given the present assumptions, neutralized, ambiguous, and precise phonolexical representations produce the same ordinal accuracy predictions when perceptual representations are neutralized, and when perceptual representations are distinctive and precise, both neutralized and ambiguous phonolexical representations produce the same predictions.

Equipped with these scenarios and their predictions, we now turn to several other studies that have also reported asymmetries in L2 lexical decision performance.[3] In understanding the findings of these studies, it is imperative to clarify how nonword stimuli are coded: in some studies they are coded according to the underlying form (e.g., Díaz et al., 2012; *[lɛmp] “lamp” is coded as an /æ/ nonword), while in others they are coded according to the surface form (e.g., Darcy et al., 2013; *[lɛmp] “lamp” is coded as an [ɛ] nonword). In the following discussion, for ease of interpretation and consistency and whenever possible, we present studies’ findings using the surface form coding scheme so that performance relative to the ordinal accuracy predictions can be evaluated. In addition, given differences between the goals of the studies reviewed below and the present study, and thus analyses focused on different types of patterns, we discuss the findings of these studies in terms of the descriptive patterns (where they are presented or can be inferred) with respect to the LD scenarios.

Melnik and Peperkamp (2019) provide data that is consistent with the predictions of the lexical coding scenarios (scenarios 5 and 6). They investigated the lexical processing of /h/-initial and vowel-initial English words by Intermediate-Advanced L1 French learners of English and native speakers of English (a combination of American, British, and Canadian varieties). Given that French learners of English will often

[2]It is worth noting that the ordinal accuracy shown in scenario 3 and 4 of **Table 2** differs slightly from the ordinal accuracy predictions made by Darcy et al. (2013). To our knowledge, the prediction associated with Darcy et al.’s (2013) perceptual coding pattern that nonwords containing the dominant category will be more accurate than words containing the non-dominant category is unattested; however, robust word > nonword accuracy patterns are reported across the literature (Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006; Díaz et al., 2012; Darcy et al., 2013; Llompart and Reinisch, 2019; Melnik and Peperkamp, 2019; Llompart, 2020, summarized in **Table 1**). For this reason, we make the assumption that words in both dominance conditions will elicit higher accuracy than will nonwords.

[3]A number of other studies have reported asymmetries in child learners (e.g., Simon et al., 2014), adult learners (e.g., Simonchyk and Darcy, 2017; Hayes-Harb and Barrios, 2019), and even early bilinguals (e.g., Amengual, 2016b) using auditory word picture matching tasks with familiar and newly-learned words.

produce English /h/-initial words such as "husband" without the initial /h/ and accept nonwords such as "usband" and "[h]officer" as the real words "husband" and "officer," the authors hypothesized that L1 French learners of English would exhibit a pattern of lexical decision accuracy consistent with them having what they called "fuzzy" phonolexical representations of /h/-initial English words (that is, they would exhibit more accurate performance for vowel-initial words than for [h]-initial words, but less accurate performance for vowel-initial nonwords than [h]-initial nonwords). This is indeed what they observed. Unexpectedly, they also observed a difference in word performance for the native English speakers similar to that reported for the learners, with [h]-initial words less accurate than vowel-initial words. However, no difference was observed for nonwords. Melnik and Peperkamp's (2019) LD findings are summarized in **Table 1**.

Patterns of LD asymmetries consistent with difficulty at the level of perceptual coding (scenarios 2, 3, and 4) have also been reported. Díaz et al. (2012) examined the lexical processing of the /æ/-/ɛ/ contrast by native Dutch late-learners of English (self-rated "high" proficiency) and a control group of native speakers of British English. Participants completed a lexical decision task involving monosyllabic English words containing /ɛ/ and /æ/, and an equal number of nonwords created by substituting /ɛ/ for /æ/, and vice versa (which they refer to as /ɛ/-type stimuli and /æ/-type, respectively). Analyses of A' scores revealed that native English speakers demonstrated greater sensitivity to the contrast than Dutch participants and that both groups were more sensitive to what they call /æ/-type than /ɛ/-type stimuli. However, the use of signal detection measures, such as A', makes it impossible to assess whether the interaction for surface segment by lexical status predicted the lexical coding scenarios was observed. Nonetheless, the mean proportion correct data presented in figure 3 of Díaz et al. (2012); recoded for surface as opposed to underlying segment, and summarized in **Table 1**) suggests that both groups of participants were more likely to correctly accept words containing [ɛ] "desk" than words containing [æ] "lamp" and more likely to incorrectly accept nonwords containing [æ] "d[æ]sk" than nonwords containing [ɛ] "l[ɛ]mp." While not the focus of this study, the descriptive pattern of performance of the Dutch (and the native English) speakers reported by Díaz et al. (2012) is compatible with the phonetic coding scenarios, and corroborates findings from eye tracking studies involving this contrast and learner population (discussed above).

Several studies involving Spanish-Catalan bilinguals have examined the robustness of Catalan-dominant and Spanish-dominant bilinguals' lexical encoding of the Catalan-specific /e/-/ɛ/ vowel contrast, which is known to be particularly difficult for the Spanish-dominant group (Sebastián-Gallés and Baus, 2005; Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006). These studies have employed lexical decision tasks involving words with /e/ or /ɛ/ and nonwords counterparts created by substituting the other member of the contrast (e.g., [finestrə] "window"/*[finɛstrə] and [gəʎedə] "bucket"/*[gəʎɛdə]). In these experiments participants were warned that nonwords would involve a single vowel change. The nonword acceptance for [e] and [ɛ] were high in both studies, particularly for the Spanish-dominant group. Moreover, mean proportion correct showed an asymmetrical pattern for nonwords in both groups. Sebastián-Gallés et al. (2006) report lower proportion correct for [e] than [ɛ] nonwords in addition to high accuracy on words for both groups of bilinguals. Sebastián-Gallés et al. (2005) report very similar findings in their figure 1 (see **Table 1** for a summary of Sebastián-Gallés et al. (2005) and Sebastián-Gallés et al. (2006)). In both studies, the behavioral data was coded for the underlying segment (rather than surface segment) and statistical analyses were conducted with A' scores as the dependent variable. As a result, it is not possible to know whether the asymmetries they observed resulted in a significant interaction of lexical status and surface segment. However, assuming that the nonword difference was robust, and that [e] is the dominant category and [ɛ] is the non-dominant category for Spanish-dominant Spanish-Catalan bilinguals whose native language has /e/ but lacks /ɛ/, this direction of the nonword effect (more accurate nonword performance for [ɛ] than [e]) would be more compatible with the lexical coding than perceptual coding scenarios. The authors attribute effects in Catalan-dominant bilinguals to exposure to variable input due to experience with Spanish-accented Catalan in the bilingual speech community where the research has been conducted.

Llompart (2020) studied the lexical decision behavior of two groups of L1 German learners of English (Intermediate and Advanced) on the /æ/-/ɛ/ contrast, with nonwords derived from real words by swapping the segment of interest (e.g., lemon/*l[æ]mon and dragon/*dr[ɛ]gon). Performance for word stimuli (not presented separately by surface segment in the manuscript) was at ceiling for the two groups, and analyses focused on nonword performance. As expected, both groups of learners were more accurate in their reject of nonwords for filler than for the test contrast and the advanced group outperformed the intermediate group in nonword performance for the /æ/-/ɛ/ contrast. While the study's focus was on relating categorization and vocabulary knowledge to lexical decision performance, the authors did provide data and an additional analysis of these nonword effects broken down by surface segment. We are told that "both groups showed numerically lower accuracies for items in which /æ/ was mispronounced as [ɛ] (e.g., *dr[ɛ]gon; Advanced: 51.03% correct (SD = 50.05); Intermediate: 25.44% (SD = 43.59)) than when the substitution pattern was the opposite (e.g., *l[æ]mon; Advanced: 66.15% correct (SD = 47.37); Intermediate: 47.74% (SD = 49.65))." (p. 6). Assuming that /ɛ/ is the dominant category for German learners of English (e.g., Llompart and Reinisch, 2019), this result would appear to be most compatible with the lexical coding scenarios. This pattern of performance in German learners of English is consistent with LD data presented in Llompart and Reinisch (2019), though that study reported only d' scores. The Llompart and Reinisch (2019) and Llompart (2020) LD findings are summarized in **Table 1**.

It has been generally assumed the representation of native phonemes will lack the fuzziness of novel L2 phonemes' representations, and that native speakers should therefore not exhibit the asymmetries that have been associated with L2 learners' lexical decision performance (e.g., Darcy et al., 2013).

This assumption appears to be implied in studies lacking control groups of native speakers (e.g., Hayes-Harb and Barrios, 2019; Llompart, 2020; Llompart and Reinisch, 2019; Llompart and Reinisch, 2020). However, inspection of reported response patterns reveals that native speakers quite often show asymmetric lexical decision performance (Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006; Díaz et al., 2012; Darcy et al., 2013; Melnik and Peperkamp, 2019). Recognizing the prevalence of such lexical decision asymmetries among native speakers, Melnik and Peperkamp (2019) note that it is "important to always compare the learners' performance to that of native speakers, such as to clearly identify asymmetries that are specific to L2 processing" (p. EL17).

Research employing auditory LD tasks to investigate L2 phonolexical processing has varied in the ways in which the data and patterns are presented, in addition to the ways in which various asymmetries are interpreted with respect to the unique influence of perceptual and/or phonolexical representations on L2 learners' performance. The first goal of the present research is thus to report LD data for multiple L2 contrasts across multiple L1 groups (representing previously unstudied L1-L2 combinations) and to evaluate the findings (taking into account patterns of performance on both word and nonword stimuli) with respect to the predictions of the scenarios described above. Research in this area has also varied in whether or not learner performance is compared to performance by native speakers. To the extent that asymmetries in LD performance by learners are interpreted as evidence of nonnative-like perceptual and/or phonolexical representations, they must be considered in relation to performance by native speakers (see also Melnik and Peperkamp, 2019). The second goal of this study is to document potential native speaker LD asymmetries to allow for comparison of performance by native speakers and L2 learners, and the inclusion of multiple contrasts and multiple groups of L2 learners allows more opportunities for this comparison than does a more targeted and less exploratory study. The third goal of this work is to add to the representation of individuals who have emigrated to L2-dominant settings (in this case, late learners of English in the United States) in the literature (see Extra and Verhoeven, 2011; Paradis et al., 2020). With a few exceptions (e.g., Darcy and Thomas, 2019), studies on this particular topic have focused on instructed learners in "foreign language" settings (e.g., Darcy et al., 2013; Cook and Gor, 2015) or early bilinguals in a bilingual speech community (e.g., Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006; Amengual, 2016a; Amengual, 2016b).

We selected two segmental contrasts known to be difficult for L2 learners of English from a variety of native language backgrounds: the /æ/-/ɛ/ vowel contrast and the /l/-/ɹ/ liquid contrast (e.g., Flege et al., 1997; Aoyama et al., 2004). Moreover, Cutler (2005) notes that failing to maintain these particular contrasts in English can lead to a substantial increase in lexical competition. The phonolexical processing of English /æ/ and /ɛ/ has been extensively studied in the context of Dutch and German (Weber and Cutler, 2004; Escudero et al., 2008; Díaz et al., 2012); however, this contrast is known to pose a challenge for learners of English from a wide range of language backgrounds, including native speakers of Mandarin and Korean. Native speakers of Mandarin have been shown to experience difficulty identifying, discriminating, and producing English /æ/ and /ɛ/ (Wang, 1997; Chen et al., 2001; Jia et al., 2006), neither of which is nominally present in the Mandarin vowel inventory. Native speakers of Korean also experience difficulty perceiving and producing the English /æ/-/ɛ/ contrast (Tsukada et al., 2005; Kim, 2010; Hong, 2012). Unlike Mandarin, however, the Korean vowel inventory nominally contains the vowel /ɛ/, and some evidence suggests that /ɛ/ might be expected to behave like the dominant vowel of the two. For example, Yang (1996) found that Korean /ɛ/ is acoustically more similar to English /ɛ/ than to English /æ/, and Flege et al. (1997) reported that native Korean speakers produce English /ɛ/ more accurately than English /æ/. However, other studies have found higher perception and production accuracy for English /æ/ than /ɛ/ (Tsukada et al., 2005; Cho and Jeong, 2013). With respect to predictions regarding the dominance status of the two vowels, we know of no data that points to the dominance of /ɛ/ or /æ/ for the L1 Mandarin speakers, and for L1 Korean speakers, evidence regarding dominance is contradictory.

Mandarin is characterized as having a lateral approximant phoneme /l/, but not /ɹ/ (e.g., Brown, 2000). Nonetheless, Brown (1998) demonstrated that native Mandarin speakers who were late learners of English and living in North America exhibited near-ceiling accuracy on both AX discrimination and forced-choice picture selection (e.g., hear "rake", choose between pictures of a "lake" or a "rake") tasks, with highly accurate performance maintained across onset, cluster, and coda positions. Brown (1998) notes that Mandarin /l/ does not vary allophonically between [l] and [ɹ], which may reduce the likelihood that native Mandarin speakers neutralize English /l/ and /ɹ/ to a single category. Korean has a singleton liquid phoneme--sometimes characterized as /l/ (e.g., Brown, 2000)--that is realized as [ɾ] (initially) and [l] (elsewhere). On the one hand, the status of these two phones in Korean as conditioned variants of the same phoneme might make the English /ɹ/ - /l/ contrast difficult for native speakers of Korean. Consistent with this prediction, Brown (2000) demonstrated that native Korean speakers less accurately perceived and lexically encoded the English /l/- /ɹ/ contrast than did native Mandarin speakers (who do not have the allophonic experience with [l] and [ɹ] that would encourage neutralization). However, the geminate liquid in Korean is realized as [ll] intervocalically, resulting in a [ɾ] - [ll] (singleton-geminate) contrast intervocalically, which Kim (2007) characterizes as a latent /ɹ/ - /l/ contrast. And indeed, native speakers of Korean have exhibited fairly accurate identification of English /l/ and /ɹ/, in intervocalic position (Ingram and Park, 1998; Hazan et al., 2006). Using a cross-language identification and category goodness task, Schmidt (1996) showed that native Korean speakers identify both English /l/ and /ɹ/ as Korean /l/ in initial position, but that English /l/ is more similar to Korean /l/ than is English /ɹ/ (see also Park, 2013 for discussion of the "new" versus "similar" status of English /l/ and /ɹ/ for native speakers of Korean). Concerning the dominance status of English of /l/ and /ɹ/ for these two groups of learners, the limited available evidence of similarities between English and both Mandarin and Korean /l/ suggest that English /l/ may be dominant for native Mandarin and Korean speakers.

TABLE 3 | Summary of participant characteristics.

	Mean age in years (SD; range)	Sex	Mean LOR in months (SD); range	Self-rated English Listening	(Additional) L2s	Counter-balance List
Native English (*n* = 56) 18 P, 38 SP	26.7 (10.7; 18–60)	20 male 35 female 1 non-binary	n/a	n/a	*23 reported one or more L2s* Arabic(1), ASL(1), Cantonese(1), Danish(1), Farsi(1), French(3), German (2), Italian(1), Japanese(4), Laotian(1), Mandarin (3), Portuguese(1), Spanish (14), Swedish (1), Thai (1)	28 List 1 28 List 2
Native Mandarin (*n* = 18) 6 P, 12 WoM	31.8 (8.3; 19–51)	8 male 10 female	97.4 (57.0; 36–267); missing = 2	3 fair 15 good	*8 reported additional L2s* Cantonese(1), French(2), German(2), Japanese(4), Korean (2), Thai (1)	9 List 1 9 List 2
Native Korean (*n* = 24) 4 P, 20 WoM	44.0 (9.1; 28–62)	4 male 20 female	175.3 (125.6; 6–480)	13 fair 11 good	*5 reported additional L2s* Japanese(4), Mandarin (1)	11 List 1 13 List 2

P, Prolific; WoM, Word of mouth; SP, Linguistics study pool; LOR, length of residence in the United States.

TABLE 4 | Lexical characteristics of the /æ/-/ɛ/ and /l/-/ɹ/ stimulus sets.

	Words			Nonwords	
	Mean (SD) Subtlex frequency	Mean (SD) # neighbors	Mean (SD) neighbor frequency	Mean (SD) # neighbors	Mean (SD) neighbor frequency
Underlying /æ/	25.33 (11.77)	4.75 (3.19)	5.25 (1.91)	3.67 (3.77)	21.39 (20.20)
Underlying /ɛ/	25.33 (20.89)	4.00 (4.112)	7.08 (9.45)	4.00 (4.00)	23.72 (15.97)
/æ/-/ɛ/ comparison	$t(22) = 0.000, p = 1.000$	$t(22) = 0.499, p = 0.623$	$t(22) = -0.658, p = 0.517$	$t(22) = -0.210, p = 0.836$	$t(22) = -0.313, p = 0.757$
Underlying /l/	25.83 (23.87)	3.42 (2.61)	7.08 (8.28)	2.00 (1.28)	21.64 (24.43)
Underlying /ɹ/	17.83 (8.32)	3.58 (3.12)	9.17 (10.18)	4.08 (3.99)	19.13 (17.36)
/l/-/ɹ/ comparison	$t(22) = 1.096, p = 0.285$	$t(22) = -0.142, p = 0.888$	$t(22) = -0.550, p = 0.588$	$t(22) = -1.723, p = 0.099$	$t(22) = 0.291, p = 0.774$
Filler	22.25 (8.58)	5.08 (2.84)	9.26 (8.69)	3.75 (2.53)	13.79 (6.01)

METHODS

Participants

There are three groups of participants in this approximately 15 min online study: 18 native Mandarin and 24 native Korean L2 learners of English, and a control group of 56 native speakers of English, with participants randomly assigned to two counterbalancing conditions. We pursued three avenues for participant recruitment: our department's participant pool, Prolific (www.prolific.co), and word-of-mouth. The participant pool connects students enrolled in linguistics courses to studies for course credit. Prolific connects participants to paid research studies. Due to limitations associated with Prolific's participant screening and recruitment policies, data from several participants recruited via Prolific was discarded because their responses to the post-task questionnaire indicated that they did not meet the study's inclusionary criteria. Participants recruited via Prolific were paid between $4.00 and $5.00 USD, with variation in compensation resulting from author experimentation with the system. Word-of-mouth recruitment involved emailing participants from previous studies who had opted into our recruitment list, in addition to asking colleagues to distribute a recruitment message to potential participants on our behalf. These participants were compensated with a $5 Amazon gift card.

The control group of 56 native speakers of English was recruited through the Linguistics study pool and Prolific and self-identified as native/first language speakers of English only. Data from an additional three native English speakers recruited via Prolific was excluded because they indicated that they were not familiar with some of the English words used in the study. Given the potential for systematic differences between participants recruited via the multiple avenues, we attempted to balance where participants were recruited from and counterbalanced list assignment. The native Mandarin and Korean speakers were born in China and Korea and considered Mandarin and Korean, respectively, to be their only first and native languages. They identified English a second language, and in order to ensure substantial exposure to North American dialects of English, were living in the United States at the time of the study, and had arrived in the United States no earlier than 12 years of age. Additionally, participants selected for inclusion in this study self-rated their English listening ability as "fair" or "good," on the four-point scale labeled "poor - fair - good - near-native." They exhibited a range of ages of English language acquisition, age of arrival in the United States, and length of residence in the United States (see **Table 3** for participant characteristics).

Materials

Native Mandarin and native Korean participants who were recruited by word of mouth (rather than a participant management platform) completed a pre-task questionnaire. They were asked a series of questions to confirm that they met the inclusionary criteria (native language, countries of birth and current residence, and status of English as a second language), and also took a brief Mandarin or Korean vocabulary test serving as a screening to reduce participation by individuals who are not speakers of these languages.

For the auditory lexical decision task, three sets of English word-nonword pairs were created by first querying the MRC Psycholinguistics Database (Coltheart, 1981) using the following criteria: two-syllable English words beginning with CV, a stress-unstress pattern, a subtlex frequency between 10 and 100. We then removed all proper nouns. From the resulting set, we selected words with initial /l/ or /ɹ/ (the /l/-/ɹ/ set) or with /æ/ or /ɛ/ in the first syllable (the /æ/-/ɛ/ set), and removed from these sets all words whose /l/-/ɹ/ or /æ/-/ɛ/ counterparts were real words (e.g., "/l/iver" and "/ɹ/iver"). In addition, we selected a separate set of candidate filler words that met all of the same criteria and created nonword counterparts for these words (e.g., "/k/otton"-"/p/otton"). We then assessed the neighborhood density, neighborhood frequency of all resulting words and nonwords, in addition to subtlex per million frequency of the words, and crafted the three stimulus sets, attempting to balance these lexical characteristics across the sets and with the aim of including only nouns (due to a limited number of word options, some monomorphemic verbs and -y adjectives are included in the sets). Because our stimuli met very specific criteria and we wanted to ensure there was a sufficient number of items for each category, there was overlap between some of the test and filler words (e.g., "rabbit" was used both as an [æ] and an [ɹ] word). Care was taken to ensure that multiple tokens of the same word were spoken by different talkers on a particular list. The lexical characteristics of the resulting set are presented in **Table 4**. Comparison of the lexical characteristics of /ɹ/ vs. /l/ and /æ/ vs. /ɛ/ stimuli reveal no significant differences. The complete list of stimuli provided in **Supplementary Table 1**, and the auditory lexical decision task materials are available at https://osf.io/9mnvg/.

The word and nonword stimuli were recorded by two female native speakers of American English. Three recordings of each were made, and Praat was used to identify and extract each auditory stimulus. The first production was chosen for presentation in the study unless it contained artifacts. Stimuli were scaled such that their average intensity was 65 dB. A Praat script provided formant values at the segment (vowel or liquid) midpoints (Lennes, 2003). Analysis of the acoustics of the stimuli revealed greater variability in F1/F2 for /æ/-/ɛ/ and in F3 for /l/-/ɹ/ in nonwords than words, leading to potentially confounding differences in stimulus acoustics. We thus eliminated responses to all tokens (words and nonwords) that were more than two standard deviations greater or less than the F1 or F2 means (/æ/-/ɛ/) or the F3 mean (/l/-/ɹ/) of the word tokens, separately for each talker. This resulted in the exclusion of participants' responses to six word tokens and 14 nonword tokens from the analyses (see **Supplementary Table 2**) and amounted to the exclusion of 8.3% of the data from each group (560 of 6,780 observations, 180 of 2,160 observations, and 238 of 2,880 observations, for the native English, Mandarin, and Korean speakers, respectively).

In a post-task questionnaire, participants were asked to provide basic demographic information (e.g., age, sex) in addition to information about their native and second language experience, location of birth and current residence. The native Mandarin and Korean speakers were additionally asked to detail their English language experience. They were asked to self-assess their own speaking, listening, reading, and writing ability in English on a four-point scale (poor to near-native). In addition, they were asked to indicate which of the stimulus words were unfamiliar to them by checking a box next to the word.

Procedures

All parts of the study were conducted online via computer and headphones, with participants using their own equipment in locations of their choice. Participants first were presented with the informed consent text, and registered their consent to participate by pressing a button labeled "Agree." This part of the study was conducted in Qualtrics. Participants were next directed to Pavlovia, where the listening task was hosted. The listening task involved an auditory lexical decision. Participants were told that they would hear many words, some of which were real English words and some not. Their task was to decide as quickly and accurately as possible whether the word they heard was a real English word. Participants were instructed to press the "y" key on the keyboard to indicate a YES response, and a "n" key to indicate a NO response. The task was self-paced, and a button press was required to advance. The next trial was presented 1s after the participant's response was made. The listening task consisted of a total of 120 test trials. Test trials were preceded by 4 practice trials (disk, d[u]sk, simple, s[u]mple) without feedback. The entire listening task took approximately 10 min. Upon completion of the listening task, participants were directed back to the Qualtrics platform to complete the post-task questionnaire, and then they were either directed back to the appropriate participant management platform for credit (departmental participant pool) or payment (Prolific), or asked to provide their name and email for the purpose of sending a gift card.

RESULTS

The data and analysis code are available at https://osf.io/psxw9/. We took several steps to ensure the quality of the data. In addition to excluding responses to some stimuli based on their acoustic properties (see above), we also considered participants' self-reported familiarity with the stimulus words. Because participants' familiarity with the words is crucial to our ability to interpret their lexical decision responses, we excluded responses from each participant for all words and associated nonwords that were unfamiliar for that individual. Words identified as unfamiliar by one or more Mandarin native speakers included lousy(3), medal(1), pattern(1), radar(2), ransom(7), reckon(7), rhythm(1), ruin(1), tunnel(1), and warrant(2). This amounted to the exclusion of an additional 66 observations (3% of the data). One or more Korean native speakers indicated the following words were unfamiliar: battle(1), lecture(1), lobby(1), legend(1), rabbit(1), ransom(3), reckon(10), rotten(1), rubber(1), supper(2), and

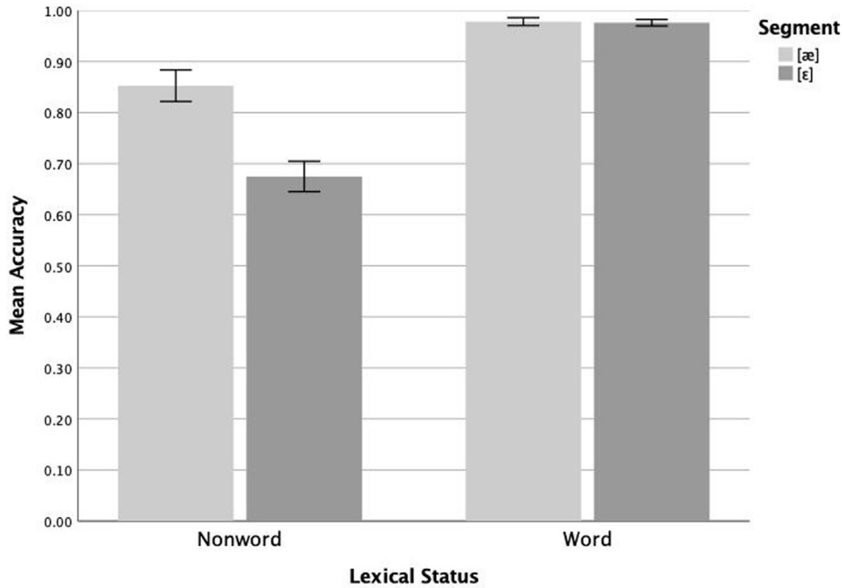


FIGURE 1 | Mean lexical decision accuracy (proportion correct) on [æ] and [ɛ] items by native English speakers (*n* = 56). Whiskers represent 1 SE.

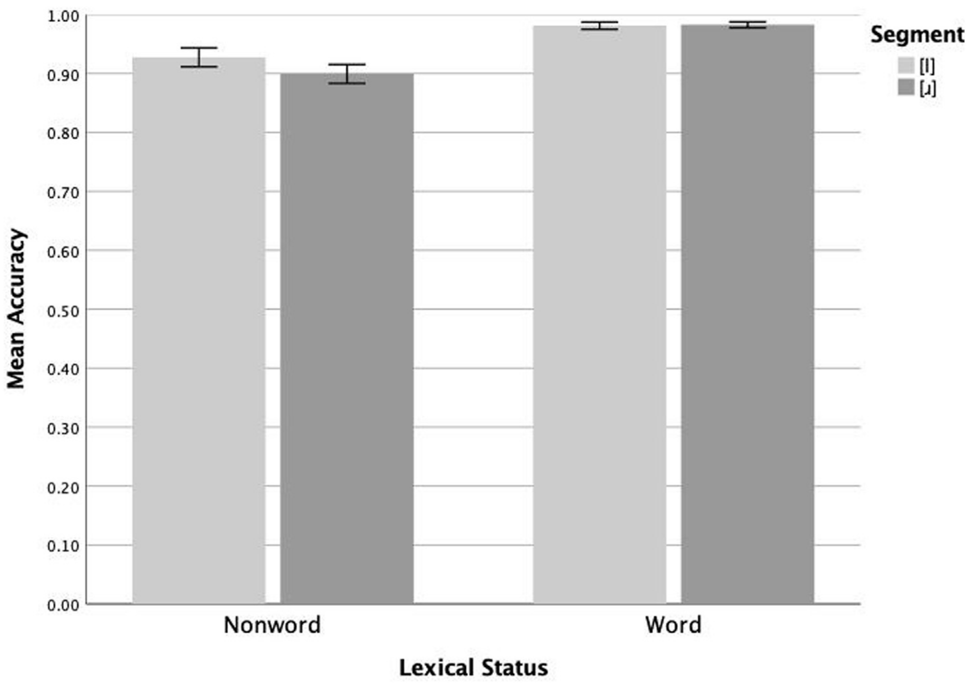


FIGURE 2 | Mean lexical decision accuracy (proportion correct) on [l] and [ɹ] items by native English speakers (*n* = 56). Whiskers represent 1 SE.

warrant(2). Exclusion of these words and their corresponding nonwords resulted in the exclusions of 78 of 2,880 observations (2.7% of the data) for the Korean native speakers.

To further control for potentially problematic stimuli, following the exclusion of tokens with outlier acoustics and unfamiliar words and corresponding nonwords just described, we also excluded tokens where the mean proportion correct performance was greater or less than 2 SD of their group mean for a given target segment ([æ], [ɛ], [l], [ɹ], or filler) and lexical status (word, nonword). This resulted in the exclusion of twelve tokens for native English speakers (336 observations), 10 tokens for native Mandarin speakers (99 observations), and 12

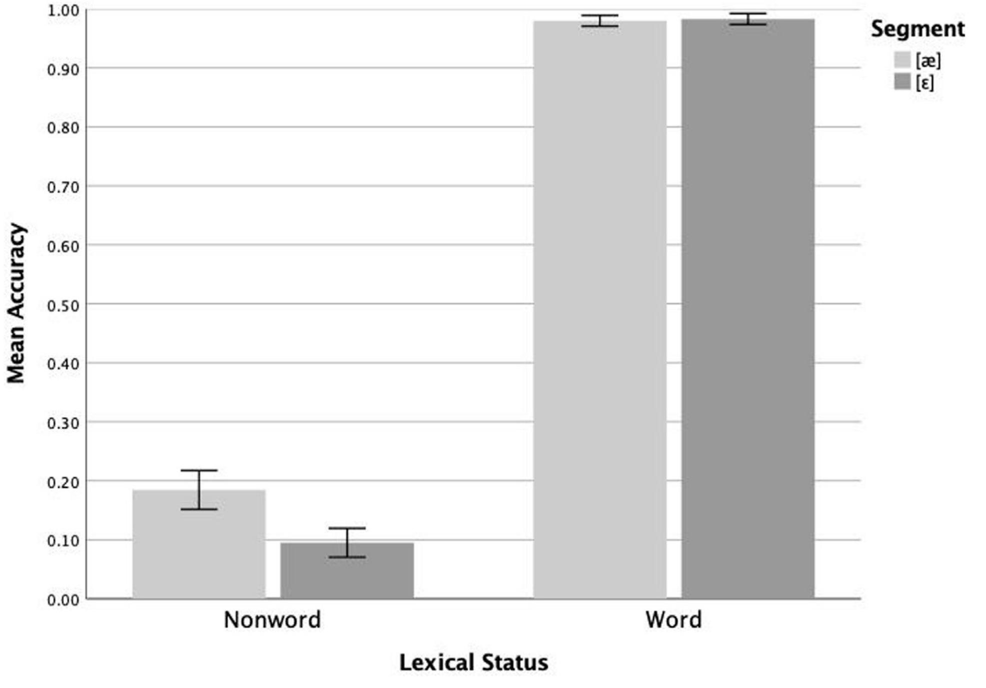


FIGURE 3 | Mean lexical decision accuracy (proportion correct) on [æ] and [ɛ] items by native Mandarin speakers (n = 18). Whiskers represent 1 SE.

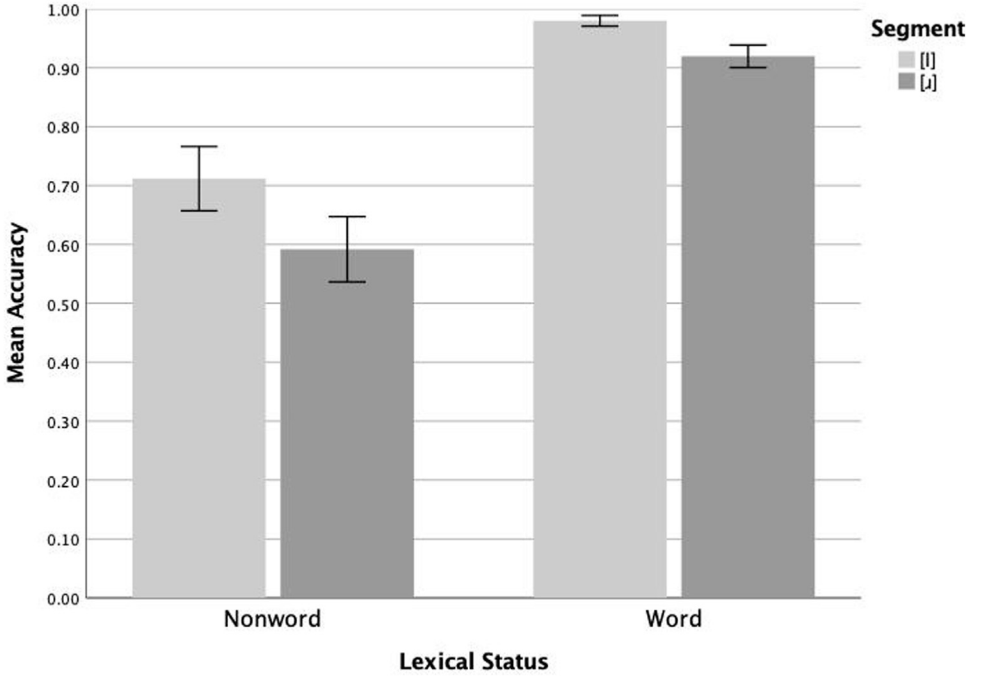


FIGURE 4 | Mean lexical accuracy (proportion correct) on [l] and [ɹ] items by native Mandarin speakers (n = 18). Whiskers represent 1 SE.

tokens for native Korean speakers (140 observations; the excluded tokens are listed in **Supplementary Table 2**). Following all exclusions, data analysis was conducted on the remaining 5,824 observations provided by 56 native speakers of English (86.7%), 1827 observations from 18 native speakers of Mandarin (84.5%), and 2,435 observations from 24 native speakers of Korean (84.5%).

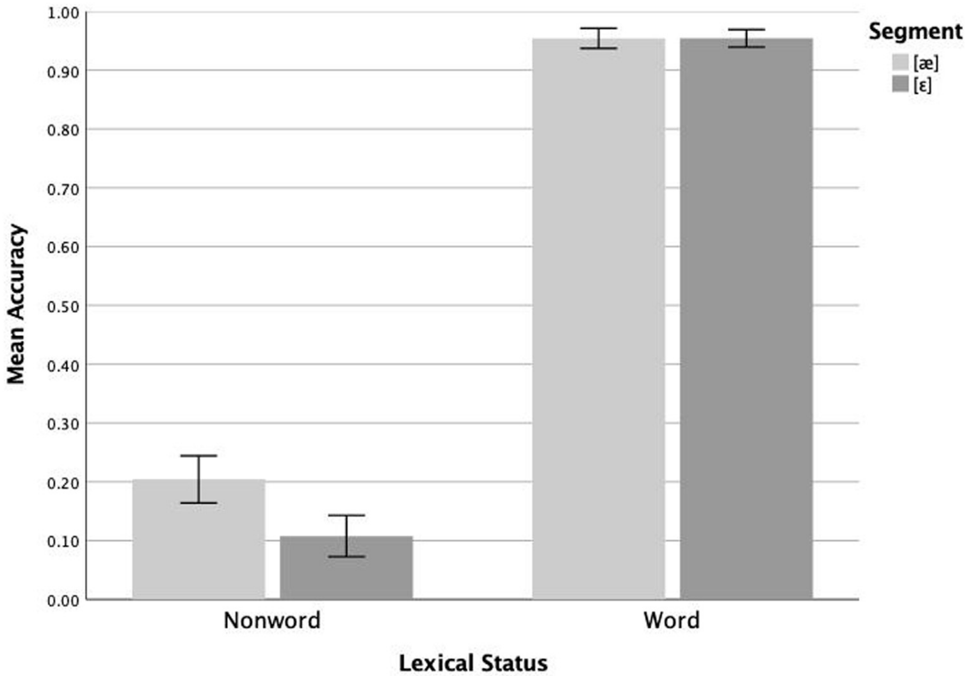


FIGURE 5 | Mean lexical decision accuracy (proportion correct) on [æ] and [ɛ] items by native Korean speakers (n = 24). Whiskers represent 1 SE.

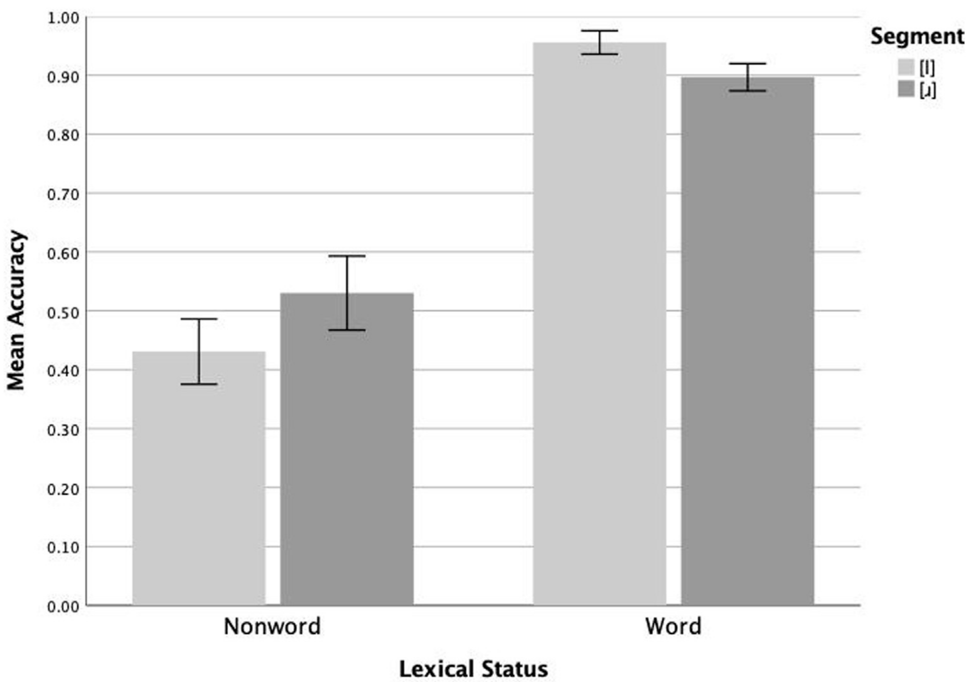


FIGURE 6 | Mean lexical decision accuracy (proportion correct) on [l] and [ɹ] items by native Korean speakers (n = 24). Whiskers represent 1 SE.

Native English Speakers

Proportion correct responses (YES to words and NO to nonwords) of the native English speakers were submitted to two separate repeated measures ANOVAs. Because filler items were designed only to distract, involved a variety of segmental contrasts, and were not associated with any predictions regarding relative difficulty, they are not considered in these analyses (mean proportion correct responses by native English speakers to filler nonwords was 0.913 and to filler words was 0.986). In the first repeated measures ANOVA, surface segment (two levels: [æ], [ɛ]) and lexical status (two levels: nonword, word) were within-subjects independent variables and proportion correct was the dependent variable. The main effect of surface segment was significant ($F(1,55) = 35.979$, $p < 0.005$; $\eta_p^2 = 0.395$), as were the main effect of lexical status ($F(1,55) = 53.466$, $p < 0.005$; $\eta_p^2 = 0.493$) and the interaction of the two ($F(1,55) = 34.399$, $p < 0.005$, $\eta_p^2 = 0.395$). Follow-up pairwise analyses revealed that native English-speaking participants responded correctly more often to [æ] (mean = 0.853) than to [ɛ] (mean = 0.675) nonwords ($F(1,55) = 39.179$, $p < .005$, $\eta_p^2 = 0.416$), but no significant difference in responses to [æ] (mean = 0.978) and [ɛ] (mean = 0.976) words ($F(1,55) = 0.047$, $p = 0.828$, $\eta_p^2 = 0.001$). These results are plotted in **Figure 1**. The pattern observed here does not readily match any of the predicted scenarios, perhaps due to the near-ceiling performance on word stimuli. However, we do observe an asymmetry in performance on nonwords, with higher accuracy for [æ] nonwords (those derived from words containing /ɛ/) than for [ɛ] nonwords (derived from words containing /æ/).

For the /l/-/ɹ/ data, a second repeated measures ANOVA with surface segment and lexical status as within-subjects variables and proportion correct as the dependent variable revealed a non-significant main effect of surface segment ($F(1,55) = 2.192$, $p = 0.144$, $\eta_p^2 = 0.038$), a significant main effect of lexical status ($F(1,55) = 24.973$, $p < 0.005$, $\eta_p^2 = 0.312$), and a non-significant interaction of the two ($F(1,55) = 2.512$, $p = 0.119$, $\eta_p^2 = 0.044$). Follow-up pairwise analyses revealed that native English-speaking participants did not differ in their accuracy for [l] (mean = 0.928) and [ɹ] (mean = 0.899) nonwords ($F(1,55) = 2.680$, $p = 0.107$, $\eta_p^2 = 0.046$), or for [l] (mean = 0.981) and [ɹ] (mean = 0.983) words ($F(1,55) = 0.051$, $p = 0.822$, $\eta_p^2 = 0.001$). These results are plotted in **Figure 2**. This response pattern, with near-ceiling performance for words and slightly lower but symmetric accuracy for nonwords, is consistent with the predictions of scenario 8, where participants benefit from perceptual and phonolexical representations that are differentiated and precise (or the functionally equivalent "not X"; scenario 7).

Native Mandarin Speakers

Mean proportion correct for native Mandarin speakers overall was 0.698. Mean proportion correct for filler words was 0.926 and filler nonwords 0.612. Analysis of the /æ/-/ɛ/ data from the native Mandarin speakers revealed significant main effects of surface segment ($F(1,17) = 6.171$, $p = 0.024$, $\eta_p^2 = 0.266$) and lexical status ($F(1,17) = 1{,}206.335$, $p < 0.005$, $\eta_p^2 = 0.986$), and a significant interaction of the two ($F(1,17) = 8.035$, $p = 0.011$, $\eta_p^2 = 0.321$). Follow-up pairwise analyses revealed that these participants responded correctly more often to [æ] (mean = 0.184) than to [ɛ] (mean = 0.095) nonwords ($F(1,17) = 7.820$, $p = 0.012$, $\eta_p^2 = 0.315$), with no significant difference in responses to [æ] (mean = 0.980) and [ɛ] (mean = 0.983) words ($F(1,17) = 0.080$, $p = 0.781$, $\eta_p^2 = 0.005$). These results are plotted in **Figure 3**. Like the native English speakers, the native Mandarin speakers exhibited highly accurate performance on [æ] and [ɛ] words. However, their relatively low accuracy on the nonword stimuli points to a YES bias in their responses. In addition, like the native English speakers, they show an asymmetry in responses to these nonwords with more accurate performance on [æ] than [ɛ] nonwords.

Analysis of the /l/-/ɹ/ data from the native Mandarin speakers revealed significant main effects of surface segment ($F(1,17) = 13.586$, $p = 0.002$, $\eta_p^2 = 0.444$) and lexical status ($F(1,17) = 37.385$, $p < 0.005$, $\eta_p^2 = 0.687$), and a non-significant interaction of the two ($F(1,17) = 1.221$, $p = 0.285$, $\eta_p^2 = 0.067$). Follow-up pairwise analyses revealed that these participants responded correctly significantly more often to [l] (mean = 0.712) than to [ɹ] (mean = 0.592) nonwords ($F(1,17) = 6.603$, $p = 0.020$, $\eta_p^2 = 0.280$), and significantly more often to [l] (mean = 0.980) than to [ɹ] (mean = 0.920) words ($F(1,17) = 7.743$, $p = 0.013$, $\eta_p^2 = 0.313$). These results are plotted in **Figure 4**. In contrast to the native English speakers, the native Mandarin speakers exhibited asymmetries for both words and nonwords with higher accuracy on [l] than [ɹ] stimuli. This pattern of responses is consistent with scenarios 2, 3 and 4, where a neutralized perceptual representation of [ɹ] as [l] contacts phonolexical representations that are ambiguous (scenario 2), 'not X' (scenario 3) or precise (scenario 4).

Native Korean Speakers

Mean proportion correct for native Korean speakers overall was 0.669. Mean proportion correct for filler words was 0.955 and filler nonwords 0.705. Analysis of the /æ/-/ɛ/ data from the native Korean speakers revealed significant main effects of surface segment ($F(1,23) = 7.182$, $p = 0.013$, $\eta_p^2 = 0.238$) and lexical status ($F(1,23) = 464.043$, $p < 0.005$, $\eta_p^2 = 0.953$), and a marginal interaction of the two ($F(1,23) = 3.949$, $p = 0.059$, $\eta_p^2 = 0.147$). Follow-up pairwise analyses reveal that these participants responded correctly more often to [æ] (mean = 0.204) than to [ɛ] (mean = 0.108) nonwords ($F(1,23) = 6.651$, $p = 0.017$, $\eta_p^2 = 0.224$), with no significant difference in responses to [æ] (mean = 0.954) and [ɛ] (mean = 0.954) words ($F(1,23)<0.005$, $p = 0.994$, $\eta_p^2<0.005$). These results are plotted in **Figure 5**. Like the native English and native Mandarin speakers, the native Korean speakers exhibited highly accurate performance on the [æ] and [ɛ] word stimuli, as well as an asymmetry in responses to these nonwords with more accurate performance on [æ] than [ɛ] nonwords. Like the Mandarin speakers, the Korean speakers exhibited a bias towards YES responses.

Analysis of the /l/-/ɹ/ data from the native Korean revealed a non-significant main effect of surface segment ($F(1,23) = 0.679$, $p = 0.418$, $\eta_p^2 = 0.029$) and a significant main effect of lexical status ($F(1,23) = 58.740$, $p < .005$, $\eta_p^2 = 0.719$), and a significant interaction of the two ($F(1,23) = 11.961$, $p = 0.002$, $\eta_p^2 = 0.342$). Follow-up pairwise analyses reveal that these participants responded correctly significantly more often to [ɹ] (mean = 0.530) than to [l] (mean = 0.431) nonwords ($F(1,21) = 5.512$, $p = 0.028$, $\eta_p^2 = 0.193$), and significantly more often to [l] (mean = 0.956) than [ɹ] (mean = 0.897) words ($F(1,23) = 7.734$, $p = 0.011$, $\eta_p^2 = 0.252$). These results are plotted in **Figure 6**. In contrast to the native English and the native Mandarin speakers, the native

Korean speakers exhibited asymmetries for both words and nonwords, but in opposite directions, with higher accuracy on [ɹ] than [l] nonwords but higher accuracy on [l] than [ɹ] words. This pattern of responses is consistent with scenarios 5 and 6, where distinctive perceptual representations of [ɹ] and [l] contact phonolexical representations containing familiar /l/ and neutralized (scenario 5) or ambiguous (scenario 6) phonolexical representations of /ɹ/.

DISCUSSION

Our first goal was to report LD data for the same two L2 contrasts across three L1 groups, taking into account patterns of performance on both word and nonword stimuli so as to evaluate the findings with respect to the predictions of the scenarios described above. The LD task revealed that native English, Mandarin, and Korean speakers all exhibit near-ceiling acceptance of English words containing [æ] and [ɛ], and are less accurate when it comes to rejecting [æ] and [ɛ] nonwords. While native English speakers' performance on nonwords was more accurate than that of the two learner groups, all three groups exhibited the same nonword asymmetry with more accurate responses to [æ] than to [ɛ] nonwords. Participants' near-ceiling performance on these word stimuli is reminiscent of the findings of a number of other studies (Sebastián-Gallés and Baus, 2005; Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006; Díaz et al., 2012; Llompart, 2020), and is often cited as the reason for reporting analyses of A' or d' scores (collapsing across words and nonwords; Sebastián-Gallés and Baus, 2005; Díaz et al., 2012; Llompart and Reinisch, 2019) or reporting analyses of nonword data only (Llompart, 2020). However, as noted earlier, in the absence of certainty regarding the dominance of the phonemes, an asymmetry for nonwords only is ambiguous with respect to whether it provides evidence for difficulties at the phonetic or phonolexical levels.

A very different picture emerged regarding the /l/-/ɹ/ contrast. Native speakers' performance was consistent with the predictions of scenario 8 (and 7), where both phonetic and phonolexical representations unambiguously encode the contrast. The native Mandarin speakers' performance matched the ordinal accuracy predictions of the perceptual coding scenarios (scenarios 2, 3, and 4), where the learner is understood to experience difficulty at the level of speech perception. In contrast, the native Korean speakers' performance was consistent with the ordinal accuracy predictions of the lexical coding scenarios (scenarios 5 and 6). We thus have evidence that the same materials can elicit three distinct patterns of performance depending on L1 background. The native speakers of English performed as is presumably expected of native speakers (symmetric and highly accurate performance in all conditions); the native Mandarin participants' performance suggested difficulty at the perceptual level with /l/ behaving as the dominant category; and the native Korean participants' performance was consistent with difficulty at the level of lexical coding, also with /l/ behaving as the dominant category. That /l/ appears to be the dominant category for the native Mandarin and Korean speakers is unsurprising given the status of /l/ in Mandarin. In the case of Korean, the combination of word and nonword asymmetries found here suggests both that /l/ is dominant for this group and that the locus of difficulty is at the lexical encoding level. A question that remains is whether the data presented here allows us to distinguish between the predictions of scenarios 2 vs. 3 and 4 or 5 vs. 6 (which produce the same ordinal accuracy predictions but differ in the magnitude of the asymmetries). The simple answer is "no", though the availability of these more nuanced predictions among the set of scenarios proposed here beg the question of how such distinctions might arise. It is possible that developmental data involving learners at different stages of L2 acquisition will provide evidence that distinguishes these scenarios (see, e.g., Darcy et al., 2013; see also Broersma and Cutler, 2011 for discussion concerning the assessment of the amount of spurious activation experienced by learners). Future work is needed to determine whether these predictions are indeed evidenced in learners.

Our second goal was to allow for the comparison of performance by native and nonnative speakers by documenting native speaker LD performance (in addition to the performance of two different nonnative speaker groups). As already noted, doing so has revealed both expected and unexpected patterns in native English speakers. With respect to the /l/-/ɹ/ contrast, native English speakers performed as expected with high accuracy for both word and nonword stimuli, suggesting precise and detailed perceptual and lexical coding of the contrast. Unexpectedly, near-ceiling effects on word stimuli coupled with high false alarm rates and asymmetric performance on nonwords were observed for the /æ/-/ɛ/ contrast in all three groups. While asymmetries in LD performance by L2 learners have been attributed to difficulties with perceptual and/or phonolexical processing, asymmetric LD performance has also been observed in native speakers (Díaz et al., 2012; Darcy et al., 2013; Melnik and Peperkamp, 2019), as well as in learners processing of words containing so-called "easy" contrasts (e.g., native German speakers exhibited higher d' scores for /ɪ/-items (winter/*w[i]nter) than /i/-items (needle/*n[ɪ]ddle) in Llompart and Reinisch, 2019). Specifically for the /æ/-/ɛ/ contrast and native English speakers, high false alarm rates for [æ] and [ɛ] nonwords have been attested (Broersma and Cutler, 2011; Díaz et al., 2012). Curiously, native English speakers' /æ/-/ɛ/ response patterns here are similar to those reported by Díaz et al. (2012), except that the American English speakers in the present study exhibit an asymmetry in favor of [æ] nonwords, and the British English speakers in their study perform more accurately on [ɛ] nonwords. Díaz et al. (2012) speculated that the differential performance for word and nonwords containing [æ] and [ɛ] in their native English speakers may be due to 1) higher frequency of /ɛ/ than /æ/ in English, and/or 2) possible idiosyncrasies in speaker's pronunciation of /æ/ and /ɛ/ (noting, however, that there were no such asymmetries for same speaker in perception-only task with different speech tokens). The explanation based on frequency is unlikely to account for the pattern we observed here, since the nonword asymmetry goes in the opposite direction; however, the question of stimulus phonetics is an important one that is expected to interact with speech perception behavior. It is worth noting that these vowel segments [æ] and [ɛ] are heavily overlapped in the acoustic space as produced by native speakers of American English (Hillenbrand et al., 1995), and the acoustic properties of the segments involved in the contrast may play a role.

It is also possible that native (and nonnative) speaker asymmetries in lexical activation can be understood in the context of asymmetries in speech perception. Peripheral vowels are known to behave as perceptual anchors and to be more readily detected than less peripheral vowels (Polka and Bohn, 2003; Polka and Bohn, 2011), and that formant proximity and stimulus prototypicality influence perceptual asymmetries (Liu et al., 2021). Having observed an asymmetry in German learners of English in the lexical processing of the /æ/-/ɛ/ contrast with more accurate performance for [æ] nonwords over [ɛ] nonwords (like the one we report here for our three listener groups), Llompart and Reinisch (2019) and Llompart (2020) suggest that the relatively more peripheral position of /æ/ than /ɛ/ make it a better perceptual anchor. Llompart and Reinisch (2019) also attribute the unexpected asymmetry observed for the "easy" /i/-/ɪ/ contrast noted above to the more peripheral nature of /i/. At this point it is unclear why the native speakers exhibited the particular asymmetry found here, though stimulus properties, including the statistical distributions of the phonemes involved, lexical properties of words containing those phonemes, and the acoustic-phonetic properties of the stimuli themselves, may all influence LD performance. Ultimately, however, the utility of LD tasks in the investigation of L2 phonolexical processing depends on an understanding of why asymmetries are sometimes also observed in native speakers, and future work must explore the reasons behind native speaker asymmetries in order to determine the appropriate interpretation of L2 learners' asymmetric patterns.

Our third goal was to increase the representation of late learners of English in the US in the body of research on L2 phonolexical processing. By recruiting via Prolific and word-of-mouth, we were able to include participants outside of the undergraduate and graduate student population that is typically represented in studies of both native speakers and language learners. While we did not systematically collect information about education level or reasons for emigrating to the United States, interaction with participants as well as their responses to open-ended questions about their language backgrounds revealed that many of the native Mandarin and Korean participants emigrated to the United States for reasons other than post-secondary education, and therefore represent a variety of English language learners in the United States. Indeed, a consequence of this recruitment strategy is that our participants' linguistic backgrounds–in particular, the circumstances under which they acquired English–varied widely, with potential impacts on their performance in the present study at individual and/or group levels, emphasizing the need for further investigation of both individual differences and learner backgrounds in the study of L2 phonolexical processing.

The LD task has been widely employed in the study of the lexical processing of language learners. However, a strong response bias often leads to high hit rates (and ceiling effects for words) and high false alarm rates for nonwords; for this reason, researchers have often chosen to report signal detection measures such as A' or d' scores (e.g., Sebastián-Gallés and Baus, 2005; Díaz et al., 2012; Llompart and Reinisch, 2019). These signal detection measures factor out response bias by simultaneously taking into account the proportion of accurate YES responses to words (hit rate) and the proportion of inaccurate YES responses to nonwords (false alarm rate) and provide a single measure of sensitivity. However, the practice of reporting only these measures is problematic, as it obscures away from the raw data, and more importantly for our purposes because it conceals possible asymmetries in lexical processing which we have argued may be helpful for understanding the locus of difficulty for learners with respect to novel phonological contrasts.

Ceiling effects for words are also problematic for understanding LD data with respect to the scenarios we have fleshed out, since asymmetries in words and their direction relative to nonword asymmetries are required to distinguish between the perceptual coding and lexical coding scenarios, in the absence of strong a priori reasons to believe that a particular category will be dominant for a learner population. Take, for example, the case in which robust asymmetries are observed for nonwords only. On the one hand, data of this sort will almost certainly result in an interaction between lexical status and surface segment, which may be indicative of lexical coding difficulty (see Darcy et al., 2013). However, nonword differences might alternatively reflect difficulty at the level of phonetic coding if ceiling effect for word stimuli (due either to overall accurate performance on the task or a strong word bias) obscure differences in performance. As a result, word data that is not at ceiling would seem to be crucial for understanding these asymmetries. Moreover, the fact that robust asymmetries have been observed for word stimuli when nonword asymmetries are absent (Melnik and Peperkamp, 2019), also provides empirical grounds for not ignoring word data. It is possible that presenting the auditory stimuli in noise,[4] or manipulating characteristics of control or filler items, would make the task more challenging or help to reduce response bias, moderating ceiling effects for words.

As highlighted above, a firm understanding of which member of a new contrast functions as the dominant category vs. the non-dominant category is crucial for being able to distinguish the perceptual coding from the lexical coding scenarios, particularly in the face of ceiling effects for words. Despite its importance, there has been little discussion of the best diagnostics for determining dominance. Some practices include inventory comparisons, consultation of existing perception data, or production acoustics. The problem is not unlike the challenge of

[4]Cutler et al. (2004) report data L1 English and L2 Dutch listeners suggesting that English phoneme identification is impacted by noise to an equal extent, suggesting that embedding words in noise may be a feasible option.

predicting phonological similarity (see, e.g., Barrios et al., 2016), and requires further exploration in this literature.

An additional limitation of this study is that we did not systematically evaluate the prosodic positions of the studies segments. The effect of position on the difficulty posed by the English /l/-/ɹ/ contrast has been documented for native speakers of both Mandarin and Korean (Ingram and Park, 1998; Hazan et al., 2006). More generally, phonological context and its effects on L2 speech perception is likely to affect the lexical encoding of both consonants and vowels, and should be investigated in future work in this area.

In conclusion, the auditory lexical decision task is attractive as a time- and cost-effective method that holds the promise of simultaneously providing information about perceptual and lexical coding. It is further appealing because it can be readily carried out online, thus reaching more than the typical convenience sample of participants. The proliferation of studies reporting LD data in recent years has resulted in inevitable variability in implementation and interpretation, and has revealed some potential pitfalls of the method with respect to clarifying the locus of difficulty in L2 lexical processing. Here we spell out the predictions of several logically possible scenarios, and provide new data for the /æ/-/ɛ/ contrast that illustrates the difficulty of interpreting asymmetries for nonwords only, in addition to the analytical challenge that arises when native speakers also exhibit LD asymmetries. The /l/-/ɹ/ contrast materials elicited evidence of three distinct scenarios, providing new data demonstrating the effects of L1 background on the perceptual and/or lexical coding difficulties experienced by language learners. We believe that these findings together highlight the need for further research that explores and addresses the promise and drawbacks of the LD task in the study of L2 phonolexical processing.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://osf.io/5z6y7/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Utah's Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

The project was initiated by RH-H, and joined by SB following a pilot phase. For the data reported here, both authors contributed to materials development. SB set up online data collection, and RH-H conducted the analyses. Both contributed to all parts of manuscript preparation.

FUNDING

This research was funded in part by a Kickstart Grant from the College of Humanities at the University of Utah awarded to RH-H and SB.

ACKNOWLEDGMENTS

We are grateful to our University of Utah colleagues in the Speech Acquisition Lab and the Cognitive Language Aficionados Research Group, as well as the editor and reviewers, for their feedback on this project. We are also grateful to the very many colleagues who supported our participant recruitment efforts, and to the study participants, for their time and attention.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcomm.2021.689470/full#supplementary-material

REFERENCES

Amengual, M. (2016a). Cross-Linguistic Influence in the Bilingual Mental Lexicon: Evidence of Cognate Effects in the Phonetic Production and Processing of a Vowel Contrast. *Front. Psychol.* 7, 617. doi:10.3389/fpsyg.2016.00617

Amengual, M. (2016b). The Perception of Language-Specific Phonetic Categories Does Not Guarantee Accurate Phonological Representations in the Lexicon of Early Bilinguals. *Appl. Psycholinguist.* 37, 1221–1251. doi:10.1017/S0142716415000557

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived Phonetic Dissimilarity and L2 Speech Learning: the Case of Japanese /r/ and English /l/ and /r/. *J. Phonet.* 32, 233–250. doi:10.1016/s0095-4470(03)00036-6

Barrios, S., Jiang, N., and Idsardi, W. J. (2016). Similarity in L2 Phonology: Evidence from L1 Spanish Late-Learners' Perception and Lexical Representation of English Vowel Contrasts. *Second Lang. Res.* 32, 367–395. doi:10.1177/0267658316630784

Best, C. T. (1995). "A Direct Realist View of Cross-Language Speech Perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Editor W. Strange (Timonium, MD: York Press), 171–204.

Broersma, M., and Cutler, A. (2011). Competition Dynamics of Second-Language Listening. *Q. J. Exp. Psychol.* 64, 74–95. doi:10.1080/17470218.2010.499174

Broersma, M. (2012). Increased Lexical Activation and Reduced Competition in Second-Language Listening. *Lang. Cogn. Process.* 27, 1205–1224. doi:10.1080/01690965.2012.660170

Brown, C. A. (1998). The Role of the L1 Grammar in the L2 Acquisition of Segmental Structure. *Second Lang. Res.* 14 (2), 136–193. doi:10.1191/026765898669508401

Brown, C. A. (2000). "The Interrelation Between Speech Perception and Phonological Acquisition from Infant to Adult," in *Second Language Acquisition and Linguistic Theory*. Editor J. Archibald (Malden, Massachussetts: Wiley-Blackwell), 4–64.

Chen, Y., Robb, M., and Gilbert, H. (2001). Vowel Production by Mandarin Speakers of English. *Clin. Linguist. Phonet.* 15, 427–440.

Cho, M.-H., and Jeong, S. (2013). Perception and Production of English Vowels by Korean Learners: A Case Study. *Stud. Phonet. Phonol. Morphol.* 19, 155–177. doi:10.17959/sppm.2013.19.1.155

Chrabaszcz, A., and Gor, K. (2014). Context Effects in the Processing of Phonolexical Ambiguity in L2. *Lang. Learn.* 64, 415–455. doi:10.1111/lang.12063

Chrabaszcz, A., and Gor, K. (2017). Quantifying Contextual Effects in Second Language Processing of Phonolexically Ambiguous and Unambiguous Words. *Appl. Psycholinguist.* 38, 909–942. doi:10.1017/s0142716416000497

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Q. J. Exp. Psychol. Section A* 33, 497–505. doi:10.1080/14640748108400805

Cook, S. V., and Gor, K. (2015). Lexical Access in L2. *Ml* 10, 247–270. doi:10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations lead to Fuzzy Form-to-Meaning Mappings. *Front. Psychol.* 7, 1345. doi:10.3389/fpsyg.2016.01345

Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric Mapping from Phonetic to Lexical Representations in Second-Language Listening. *J. Phonet.* 34, 269–284. doi:10.1016/j.wocn.2005.06.002

Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English Phoneme Confusions by Native and Non-Native Listeners. *J. Acoust. Soc. Am.* 116, 3668–3678. doi:10.1121/1.1810292

Cutler, A. (2005). "The Lexical Statistics of Word Recognition Problems Caused by L2 Phonetic Confusion," in *INTERSPEECH-2005*, 413–416.

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric Lexical Access and Fuzzy Lexical Representations in Second Language Learners. *Mental Lexicon* 8, 372–420. doi:10.1075/ml.8.3.06dar

Darcy, I., and Thomas, T. (2019). When Blue Is a Disyllabic Word: Perceptual Epenthesis in the Mental Lexicon of Second Language Learners. *Bilingualism* 22, 1141–1159. doi:10.1017/S1366728918001050

Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). Individual Differences in Late Bilinguals' L2 Phonological Processes: From Acoustic-Phonetic Analysis to Lexical Access. *Learn. Individ. Differ.* 22, 680–689. doi:10.1016/j.lindif.2012.05.005

Díaz, B. (2021). *Personal Communication with R. Hayes-Harb on Mar 23, 2021.*

Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel Second-Language Words and Asymmetric Lexical Access. *J. Phonet.* 36, 345–360. doi:10.1016/j.wocn.2007.11.002

Extra, G., and Verhoeven, L. (2011). *Bilingualism and Migration*. Berlin, NY: Walter de Gruyter. Available at: https://play.google.com/store/books/details?id=ot6vWLDxMl0C.

Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of Experience on Non-Native Speakers' Production and Perception of English Vowels. *J. Phonet.* 25, 437–470. doi:10.1006/jpho.1997.0052

Hayes-Harb, R., and Barrios, S. (2019). "Investigating the Phonological Content of Learners' "Fuzzy" Lexical Representation for New L2 Words," in Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference. Editors J. M. Levis, C. L. Nagle, and E. Todey, Ames, IA: Iowa State University, 55–69. Available at: https://apling.engl.iastate.edu/wp-content/uploads/sites/221/2019/07/10th-Annual-PSLLT-Proceedings-FINAL.pdf.

Hayes-Harb, R., and Masuda, K. (2008). Development of the Ability to Lexically Encode Novel Second Language Phonemic Contrasts. *Second Lang. Res.* 24, 5–33. doi:10.1177/0267658307082980

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., and Chung, H. (2006). The Use of Visual Cues in the Perception of Non-Native Consonant Contrasts. *J. Acoust. Soc. Am.* 119, 1740–1751. doi:10.1121/1.2166611

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic Characteristics of American English Vowels. *J. Acoust. Soc. America* 97 (5), 3099–3111. doi:10.1121/1.411872

Hong, S. (2012). The Relative Perceptual Easiness Between Perceptually Assimilated Vowels for University-Level Korean Learners of American English and Measurement Bias in an Identification Test. *Stud. Phonet. Phonol. Morphol.* 18, 491–511. doi:10.17959/sppm.2012.18.3.491

Ingram, J. C. L., and Park, S.-G. (1998). Language, Context, and Speaker Effects in the Identification and Discrimination of English /r/ and /l/ by Japanese and Korean Listeners. *J. Acoust. Soc. Am.* 103, 1161–1174. doi:10.1121/1.421225

Jia, G., Strange, W., Wu, Y., Collado, J., and Guan, Q. (2006). Perception and Production of English Vowels by Mandarin Speakers: Age-Related Differences Vary with Amount of L2 Exposure. *J. Acoust. Soc. Am.* 119, 1118–1130. doi:10.1121/1.2151806

Kim, J.-E. (2010). Perception and Production of English Front Vowels by Korean Speakers. *Phonetics Speech Sci.* 2, 51–58. Available at: http://www.koreascience.or.kr/article/JAKO201019455946033.page.

Kim, J. G. (2007). "Evidence of /l/-/r/ Contrast in Korean," in ICPhS 2007 Proceedings. Editors J. Trouvain and W. J. Barry (Saarbrücken, Germany: University des Saarlandes), 649–652.

Lennes, M. (2003). Analyze Formant Values from Labeled Segments in Files [Praat Script]. Available at: http://phonetics.linguistics.ucla.edu/facilities/acoustic/collect_formant_data_from_files.txt (Accessed February 20, 2020).

Liu, Y. Y., Polka, L., Masapollo, M., and Ménard, L. (2021). Disentangling the Roles of Formant Proximity and Stimulus Prototypicality in Adult Vowel Perception. *JASA Express Lett.* 1 (1), 015201. doi:10.1121/10.0003041

Llompart, M. (2021). *Personal communication with R. Hayes-Harb and S. Barrios on Mar 26, 2021.*

Llompart, M. (2020). Phonetic Categorization Ability and Vocabulary Size Contribute to the Encoding of Difficult Second-Language Phonological Contrasts into the Lexicon. *Bilingualism* 24, 481–496. doi:10.1017/S1366728920000656

Llompart, M., and Reinisch, E. (2017). Articulatory Information Helps Encode Lexical Contrasts in a Second Language. *J. Exp. Psychol. Hum. Percept. Perform.* 43, 1040–1056. doi:10.1037/xhp0000383

Llompart, M., and Reinisch, E. (2019). Robustness of Phonolexical Representations Relates to Phonetic Flexibility for Difficult Second Language Sound Contrasts. *Bilingualism* 22, 1085–1100. doi:10.1017/S1366728918000925

Llompart, M., and Reinisch, E. (2020). The Phonological Form of Lexical Items Modulates the Encoding of Challenging Second-Language Sound Contrasts. *J. Exp. Psychol. Learn. Mem. Cogn.* 46, 1590–1610. doi:10.1037/xlm0000832

Melnik, G. A., and Peperkamp, S. (2019). Perceptual Deletion and Asymmetric Lexical Access in Second Language Learners. *J. Acoust. Soc. America* 145, EL13–EL18. doi:10.1121/1.5085648

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The Influence of Native-Language Phonology on Lexical Access: Exemplar-Based versus Abstract Lexical Entries. *Psychol. Sci.* 12, 445–449. doi:10.1111/1467-9280.00383

Park, H. (2013). Detecting Foreign Accent in Monosyllables: The Role of L1 Phonotactics. *J. Phonet.* 41 (2), 78–87. doi:10.1016/j.wocn.2012.11.001

J. Paradis, X. Chen, H. Ramos, and M. Crago (Editors) (2020). The Language, Literacy and Social Integration of Refugee Children and Youth [Special Issue]. *Appl. Psycholinguist* 41 (6), 1251–1254. doi:10.1017/S0142716420000788

Polka, L., and Bohn, O.-S. (2003). Asymmetries in Vowel Perception. *Speech Commun.* 41, 221–231. doi:10.1016/S0167-6393(02)00105-X

Polka, L., and Bohn, O.-S. (2011). Natural Referent Vowel (NRV) Framework: An Emerging View of Early Phonetic Development. *J. Phonet.* 39, 467–478. doi:10.1016/j.wocn.2010.08.007

Schmidt, A. M. (1996). Cross-Language Identification of Consonants. Part 1. Korean Perception of English. *J. Acoust. Soc. Am.* 99 (5), 3201–3211. doi:10.1121/1.414804

Sebastián-Gallés, N., and Baus, C. (2005). "On the Relationship between Perception and Production in L2 Categories," in *Twenty-first Century Psycholinguistics: Four Cornerstones*. Editor A. Cutler (Mahwah, New Jersey: Erlbaum), 279–292.

Sebastián-Gallés, N., Echeverría, S., and Bosch, L. (2005). The Influence of Initial Exposure on Lexical Representation: Comparing Early and Simultaneous Bilinguals. *J. Mem. Lang.* 52, 240–255. doi:10.1016/j.jml.2004.11.001

Sebastian-Gallés, N., Rodríguez-Fornells, A., De Diego-Balaguer, R., and Díaz, B. (2006). First- and Second-Language Phonological Representations in the Mental Lexicon. *J. Cogn. Neurosci.* 18, 1277–1291. doi:10.1162/jocn.2006.18.8.1277

Simon, E., Sjerps, M. J., and Fikkert, P. (2014). Phonological Representations in Children's Native and Non-Native Lexicon. *Bilingualism* 17, 3–21. doi:10.1017/S1366728912000764

Simonchyk, A., and Darcy, I. (2017). "Lexical Encoding and Perception of Palatalized Consonants in L2 Russian," in Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference. Editors M. G. O'Brien and J. M. Levis (Ames, IA: Iowa State University), 121–132.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., and Flege, J. (2005). A Developmental Study of English Vowel Production and Perception by Native

Korean Adults and Children. *J. Phonet.* 33, 263–290. doi:10.1016/j.wocn.2004.10.002

Wang, X. (1997). The Acquisition of English Vowels by Mandarin ESL Learners: A Study of Production and Perception. Available at: http://summit.sfu.ca/system/files/iritems1/7404/b18794348.pdf (Accessed March 19, 2021).

Weber, A., and Cutler, A. (2004). Lexical Competition in Non-Native Spoken-word Recognition. *J. Mem. Lang.* 50, 1–25. doi:10.1016/S0749-596X(03)00105-0

White, E. J., Titone, D., Genesee, F., and Steinhauer, K. (2017). Phonological Processing in Late Second Language Learners: The Effects of Proficiency and Task. *Bilingualism* 20, 162–183. doi:10.1017/S1366728915000620

Yang, B. (1996). A Comparative Study of American English and Korean Vowels Produced by Male and Female Speakers. *J. Phonet.* 24, 245–261. doi:10.1006/jpho.1996.0013

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 July 2021
doi: 10.3389/fpsyg.2021.688356

Check for updates

Vocabulary Size Is a Key Factor in Predicting Second Language Lexical Encoding Accuracy

Danielle Daidone[1]* and Isabelle Darcy[2]

[1] Department of World Languages and Cultures, University of North Carolina Wilmington, Wilmington, NC, United States,
[2] Department of Second Language Studies, Indiana University, Bloomington, IN, United States

This study investigates the relationship between the accuracy of second language lexical representations and perception, phonological short-term memory, inhibitory control, attention control, and second language vocabulary size. English-speaking learners of Spanish were tested on their lexical encoding of the Spanish /ɾ-r/, /ɾ-d/, /r-d/, and /f-p/ contrasts through a lexical decision task. Perception ability was measured with an oddity task, phonological short-term memory with a serial non-word recognition task, attention control with a flanker task, inhibitory control with a retrieval-induced inhibition task, and vocabulary size with the X_Lex vocabulary test. Results revealed that differences in perception performance, inhibitory control, and attention control were not related to differences in lexical encoding accuracy. Phonological short-term memory was a significant factor, but only for the /r-ɾ/ contrast. This suggests that when representations contain sounds that are differentiated along a dimension not used in the native language, learners with higher phonological short-term memory have an advantage because they are better able to hold the relevant phonetic details in memory long enough to be transferred to long-term representations. Second language vocabulary size predicted lexical encoding across three of the four contrasts, such that a larger vocabulary predicted greater accuracy. This is likely because the acquisition of more phonologically similar words forces learners' phonological systems to create more detailed representations in order for such words to be differentiated. Overall, this study suggests that vocabulary size in the second language is the most important factor in the accuracy of lexical representations.

Keywords: lexical encoding, vocabulary size, phonological short-term memory, inhibitory control, attention control, L2 perception, L2 Spanish

OPEN ACCESS

Edited by:
Kira Gor,
University of Maryland, College Park,
United States

Reviewed by:
Miquel Llompart,
University of Erlangen-Nuremberg,
Germany
Gabriela Simon-Cereijido,
California State University, Los Angeles, United States

****Correspondence:***
Danielle Daidone
daidoned@uncw.edu

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *30 March 2021*
Accepted: *15 June 2021*
Published: *22 July 2021*

Citation:
Daidone D and Darcy I (2021) Vocabulary Size Is a Key Factor in Predicting Second Language Lexical Encoding Accuracy. Front. Psychol. 12:688356. doi: 10.3389/fpsyg.2021.688356

INTRODUCTION

Models of second language (L2) speech perception have typically focused on the effect of the first language (L1) at the level of phonetic or phonological categories (e.g., Flege, 1995; Best and Tyler, 2007), with the implicit assumption in the field being that the accuracy of category perception directly translates to the accuracy of these sounds in the lexicon, that is, of lexical representations. However, recent empirical studies have found variation in the relationship between accuracy in perception and lexical representations (Elvin, 2016; Simonchyk and Darcy, 2017; Llompart, 2021b), while others have found that even accurate perception is not a guarantee of accurate lexical encoding (e.g., Darcy et al., 2013; Daidone and Darcy, 2014; Amengual, 2016). Thus, the relationship between perception ability and lexical encoding is not straightforward. This suggests that the phonological forms of words in the L2 mental lexicon (i.e., L2 phonolexical forms) may generally be less detailed, or "fuzzy" (Hayes-Harb and Masuda, 2008; Cook, 2012; Darcy et al., 2013), above and beyond what would be expected from perception ability alone. Consequently, there must be other factors at play that influence learners' ability to encode the sounds of L2 words in long term memory. Identifying these factors is important theoretically, as examining such individual differences can give a window into the mechanisms necessary for the establishment of lexical representations. Additionally, identifying these factors is the first step in determining how to aid learners in acquiring more accurate L2 lexical representations.

It is likely that variability in lexical encoding accuracy may be due to learners' differing abilities to select the relevant information in the signal, hold sounds in memory, or reduce the influence of their L1 phonological grammar during word learning. Previous studies have shown that phonological short-term memory (e.g., Aliaga-García et al., 2011), inhibitory control (e.g., Lev-Ari and Peperkamp, 2013, 2014; Darcy et al., 2016), attention control (e.g., Darcy et al., 2014), and L2 vocabulary size (e.g., Bundgaard-Nielsen et al., 2011, 2012) are all possibly involved in enhancing the processing of L2 sounds or modulating cross-linguistic phonological influence on perception or production. However, the link between learners' phonological short-term memory, inhibitory control, attention control, and L2 vocabulary size and the accuracy of their lexical representations has largely been unexplored.

Background

For native speakers, the representations of words accurately reflect the sound system of the language being processed, and the process of selecting appropriate representations in the lexicon is efficient and largely error-free. For L2 learners, however, this is not necessarily the case, as they may have difficulty in both the accurate storage and processing of L2 words. While this is often attributable to difficulty with the perception of novel L2 contrasts, the relationship between accuracy in perception and accuracy in lexical encoding has been shown to vary by proficiency level and the language pairing or contrasts under investigation (Elvin, 2016; Simonchyk and Darcy, 2017; Llompart, 2021b). For example, Simonchyk and Darcy (2017) examined the relationship between perception and lexical encoding for plain versus palatalized consonants for English-speaking learners of Russian at different levels of proficiency. They found that there was no relationship between intermediate learners' error rates in an ABX perception task and their error rates in an auditory word-picture matching task. In contrast, for advanced learners, higher ABX error rates were positively correlated with higher errors rates in the auditory word-picture matching task. In other words, those learners with better perception were also more accurate at lexical encoding, but only if they were at an advanced proficiency level. Llompart (2021b) reported the opposite result, finding that differences in perception ability were only significant predictors of lexical encoding of words with /ɛ/ and /æ/ for his intermediate German-speaking learners of English, not his advanced learners. Thus, a learner's perception ability alone is not sufficient to predict the accuracy of their lexical representations.

Additionally, even learners with accurate perception experience difficulties with L2 lexical encoding (Sebastián-Gallés and Baus, 2005; Díaz et al., 2012; Darcy et al., 2013; Amengual, 2016). In Darcy et al. (2013), ABX tasks determined that English-speaking learners of German were able to discriminate front and back rounded vowels, and English-speaking learners of Japanese were able to discriminate singleton and geminate consonants. Nevertheless, in a lexical decision task, intermediate learners in both groups and advanced Japanese learners had trouble rejecting non-words if the real word contained a new L2 category; for example, they accepted **kipu* /kipɯ/ as a word when the real word is *kippu* /kippɯ/ 'ticket'. Even highly proficient early bilinguals have been found to exhibit this tendency to perform less well on lexical tasks than would be expected from their accuracy on perceptual tasks. Amengual (2016) reported that Spanish-Catalan bilinguals had high accuracy on forced-choice identification and AX discrimination tasks, but had difficulty rejecting non-words with the incorrect vowel from the /e-ɛ/ contrast. Another study on Spanish-Catalan bilinguals by Sebastián-Gallés and Baus (2005) had participants complete a categorical perception task, which looked at their perceptual boundary for /e-ɛ/; a gating task, which examined how much of the word was necessary to be heard for it to be correctly chosen; and a lexical decision task, which looked at whether participants could correctly reject non-words with /e/ and /ɛ/ switched. They found that while 68.3% of the participants scored within the native Catalan range for the perception task, only 46.6% did so for the gating task. A mere 18.3% had native-like performance on the lexical decision task. These results show that exhibiting a native-like perceptual boundary between these two vowels in isolation did not entail that they were represented correctly in words. Díaz et al. (2012) found similar results when testing Dutch-speaking late learners of English. While almost half the L1 Dutch participants in their study scored within the native range for a categorization task testing the English /æ-ɛ/ contrast, only a few scored within the native range for tasks tapping lexical knowledge, suggesting that for most participants their lexical representations containing /æ/ or /ɛ/ were not as accurate as their perception of those vowels.

While it is always possible that the discrimination tasks researchers have used were not sensitive enough to expose learners' continued difficulties with novel L2 sounds, even L2 words that do not contain confusable phonemes have been shown to be less effectively recognized (Cook, 2012; Cook and Gor, 2015; Cook et al., 2016). Cook et al. (2016) administered a translation judgment task to English-speaking learners of Russian in which participants heard a word such as /malatok/ 'hammer' followed by the English translation (< HAMMER >) presented visually. Participants then decided whether the English word was the correct translation of the Russian word. In some cases, the auditory stimulus was not the translation of the following English word, but rather a phonologically similar word, such as /malako/

'milk.' Importantly, these words did not differ based on contrasts that were difficult for L2 learners. They found that unlike native speakers, learners were willing to accept phonologically similar words as a match to the translation, and the more similar the words were to the correct translation, the more likely they were to accept them.

It is clear that the L1 phonological system affects learners' ability to accurately encode L2 words. However, given that even accurate perception does not always lead to accurate lexical encoding, the nature of L2 phonolexical representations cannot be explained solely by interference from the L1 phonological system. Therefore, other factors must also be playing a role in the accuracy of these representations. We investigate the following four factors in addition to perception: phonological short-term memory, inhibitory control, attention control, and vocabulary size in the second language. Reasons that make these factors good candidates to impact the accuracy of L2 lexical representations are outlined below.

Phonological Short-Term Memory

One cognitive ability that may be related to learners' individual differences in L2 phonolexical representations is phonological short-term memory (PSTM), which is the phonological loop component of working memory. The phonological loop allows for the storage and manipulate of auditory information; it is capable of maintaining auditory memory traces for up to a few seconds before they decay, unless they are renewed by sub-vocal articulatory rehearsal (Baddeley, 2003).

Researchers that have examined the relationship between individual differences in PSTM and perception have reported that learners with higher PSTM generally have more accurate perception of vowels and consonants (MacKay et al., 2001; Aliaga-García et al., 2011; Lengeris and Nicolaidis, 2014; Cerviño-Povedano and Mora, 2015; Darcy et al., 2015). For example, Lengeris and Nicolaidis (2014) found that Greek learners of English with higher PSTM were more accurate at identifying English consonants, in noise and in quiet. Aliaga-García et al. (2011) reported that bilingual Catalan-Spanish learners in the high PSTM group had more accurate perception of synthesized English vowel stimuli than did the learners in the low PSTM group, although Safronova and Mora (2012) did not reproduce this finding. The results of these studies suggest that higher PSTM may help learners develop more target-like cue-weighting and therefore more native-like perception, as suggested by Cerviño-Povedano and Mora (2015). They reported that Spanish-speaking learners of English with higher PSTM were less likely to over-rely on duration as a cue to the English /i-ɪ/ contrast.

PSTM has also been shown to be related to accuracy and gains over time in L2 production (O'Brien et al., 2007; Nagle, 2013; Mora and Darcy, 2016). For example, Nagle (2013) examined the relationship between the pronunciation ratings given to English-speaking learners of Spanish and their PSTM. He found a moderate positive correlation between the two, such that higher PSTM was related to higher pronunciation ratings from native Spanish speakers. A positive relationship between PSTM and pronunciation accuracy was also evidenced by Mora and Darcy (2016) for Spanish-speaking learners of English.

Overall, the majority of studies have shown that higher PSTM is related to more accurate L2 perception and production, accounting for a small but significant portion of the variance or evidencing at least a moderate correlation. Researchers hypothesize that this is because learners who have a greater ability to encode and maintain detailed and accurate short-term representations of sounds subsequently transfer these more target-like representations to long-term memory and into lexical representations. In turn, the enhanced development of new L2 phonetic categories stems from these more accurate long-term representations of words (Speciale et al., 2004; Nagle, 2013). While this proposed connection between PSTM and lexical encoding has not previously been examined empirically, this hypothesis suggests that there should be a positive relationship between variance in PSTM and the accuracy of L2 phonolexical encoding in the current study.

Inhibitory Control

Another factor that may affect lexical encoding is inhibitory control. In general, inhibitory control is a type of executive function that allows an individual to suppress a dominant internal response or override the pull of an external stimulus and instead respond in a more appropriate manner (Diamond, 2013). Various taxonomies of inhibition, interference control, or executive functions more broadly have been proposed, with a lack of general agreement between studies on the use of terms (Nigg, 2000; Friedman and Miyake, 2004; Miyake and Friedman, 2012). For the present study, the most relevant type of inhibition is that referred to by Friedman and Miyake (2004) as *Resistance to Distractor Interference*, or "the ability to resist or resolve interference from information in the external environment that is irrelevant to the task at hand" (p. 104). Although not necessarily termed as such within the studies themselves, a body of work has found that the results of tasks testing resistance to distractor interference is related to the amount of interference between bilinguals' L1 and L2 phonology in production and perception.

Using a retrieval-induced inhibition task, Lev-Ari and Peperkamp (2013) investigated the relationship between inhibitory control and L2 influence on the L1 phonology. They found that English-French bilinguals with lower inhibitory skill produced the voiceless stops /p t k/ with shorter, more French-like VOT values when speaking English. Those with lower inhibitory skill also categorized more tokens along a continuum between *dean* and *teen* as beginning with the voiceless /t/, suggesting that they had a more French-like VOT boundary. Thus, those with lower inhibitory skill exhibited more influence from their L2 phonology in their L1. Darcy and colleagues have used a retrieval-induced inhibition task based on the one used by Lev-Ari and Peperkamp (2013) to investigate the relationship between inhibitory control and L2 phonological accuracy. In their study on the L2 phonology of English-speaking learners of Spanish and Spanish-speaking learners of English, Darcy et al. (2016) found that learners with higher inhibitory skill were more accurate at perceiving L2 vowels and more accurate at producing L2 consonants. However, Mora and Darcy (2016) found no relationship between inhibitory control and L2 pronunciation accuracy for learners of English who were L1 Spanish speakers or L1 Spanish-L1 Catalan bilinguals. In a similar study, Darcy and Mora (2016) did find that stronger inhibitory control was related to more accurate perception by L1 Spanish learners of English, although not if they were L1 Spanish-L1 Catalan bilinguals. Ghaffarvand Mokari and Werner (2019) also tested inhibitory skill with a version of the retrieval-induced inhibition task. They reported a positive relationship with inhibitory control and perception for the acquisition of British English vowels by Azerbaijani learners. Inhibitory control was significantly correlated with gain scores, such that those with higher inhibitory skill developed more accurate L2 vowel perception.

Although the connection between inhibitory control and lexical encoding has not previously been investigated, higher inhibitory skill has often been found to be related to less L1-L2 interference in perception and production, and thus it is probable that higher inhibitory skill also is related to less L1-L2 interference in encoding phonolexical representations. Therefore, stronger inhibitory control is hypothesized to correspond to higher accuracy of L2 phonolexical representations in the present study.

Attention Control

Attention is an important component in speech learning, since the ability to attend to pertinent information in the speech signal allows an individual to better notice relevant acoustic properties and create new phonetic categories (Francis et al., 2000; Guion and Pederson, 2007). Results of research on the relationship between learners' attention control and L2 phonological accuracy have been mixed, indicating a positive relationship, a negative relationship, or no relationship between the two (e.g., Kim and Hazan, 2010; Darcy et al., 2014, 2015; Gökgöz-Kurt, 2016; Mora and Darcy, 2016; Safronova, 2016). Gökgöz-Kurt (2016) found a positive relationship between both attention switching and selective attention tasks and gain scores on a test of word-boundary palatalization in English after training. Darcy et al. (2014) tested L1 English-L2 Spanish and L1 Spanish-L2 English bilinguals' L2 phonological accuracy and their attention-switching ability. The researchers reported that attention control was related to perception and production accuracy, but only for the L1 Spanish-L2 English learners, in that greater attention control was related to more accurate perception. Surprisingly, while greater attention control was also related to higher accuracy for consonants in production, greater accuracy for vowels in production was related to *less* efficient attention control. Safronova (2016) also reported mixed results for the relationship between results on L2 phonological tasks and attention switching for Catalan-Spanish bilinguals. She found that more efficient attention control was related to more perceived distance between L1 and L2 vowels. In contrast, attention control error rate was related to higher accuracy in discrimination, but in the opposite direction as expected. Those learners with a higher error rate in classifying the stimuli according to the correct dimension were those that were more accurate in discrimination. Darcy et al. (2015) found no association between the attention switching scores of Korean learners of English and their performance on a range of L2 phonological tasks. Similarly, Ghaffarvand Mokari and Werner (2019) found no association between attention control, as measured with a Stroop task, and Azerbaijani learners' improvement on L2 English vowels from high variability phonetic training.

In sum, any relationship between attention control and L2 phonological accuracy is still unclear. The conceptualization of attention control varies greatly in the literature and different tasks are used to test this concept, making it even more difficult to draw definitive conclusions. As for the relationship between attention control and lexical encoding accuracy, it is logical to think that more efficient attention control, operationalized as selective attention or attention switching, would correspond to more accurate lexical representations, since the ability to focus attention on only relevant acoustic cues and efficiently switch attention between those dimensions that matter for L1 sounds versus L2 sounds could aid in acquisition. Nevertheless, this is still an open question that lacks clearly supported predictions based on the mixed results in the aforementioned literature. For the current study, it is tentatively hypothesized that greater selective attention control will correspond to more accurate L2 phonolexical representations.

Second Language Vocabulary Size

Another individual difference that may play a role in the development of L2 phonolexical representations is L2 vocabulary size. Research on child language acquisition has found that the development of a vocabulary triggers phonological development. Young children initially store words as more holistic phonological units, but as they add more vocabulary, this leads to more sensitivity to phonological differences between words. In turn, their phonolexical representations are refined in line with their increased phonological awareness (e.g., Metsala and Walley, 1998; Vihman and Croft, 2007). A similar phenomenon has been proposed for L2 learning, in that the creation of an L2 vocabulary is hypothesized to encourage the development of the L2 phonological system. The establishment of increasingly well-defined phonetic categories is in turn thought to feed back into more accurate phonolexical representations (Walley, 2007; Majerus et al., 2008; Bundgaard-Nielsen et al., 2011, 2012).

Several studies to date have examined the effect of vocabulary size on the accuracy of L2 perception and production. Darcy et al. (2015) tested the L1 and L2 productive vocabulary size of Korean learners of English. They found no significant correlations between L1 or L2 vocabulary size and a range of L2 phonological measures. Bundgaard-Nielsen et al. (2011) tested Japanese learners of English studying in Australia on their perceptual assimilation and discrimination of a range of English vowels. The learners did not differ in their years of English study, their length of stay in Australia, the age at which they began learning English, or the age at which they started their immersion experience, but they did differ in vocabulary size. The researchers found that the high vocabulary group consistently had more accurate discrimination of English vowel contrasts than the low vocabulary group. Bundgaard-Nielsen et al. (2012) reported a parallel result for the production of English vowels by Japanese-speaking learners. The vowels produced by the learners in the high vocabulary group as compared to the low vocabulary group were more accurately identified as the intended target by listeners, and vocabulary size as a continuous measure was a significant predictor of average intelligibility, unlike years of English study or length of stay in Australia. Similarly, Mairano and Santiago (2020) reported that vocabulary size correlated moderately with fluency measures and ratings of accentedness. In addition, one study has examined the relationship between L2 vocabulary size and lexical encoding. Llompart (2021b) found that a larger L2 vocabulary was predictive of more accurate phonolexical representations for German learners of English, but only if they were at an advanced level of proficiency.

Overall, these studies suggest that a larger L2 vocabulary leads to a more robust L2 phonological system and more accurate phonolexical representations. Thus, in this study a larger L2 receptive vocabulary size is expected to correspond to higher accuracy in L2 phonolexical encoding.

The Current Study

To sum up, individual differences in cognitive abilities and L2 vocabulary knowledge are all likely play a role in the processing and storage of L2 sounds, beyond learners' accuracy in perception. Greater PSTM may entail holding more detailed representations of L2 sounds in working memory, leading to the creation of more robust long-term representations. Increased inhibitory control may aid in suppressing the L1 phonological system during L2 processing, and stronger attention control may help learners focus attention on L2-relevant dimensions of the speech signal. Finally, a larger L2 vocabulary size may highlight the importance of L2 contrasts through the noticing of continual mismatches with phonological neighbors, leading to the refinement of existing phonolexical representations. Accordingly, the aim of the current study is to determine how well perception, PSTM, inhibitory control, attention control, and L2 vocabulary size each account for L2 lexical encoding accuracy. To investigate this question, our test case is the Spanish /ɾ-r/ ("/tap-trill/"), /ɾ-d/ ("/tap-d/"), /r-d/ ("/trill-d/"), and /f-p/ contrasts. These contrasts have been found to range in discriminability and lexical encoding accuracy for English-speaking learners.

First of all, /tap-trill/ has been found to be accurate in perception but not in lexical encoding. Rose (2010) reported that learners at all proficiency levels were highly accurate at distinguishing the tap and trill in an AXB task, and even naïve English listeners who knew no Spanish were able to discriminate the two phonemes at 80% accuracy. Likewise, Detrixhe (2015) found that intermediate learners were almost at ceiling on a discrimination task and an identification task before going abroad, and Herd (2011) reported that intermediate learners were already quite good at an identification task before training, at 81% accuracy, and improved to 89% accuracy after training. Daidone and Darcy (2014) also found that learners were generally able to perceive the /tap-trill/ distinction in an ABX task; in fact, advanced learners' accuracy did not significantly differ from that of native speakers. However, Rose (2012) found that both Spanish tap and trill are perceptually assimilated largely to English /ɹ/, which may help explain why Daidone and Darcy (2014) found learners' lexical encoding accuracy to be low. Learners accepted non-words with the incorrect rhotic in over 70% of cases, such such as accepting **quierro* [ki̯ero] as a word, when the real word contains a tap, i.e., *quiero* /ki̯eɾo/ 'I want.'

Regarding the /tap-d/ distinction, Rose (2010) found that this contrast was significantly less accurate than /tap-trill/ in perception for learners at all levels, ranging from an accuracy of 69.6% for second-semester students to 82.5% for graduate students. Daidone and Darcy (2014) similarly reported that /tap-d/ was less accurate than /tap-trill/, at 64% accuracy for intermediate learners and 82% for advanced learners. The intermediate learners tested by Herd (2011) also struggled to correctly identify tap and /d/ tokens and actually became less accurate after training, going from 70% to 66% accuracy on the identification task, making /tap-d/ the least accurate contrast of the three. Despite the low accuracy of /tap-d/ in perception, Daidone and Darcy (2014) found that it was more accurate in lexical encoding than /tap-trill/. While both intermediate and advanced learners were able to correctly accept tap and /d/ words with an accuracy rate above 90%, they accepted non-words with the incorrect sound at a rate of 65% for the intermediate group and 54% for the advanced group.

The /trill-d/ contrast has been found to be fairly accurate in both perceptual and lexical tasks. This was the most accurate contrast compared to /tap-trill/ and /tap-d/ in the perception results of Daidone and Darcy (2014), with an accuracy rate of 87% for intermediate learners and 94% for advanced learners. Herd (2011) also found that intermediate learners were significantly most accurate at identifying /trill-d/ than /tap-trill/ and /tap-d/, with an accuracy rate of 96% before training and 97% after training. The /trill-d/ contrast was also the most accurate of the three contrasts in lexical encoding in the results of Daidone and Darcy (2014), with word acceptance rates above 80% and non-word erroneous acceptance rates of 39 and 25% for intermediate and advanced learners, respectively.

The /f-p/ contrast served as a control in Daidone and Darcy (2014). Since this contrast also exists in English, it is unsurprising that /f-p/ was significantly more accurate in perception than /tap-trill/, /tap-d/, and /trill-d/ combined for the intermediate learners, and as accurate as these test contrasts for the advanced learners. In lexical encoding, non-word accuracy for /f-p/ was higher than for the test contrasts combined for both groups.

In sum, previous research has found that /trill-d/ is the most accurate in perception, followed by /tap-trill/ and /tap-d/, respectively. The /trill-d/ contrast has been shown to be the most accurate in lexical encoding as well; however, unlike in perception, /tap-d/ has been shown to be more accurate than /tap-trill/. The control contrast /f-p/ has been found to be accurate in both perception and lexical encoding. Given the range in accuracy of discrimination and lexical representations for these contrasts, and the varying relationship between these two constructs, Spanish /tap-trill/, /tap-d/, /trill-d/, and /f-p/ were judged to be a good test case for the relationship between lexical encoding and individual differences in perception, cognitive abilities, and vocabulary size.

MATERIALS AND METHODS

This study used a lexical decision task to investigate lexical encoding accuracy, an oddity task to examine perception of the contrasts appearing in the lexical task, a serial non-word recognition task to investigate PSTM, a retrieval-induced inhibition task to measure inhibitory control, a flanker task to investigate attention control, and an X_Lex vocabulary test to estimate Spanish vocabulary size, all described in detail below.

Lexical Decision Task

A standard auditory lexical decision task was used to provide information on the accuracy of participants' phonolexical representations. If representations are accurate, learners should accept real words and reject non-words with an incorrect sound. This task has previously been used to examine L2 lexical encoding (e.g., Sebastián-Gallés and Baus, 2005; Darcy et al., 2013).

The lexical decision task used in this study was the same task as employed by Daidone and Darcy (2014). In this task, participants heard a stimulus and indicated whether or not what they heard was a real word of Spanish. Non-words were created by substituting the target phoneme with the other sound in the contrast. For example, the non-word *quierro* /ki̯ero/ was created from the real word *quiero* /ki̯eɾo/ 'I want' by substituting the tap for a trill (see **Table 1** for more examples). The test contrasts were /tap-trill/, /tap-d/, and /trill-d/; these contrasts were chosen because they were expected to display a range of discriminability and lexical encoding accuracy. In addition, /f-p/ was the control contrast. This contrast was included because an /f-p/ contrast also exists in English, and thus should be relatively easy for learners to discriminate and encode lexically. Furthermore, /f/ and /p/ are similar in place of articulation but differ in manner of articulation, which parallels the test contrasts in that all are similar in place of articulation but differ in manner. **Table 1** provides two example words and their non-word counterparts for each condition. The full list of words used in the lexical decision task is available in **Supplementary Material**.

In order to find lexical items for the task that would be familiar to learners, an effort was made to choose as many words as possible from the *Beginning Spanish Lexicon*, a database of words from beginner Spanish textbooks (Vitevitch et al., 2012). However, because additional words were needed that contained the target sounds, the L2 Spanish learners who participated in the experiment by Daidone and Darcy (2014) also filled out a word familiarity questionnaire containing all the words from the test and control conditions to gauge their knowledge of the stimuli. This questionnaire revealed that participants in that study were generally very familiar with the words; all contrast conditions averaged 6.3 or above on a 7-point scale (range = 6.32–6.87), with 1 indicating no knowledge of the word and 7 indicating the word was very well known. Words ranged between 2 and 4 syllables, with the target phoneme appearing in intervocalic position as the onset of the 2nd, 3rd, or 4th syllable. All of the stimuli were recorded in a sound booth by two native Spanish speakers: (1) a female speaker from Puerto Rico and (2) a male speaker from Costa Rica. The speakers produced the stimuli with a standard Spanish pronunciation, such that all taps were realized with one occlusion, all trills were realized with at least two occlusions, and /d/ was realized as an approximant [ð̞].

During each trial, a fixation cross appeared in the center of the screen, and participants had 4000 ms to respond from the beginning of the stimulus. The intertrial interval (ITI) was 1000 ms. Different versions of the task were created for right- and left-handed individuals so that a response indicating 'real word' always corresponded to a key press with the participant's dominant hand. Furthermore, two different lists were created so that a word and its non-word equivalent were never heard by the same participant. For example, because the word *quiero* appeared in List 1, the non-word *quierro* appeared in List 2. This resulted in 5 words and 5 non-words for each of the 8 contrasts (see **Table 1**) in each list, totaling 80 trials. Stimuli were evenly divided between the two speakers for each contrast, and stimuli from the same speaker was used for both the word and its non-word counterpart across lists, e.g., both *quiero* and *quierro* were spoken by the female Puerto Rican speaker. In addition to the test and control stimuli, the same 24 filler words and 24 filler non-words were also included in each list, bringing the total number of trials to 128. The task began with 10 practice trials, during which reminders of what keys to press appeared on the screen (e.g., L = Real, A = Fake), and participants were given feedback on their answers (correct, incorrect, or too slow). Participants needed to score at least 7 out of 10 to precede; otherwise, they repeated the practice trials. This task was administered through a web browser with jsPsych (de Leeuw, 2015) and took participants approximately 6 min to complete.

Oddity Task

An oddity task containing the contrasts from the lexical tasks was constructed in order to investigate the ease of discriminability of these sounds. This task was chosen instead of other common perception tasks, such as AX or ABX, because it is a cognitively more demanding task (Strange and Shafer, 2008), and therefore

TABLE 1 | Example stimuli from lexical decision task.

Condition	Contrast	Stimuli Examples			
		Word		Non-word	
		Orthography	IPA	Orthography	IPA
/tap-trill/	/r-*ɾ/	aburrido 'bored'	/a.bu.'ri.do/	aburido	/a.bu.'ɾi.do/
	/ɾ-*r/	dinero 'money'	/di.'ne.ɾo/	dinerro	/di.'ne.ro/
/tap-d/	/ɾ-*d/	cultura 'culture'	/kul.'tu.ɾa/	cultuda	/kul.'tu.da/
	/d-*ɾ/	miedo 'fear'	/'mi̯e.do/	miero	/'mi̯e.ɾo/
/trill-d/	/r-*d/	ocurre 'it occurs'	/o.'ku.re/	ocude	/o.'k.ude/
	/d-*r/	estado 'state'	/es.'ta.do/	estarro	/es.'ta.ro/
/f-p/	/f-*p/	jefe 'boss'	/'xe.fe/	jepe	/'xe.pe/
	/p-*f/	grupo 'group'	/'gɾu.po/	grufo	/'gɾu.fo/

*indicates the sound in the nonword.

was less likely to result in ceiling effects for the easier contrasts. In addition, because the chance level is lower in an oddity task (25%) compared to an AX or ABX task (50%), it was expected to yield more variation in scores.

In this task, participants heard three stimuli in a row and were instructed to choose which of the three was different, or alternately, that they were all the same. For example, if they heard *lefo-lepo-lefo*, the participant was expected to indicate that the second stimulus was different. The conditions were the same as those appearing in the lexical task, that is, /tap-trill/, /tap-d/, /trill-d/, and /f-p/. Filler trials that represented other contrasts were also included. All stimuli were disyllabic Spanish non-words. Stimuli were also non-words in English. Three non-words pairs per contrast were created with the target consonants always appearing as the onset of the second syllable, such as *terro-tedo* /tero/-/tedo/. The full list of stimuli is available in **Supplementary Material**. Stimuli were recorded by a female simultaneous Spanish-English bilingual who spoke Mexican Spanish, a male Costa Rican Spanish speaker, and a female Puerto Rican Spanish speaker. The Costa Rican speaker and the Puerto Rican speaker were the same speakers that were recorded for the lexical task. Three different Spanish speakers were recorded because using different voices reduces participants' reliance on purely episodic memory to complete the task (Ramus et al., 2010); instead, participants must categorize the sounds at a phonological level to compare across speakers. Only tokens with a standard Spanish pronunciation were selected for the task; for example, all examples of the trill had at least two clear occlusions.

For every trial, each token was spoken by a different speaker, always in the same order: (1) the female simultaneous Spanish-English bilingual who spoke Mexican Spanish, (2) the male Costa Rican Spanish speaker, (3) the female Puerto Rican Spanish speaker. Participants indicated their response by clicking on one of three robots in a row on the screen according to which one "said" something different, or by clicking on the X following the robots to indicate that all the words were the same.

Each of the stimuli pairs appeared once in the 8 possible combinations of orders (AAA, BBB, ABB, BAA, ABA, BAB, AAB, BBA). For example, the *nera-nerra* stimuli pair appeared once as *nera-nera-nera* (AAA), once as *nerra-nerra-nerra* (BBB), once in the order *nera-nerra-nerra* (ABB), etc. This resulted in 24 trials per contrast and 96 test and control trials total. In addition, 48 filler trials were included, bringing the total number of trials to 144. These filler pairs also all appeared in the 8 possible combination of orders. The interstimulus interval (ISI) in each trial was 400 ms, the ITI was 500 ms, and the timeout for the trials was set to 6500 ms from the start of the trial. Participants also completed 8 training trials in order to familiarize them with the task. Participants needed to correctly respond to at least 6 out of 8 of the practice trials to precede to the actual task, or else they repeated the practice trials. The task lasted approximately 10 min, with one break in the middle, and was administered through a web browser with jsPsych. Each block contained an equal number of trials per condition, and trials were randomized within each block.

Phonological Short-Term Memory Task

A serial non-word recognition task adapted from the one used in Zahler and Lord (in press) was employed to examine PSTM. Following Cerviño-Povedano and Mora (2015), a non-word recognition task was chosen over a non-word repetition task because the latter involves production of the stimuli, and participants' ability to articulate the Russian sounds would likely have differed. Furthermore, serial recognition is less affected by the lexical status of the stimuli than serial recall, which suggests that a recognition task is a better indicator of short-term memory ability rather than knowledge of representations stored in long-term memory (O'Brien et al., 2007).

In this task, participants heard sequences of Russian stimuli and had to decide if the two sequences were in the same order or a different order. The task became progressively harder as the two sequences that participants needed to compare became longer, starting at four stimuli in a row for each sequence and ending at seven stimuli in a row. The stimuli were CVC sequences spoken by a female native speaker of Russian (see **Supplementary Material**). Although some of the Russian stimuli were real words in Russian, all of the stimuli in this task will be referred to as non-words because they were all unknown from the participants' point of view.

Stimuli were organized into sequences. Non-words within a sequence were separated by 300 ms pauses, and the two sequences in a trial were separated by a 2000 ms pause. For the different-order trials, two stimuli in the middle of the sequence were always switched (e.g., ABCDE vs. ACBDE; ABCDE vs. ABDCE), while the first and last stimulus were always in the same position. No minimal pairs were used within a sequence and adjacent stimuli did not share any phonemes. After both sequences had finished playing, participants were shown a screen reminding them of the key presses for 'same' and 'different' and given 3000 ms to respond. The ITI was 1000 ms. Participants completed 8 trials for each of the sequence lengths (4, 5, 6, and 7 non-words), for 32 trials in total. Trials were blocked by sequence length, starting with sequences of 4 and ending with sequences of 7. Before beginning the actual task, participants had to correctly respond to 3 out of the 4 practice trials with a sequence length of 4 non-words; the practice repeated as necessary. The PSTM task was administered using jsPsych and took 7 min to complete.

Inhibitory Control Task

The task employed to investigate inhibition was a retrieval-induced inhibition task like the one used in Lev-Ari and Peperkamp (2013) and Darcy et al. (2016). This task was chosen to investigate inhibitory control because other tasks often used to measure inhibition, such as the Stroop task, can also be considered measures of selective attention to external stimuli, and a separate task was used in the current study for that measure.

This task consisted of three phases: memorization, practice, and test. Participants first were instructed to memorize the 18 words. The words were individually presented on the screen with their category (e.g., "FRUITS – apple") for 5 s. In the practice phase, participants practiced half of the words from two of the categories, each three times. The categories and words that were practiced were randomized across participants. In order to practice the words, participants were presented with a category and the first letter of a word (e.g., "FRUITS-a") with a blank textbox below. They then needed to type the relevant word into the textbox. In the test phase, participants were presented with a word (e.g., "apple") and had to indicate whether each word shown on the screen was a word that they have learned in the memorization phase. Each trial was preceded by a fixation cross in the center of the screen for 1500 ms, and once the word appeared participants had 3000 ms to respond.

Stimuli were 6 words in each of 3 categories – fruits, occupations, and animals – for a total of 18 words (see **Supplementary Material**). The words were assigned into three possible conditions: practiced, inhibited, and control. Practiced items were memorized and then practiced by the participant. Inhibited items were memorized as well, but they were not practiced by the participant. However, they belonged to the same semantic category as other words that were practiced. Control items were memorized by the participant but were not subsequently practiced by them, and none of the words in that specific category were practiced. For example, if *fruits* was the control category for a participant, they would then memorize and practice half the words from each of the *occupations* and *animals* categories.

By having participants practice only some of the words that they memorized, this task led participants to inhibit the other learned items from those categories, because retrieving words from a semantic category necessitates the suppression of other words in that category. For example, if a participant memorized "nurse" and "dentist" but then only practiced "nurse", the word "dentist" should be inhibited and thus take more time to retrieve and respond to. In contrast, a word in the *animals* category like "wolf" should not have been inhibited and therefore be faster to respond to than "dentist", while "nurse" should elicit an even faster RT since it was practiced and therefore more strongly activated.

All of the 18 words they had initially memorized were included in the test phase, as well as 18 distractor words from the same semantic categories, resulting in an equal number of 'yes' and 'no' correct answers. Two versions of the task were created so that a 'yes' response corresponded to a key press with the participant's dominant hand for both right- and left-handed individuals. This 6-min task was administered through a web browser with jsPsych.

Attention Control Task

A flanker task, a non-verbal test of selective attention, was used to investigate attention control (Eriksen, 1995). The choice to use a non-verbal task rather than a speech-based attention-switching task was made in order to ensure as much as possible that the attention control task was testing a different construct than the verbal retrieval-induced inhibition task.

In this task, participants decided which way the center arrow was facing out of a group of five arrows. In congruent trials, all arrows faced the same direction (e.g., →→→→→), while in incongruent trials the middle arrow faced the opposite duration of the flanking arrows (e.g., →→←→→). Participants' ability to select relevant information (the center arrow) and ignore distracting information (the flanking arrows) tested their spatial selective attention ability, which is operationalized as the difference between reaction times to congruent and incongruent trials (Bugg and Crump, 2012). This is also known as the conflict effect or executive control (Fan et al., 2002). The smaller the difference in reaction times to congruent and incongruent trials, the better able the participant is to focus their attention on the relevant dimension.

Each trial was preceded by a fixation cross in the middle of the screen for 400 ms, after which time the arrows appeared. Participants pressed the right arrow key to indicate a right-facing arrow in the center, and the left arrow key to indicate a left-facing arrow in the center. They had 1700 ms to respond, after which point there was a 400 ms pause before the next trial. Participants first completed a training phase with feedback. In the following test phase, the 4 possible types of trials (right-facing congruent, right-facing incongruent, left-facing congruent, and left-facing incongruent) were each repeated 20 times, for a total of 80 trials. The flanker task was run through a web browser using jsPsych, and it lasted approximately 3 min.

Spanish Vocabulary Test

The X_Lex vocabulary test was used to estimate participants' receptive Spanish vocabulary size (Meara, 2005). This task was chosen because it tests words in the 0–5,000 frequency range, and it was anticipated that targeting this frequency range would capture variation in learners' knowledge without producing floor effects. In this task, participants were presented with a randomized sampling of 100 Spanish words which were evenly distributed among the 1K, 2K, 3K, 4K, and 5K frequency bands. The test also included 20 plausible Spanish non-words to correct for any bias toward answering yes to unknown words. Participants indicated whether or not they knew a word shown on the screen by clicking on the happy face for 'yes' and the sad face for 'no'. The vocabulary task took around 5 min for participants to complete.

Participants

Participants in this study were English-speaking learners of Spanish[1]. These learners were either undergraduate Spanish majors and minors enrolled in a fifth-semester or higher-level Spanish course or graduate students that had taken graduate courses in Spanish. Most of the graduate students were teaching Spanish and studying Hispanic linguistics or Hispanic literatures and cultures. They had all grown up in monolingual households in which only English was spoken.

In total, 42 L2 learners of Spanish were tested. However, three participants were excluded from all analyses for various reasons (see Daidone, 2020). This resulted in a final count of 39 L2 learners for inclusion in the analyses. The demographic info for all remaining participants is available in **Table 2**. Participants were also excluded on a task-by-task basis when necessary. These exclusions are discussed under the analysis and results section for each task.

[1] A small number of native Spanish speakers were also tested in order to ensure that the tasks were working as expected. See Daidone (2020) for details.

TABLE 2 | Demographic information for participants.

	L1 English-L2 Spanish Learners *N* = 39
Age at testing (years)	22.4 (3.8)
Age of onset for L2 learning	13.1 (2.5)
Residence in a Spanish-speaking country (months)	2.5 (6.2)
Self-rated L2 speaking ability (0–6)	3.9 (1.7)
Self-rated L2 listening ability (0–6)	4.2 (1.5)
Self-rated L2 reading ability (0–6)	4.5 (1.3)
Self-rated L2 writing ability (0–6)	4.4 (1.5)
Gender	27 female
Handedness	3 left-handed

Means are given for rows 1–8, with standard deviations in parentheses. Counts are given for rows 9 and 10.

General Procedure

After viewing the study information sheet and consenting to take part in the study, participants completed a bilateral hearing screening. All participants needed to pass the hearing screening in order for their data to be included in the analyses. Participants next completed the lexical decision task, oddity task, and a forced choice lexical decision task that is not discussed in the current study (see Daidone, 2020, for more details). They then moved onto the serial non-word recognition task, flanker task, retrieval-induced inhibition task, and X_Lex vocabulary test. Lastly, they completed a language background questionnaire, which also included a word familiarity section for the words used in the lexical decision task. For the tasks that presented auditory stimuli, participants wore Sennheiser HD 515 over-ear headphones. The entire experiment lasted 65–75 min and individuals were paid $15 for participating.

RESULTS

Results and Analyses by Task

Lexical Decision Task Analysis and Results

The lexical decision task directly assessed the accuracy of participants' lexical representations for words containing the Spanish contrasts we examine. The ability to reject a non-word is contingent on its word counterpart being accurately represented in the lexical entry. We use *d'* ("d-prime") scores as a bias-free measure of perceptual sensitivity to the lexical status of non-words; thus, a higher *d'* indicates more accurate lexical representations for that contrast. Generally, *d'* scores below 0.75 can be interpreted as a lack of discrimination sensitivity. Scores from 0.75 to 3.0 show increasing discrimination sensitivity, and scores above 3.0 show very strong discrimination sensitivity.

Data for the lexical task were not saved for two participants due to a coding error. Trials with timeouts were excluded from the analysis. Participants needed to have responses to minimally 95% of trials in order to be included (i.e., 6 or fewer timeouts). No learner had to be excluded for timeouts.

Despite the fact that the words in the lexical tasks were chosen in order to be familiar to L2 learners, it is likely that some words were unknown, and therefore a response on these trials would not be a reliable reflection of learners' phonolexical knowledge. Because of this, learners' responses on the word familiarity section of the background questionnaire were taken into account. Vocabulary knowledge was evaluated on an individual basis for each participant. For a trial to be included, the participant had to have chosen one of the three highest options on the 6-point word familiarity scale for that word. Vocabulary knowledge was considered for non-word trials as well. The inclusion of non-word trials was evaluated based on the participant's familiarity with the corresponding real word, with the exception of the filler condition where non-words were not based on real words. For example, if the word *desarrollo* 'development' was not known, the non-word counterpart trial *desadollo* was excluded from the analysis. If participants had less than half of word or non-word trials remaining in a condition, their results were excluded from the analysis. Two participants' results were excluded for remaining with less than half of the non-word trials in the /trill-d/ condition and the /f-p/ condition, respectively. The final number of L2 learners who were included in the lexical decision task analyses was 35, with an almost even split between those who completed List 1 (17 participants) and those who completed List 2 (18 participants). **Figure 1** displays the *d'* scores for each condition, excluding trials with timeouts and unknown words. Diamonds represent mean values, and violin plots around the boxplots show the distribution of scores.

Overall, as shown in **Figure 1**, the results of the lexical task show that learners had the lowest scores for the /tap-trill/ condition, followed by the /tap-d/ condition, while the /trill-d/ and /f-p/ conditions were both the most accurate. This suggests that lexical representations are overall most accurate for the trill-d and f-p contrasts, and least accurate for the tap-trill contrast. Analyses of Variance (ANOVAs) were conducted to examine the effects of condition (contrast) and list. A two-way mixed ANOVA was run with *d'* score as the dependent variable, *condition* (/tap-trill/, /tap-d/, /trill-d/, /f-p/) as the within-subjects independent variable, and *list* (1 vs. 2) as the between-subjects independent variable. The ANOVA test and tests for checking the assumptions of an ANOVA were conducted in R using the *rstatix* package v.0.3.1 (Kassambara, 2019). All assumptions for the ANOVA

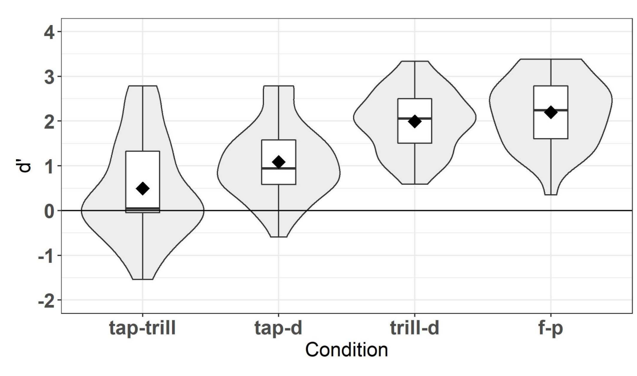


FIGURE 1 | *d'* scores for lexical decision task.

were met regarding normality, sphericity, and homogeneity of variances and covariances. The Bonferroni correction method was used to adjust *p*-values for multiple comparisons in *post hoc* tests, which were conducted with the built-in *stats* package in R version 4.0.2 (R Core Team, 2020). The ANOVA revealed that there was a significant interaction between condition and list, $F(3, 99) = 7.654$, $p < 0.001$. Condition was significant for both lists ($p < 0.001$), such that within each list, conditions differed from each other, with some slight differences. For List 1, only /trill-d/ vs. /f-p/ ($p = 0.734$) and /tap-trill/ vs. /tap-d/ ($p = 1$) did not differ from each other, while in List 2, only /trill-d/ vs. /f-p/ ($p = 1$) and /tap-d/ vs. /f-p/ ($p = 0.216$) did not differ from each other. However, *d*' scores did not differ between lists for any of the conditions (all $p > 0.1$), nor was there a main effect of list ($p = 0.568$). For this reason, it was judged appropriate to combine scores across the two lists for the individual differences analyses. The main effect of condition was significant, $F(3, 99) = 65.412$, $p < 0.001$. When lists were combined, all conditions were significantly different from each other (all $p < 0.001$) with one exception; performance on /trill-d/ was not different from /f-p/ ($p = 1$). This task largely replicated the results of Daidone and Darcy (2014) and yielded substantial variation in scores for the L2 learners, making it suitable for use in the individual differences analyses.

Oddity Task Analysis and Results

The oddity task was used to examine participants' perception ability for the Spanish contrasts that appeared in the lexical task (/tap-trill/, /tap-d/, /trill-d/, and /f-p/). For each contrast, *d*' scores were computed rather than accuracy because the learners showed a strong bias in the /tap-d/ condition and to a lesser extent in the /tap-trill/ condition toward choosing that the trials were the same, and *d*' provides a bias-free measure of perceptual sensitivity. The *d*' scores were calculated by grouping trials as same (AAA, BBB) or different (AAB, BBA, ABA, BAB, ABB, BAA). If participants recognized that one of the sounds was different, even if they did not correctly identify which sound was different, this counted as a hit, whereas if they chose any of the stimuli as different when they were all the same, this was counted as a false alarm. Trials with timeouts were excluded. Participants could not have timeouts on more than 5% of trials (i.e., 7 timeouts) in order to be included; no participant had timeouts on more than 2 trials. Therefore, all 39 learners were included in the analysis. Results of the *d*' analysis are illustrated in **Figure 2**, indicating that learners were highly accurate on the /f-p/ condition and less accurate on the /trill-d/ condition, followed by the /tap-trill/ condition and the /tap-d/ condition, respectively. This result is expected based on the findings of Daidone and Darcy (2014). Notably, the order of accuracy is not the same as for lexical decision, just as Daidone and Darcy found.

Inferential statistics confirmed that learners' performance differed by contrast. A non-parametric Friedman test was run in R using the *rstatix* package v.0.3.1 (Kassambara, 2019) because the data violated the assumptions for a repeated measures ANOVA. Specifically, the results for the /tap-trill/, /trill-d/, and /f-p/ conditions were not normally distributed, and the assumption of sphericity was also violated. The Friedman test was conducted with *d*' *score* as the dependent variable and *condition*

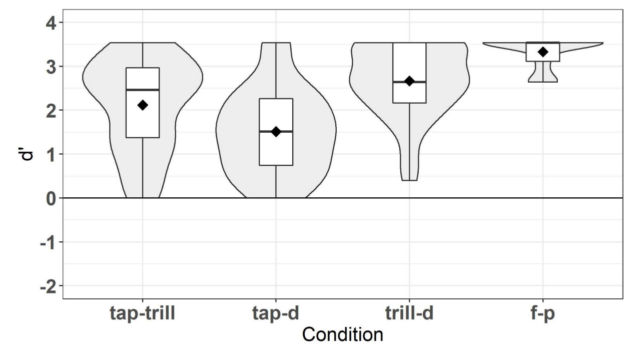


FIGURE 2 | *d'* scores for oddity task.

(/tap-trill/, /tap-d/, /trill-d/, /f-p/) as the independent variable. *Post hoc* tests were adjusted for multiple comparisons with the Bonferroni correction method. Results revealed that *d*' scores were significantly different across conditions, $\chi^2(3) = 69.846$, $p < 0.001$. Pairwise Wilcoxon signed-rank tests found that all conditions were significantly different from each other (all $p < 0.05$). Therefore, these contrasts varied in discriminability. Learners also showed substantial variation within each condition, at least for /tap-trill/, /tap-d/, and /trill-d/, making these scores acceptable for individual differences analyses.

Phonological Short-Term Memory Task Analysis and Results

The PSTM task examined how well participants were able to hold increasingly longer sequences of sounds in memory and compare them. The more accurate they were at correctly identifying if these sequences were the same or different, the greater their PSTM. In order to analyze the PSTM task, the response to each test trial was coded as 1 or 0. If the participant correctly identified the paired sequences of Russian CVC non-words as being in the same order or a different order, they received a 1 for that trial, and if they were incorrect or timed out, they received a 0. No participant had more than one timeout. Participants earned a score out of 8 for each sequence length (4, 5, 6, or 7 non-words) as the block for each sequence length contained 8 trials. In accordance with Zahler and Lord (in press), scores were then weighted by the length of the sequences, such that the score for each length block was multiplied by the length itself (4, 5, 6, or 7). For example, a participant who correctly responded to 6 trials of length 4 received a score of $6 \times 4 = 24$ for those trials. This resulted in a total possible weighted score of 176 $[(8 \times 4) + (8 \times 5) + (8 \times 6) + (8 \times 7) = 176]$. Scores ranged from 78 to 151 ($M = 114$, SD = 18).

Inhibition Task Analysis and Results

The retrieval-induced inhibition task examined participants' inhibitory skill by testing how much slower they responded to memorized words that were inhibited due to the effect of having retrieved semantically related words during the practice phase. A slower reaction time (RT) to these items indicated more inhibition, and thus higher inhibitory control. Inhibitory skill was calculated using median RTs in accordance with the technique reported by Lev-Ari and Peperkamp (2013). First, the median RT was determined for each participant for each of the three conditions in the test phase (practiced, inhibited, and control). The practiced items were those words that also appeared during the practice phase, the inhibited items were those that came from a practiced category, but did not form part of the practice phase, and control items were those that came from a category that was not part of the practice phase at all. For example, if a participant had to type the words *engineer, nurse, carpenter, grape, cherry,* and *orange* during the practice phase, then the RTs for the recognition of these words in the test phase fell under the practiced condition, the other words under the categories *occupations* and *fruits* were part of the inhibited condition, and all words in the *animals* category formed part of the control condition.

Following Darcy et al. (2016), if participants missed all instances of two or more words during the practice phase of the task, they were excluded. Four L2 learners were excluded for this reason, resulting in a total of 35 included participants. For the test part of the task, trials with an RT beyond 2 SD in either direction from the average for that participant were removed. No participant had more than two trials removed for this reason.

The data exhibited extreme outliers and violated the assumptions of sphericity and normality for all conditions for a repeated measures ANOVA. Thus, a non-parametric Friedman test was run in order to examine if *median RT* (dependent variable) differed by *condition* (practiced, inhibited, control). The Friedman test revealed that RT was significantly different across conditions, $\chi^2(2) = 12.189$, $p = 0.002$. Pairwise Wilcoxon signed-rank tests found that, as hypothesized, inhibited items were responded to more slowly than practiced items (Bonferroni adjusted $p = 0.003$). Although median RTs to control items (808 ms) were numerically slower than practiced items (761 ms) and faster than inhibited items (856 ms), median RTs to control items did not significantly differ from inhibited items (adjusted $p = 0.235$) or from practiced items (adjusted $p = 0.184$). An inhibition score for each participant was calculated by dividing the median RT for inhibited items by the median RT for control items; higher values indicate greater inhibitory skill (Lev-Ari and Peperkamp, 2013). Inhibition scores ranged from 0.71 to 1.65 ($M = 1.07$, SD = 0.19).

Attention Control Task Analysis and Results

Participants' ability to selectively attend to the center arrow while ignoring the surrounding arrows (in other words, to respond equally as quickly when the surrounding arrows did not match the direction of the center arrow) served as the measure of attention control. Two L2 learners were excluded from the analysis because they had timeouts on more than 5% of trials, leaving a total of 37 participants. The mean and SD for each participant was calculated, and RTs beyond two SDs from the mean in either direction were excluded. All participants were left with at least 36 trials out of 40 in each condition (i.e., congruent and incongruent). In order to investigate whether there was a significant difference in RTs between congruent and incongruent trials, a two-tailed paired samples t-test was run. Results showed that there was a significant effect of condition, such that congruent trials ($M = 441$ ms) were responded to faster on average than incongruent trials ($M = 467$ ms), $t(36) = -8.401$, $p < 0.001$. For each participant, the mean RT for congruent and incongruent trials was derived and the RT differences between the congruent and incongruent trials (congruent average RT - incongruent average RT) was calculated for the measure of selective attention. Scores closer to zero designate better selective attention, that is, less of a reaction time difference between the congruent and incongruent conditions, although in some cases participants' scores were unexpectedly negative, indicating faster responses to incongruent trials on average. L2 learners exhibited a range of scores, with a min of −23 ms and a max of 69 ms ($M = 27$ ms, SD = 19 ms).

Vocabulary Test Analysis and Results

Learners' ability to recognize real Spanish words at different frequency bands and reject non-words was used to estimate their L2 vocabulary size, with more acceptance of words and rejection of non-words indicating more robust vocabulary knowledge. The measure of vocabulary size was their adjusted vocabulary scores out of 5000 generated by the X_Lex vocabulary test (Meara, 2005). According to the X_Lex manual, these adjusted scores were calculated by subtracting the overall false alarm rate from the hit rate for each frequency band. For example, if a participant scored 20/20 on each of the 5 frequency bands (1K, 2K, 3K, 4K, and 5K), but responded 'yes' to 3 non-words, then their adjusted score for each frequency band would be 17/20. If the number of false alarms was higher than the hit rate, this was coded as a score of 0 for that frequency band. Accuracy was averaged across the frequency bands (0.85 for this example participant whose adjusted score was 17/20 for each frequency band) and multiplied by 5000 to result in a score out of 5000 (in the example participant's case, 4250). All 39 participants were included in the analysis. Participants' vocabulary scores ranged from 400 to 4850 ($M = 2792$, SD = 1110).

Individual Differences Analyses

In order to examine how the individual differences measures (perception, PSTM, inhibition, attention, and vocabulary size) contributed to lexical encoding accuracy, a linear regression analysis was run on the complete dataset, and individual regression analyses were run for each contrast. For all of the analyses, the individual differences measures were converted into z-scores, and lexical decision *d'* scores were used for the dependent variable. The individual differences measures did not exhibit high levels of collinearity; the variance inflation factor (VIF) for variables across all analyses was less than 2, whereas problematic collinearity would be indicated by values of 5 or higher (Heiberger and Holland, 2004, 243). For each analysis, the oddity perception measure always matched the condition used for the lexical measure; for example, in the analysis examining the impact of individual differences on the /tap-trill/ condition in the lexical decision task, only performance on the /tap-trill/ condition was included in the oddity z-score calculation.

While we originally attempted to fit a linear mixed effects model with random intercepts for participants, this resulted in a singular fit with variance and standard deviation of the random intercept both estimated at 0. Thus, we decided to run a linear regression model with fixed effects only. This analysis was run in R with the *stats* package version 3.6.2

(R Core Team, 2020), with lexical decision scores as the dependent variable and all individual differences measures (oddity, PSTM, inhibitory control, attention control, and vocabulary size) and their interactions with condition (/tap-trill/, /tap-d/, /trill-d/, and /f-p/), as well as condition alone, as the independent variables.

For the overall analysis, the multiple regression was significant, $F(23, 96) = 8.446$, $p < 0.001$. As **Table 3** illustrates, vocabulary score was the only significant predictor of overall lexical decision performance, with greater vocabulary size predicting more accurate lexical encoding. Additionally, the /tap-trill/ and /tap-d/ conditions significantly differed from the baseline /f-p/ condition, which replicates the results of the ANOVA on the lexical decision results in section 3.1.1. None of the other main effects or interactions were significant.

The multiple linear regression analyses were run on the data for each contrast in R using the *stats* package, with tables created in part with the *apaTables* package v.2.0.5 (Stanley, 2018). All confidence intervals were calculated with the bootstrap method described in Algina et al. (2008) using the *apa.reg.boot.table* function, as recommended for smaller sample sizes and data that violate the assumptions of normality or homogeneity of variances. Only 30 complete cases remained after excluding participants with missing data points.

TABLE 3 | Summary of regression analysis for lexical decision, all conditions.

Predictor	B	B 95% CI	Std Error B	*t*-Value	*p*	
(Intercept)	2.07	[1.80, 2.34]	0.137	15.093	< 0.001	***
Oddity	0.17	[−0.17, 0.50]	0.169	0.983	0.328	
PSTM	0.18	[−0.10, 0.46]	0.145	1.254	0.213	
Inhibition	0.07	[−0.21, 0.36]	0.147	0.508	0.613	
Flanker	0.14	[−0.17, 0.45]	0.157	0.908	0.366	
Vocab	0.46	[0.16, 0.75]	0.150	3.054	0.003	**
/tap-d/ condition	−1.10	[−1.47, −0.72]	0.190	−5.801	< 0.001	***
/tap-trill/ condition	−1.62	[−1.99, −1.24]	0.189	−8.533	< 0.001	***
/trill-d/ condition	−0.21	[−0.58, 0.17]	0.191	−1.089	0.279	
Oddity x/tap-d/	−0.18	[−0.64, 0.28]	0.236	−0.762	0.448	
Oddity x/tap-trill/	0.11	[−0.36, 0.58]	0.240	0.461	0.646	
Oddity x/trill-d/	−0.09	[−0.57, 0.39]	0.245	−0.36	0.720	
PSTM x/tap-d/	−0.10	[−0.49, 0.30]	0.202	−0.472	0.638	
PSTM x/tap-trill/	0.36	[−0.05, 0.77]	0.209	1.731	0.087	
PSTM x/trill-d/	−0.10	[−0.49, 0.30]	0.201	−0.472	0.638	
Inhibition x/tap-d/	−0.28	[−0.69, 0.14]	0.212	−1.297	0.198	
Inhibition x/tap-trill/	−0.05	[−0.45, 0.36]	0.209	−0.216	0.830	
Inhibition x/trill-d/	−0.03	[−0.44, 0.38]	0.209	−0.135	0.893	
Flanker x/tap-d/	−0.12	[−0.57, 0.32]	0.227	−0.541	0.590	
Flanker x/tap-trill/	0.15	[−0.28, 0.59]	0.221	0.686	0.495	
Flanker x/trill-d/	−0.01	[−0.45, 0.42]	0.223	−0.061	0.952	
Vocab x/tap-d/	0.15	[−0.27, 0.56]	0.214	0.681	0.498	
Vocab x/tap-trill/	−0.01	[−0.48, 0.47]	0.243	−0.035	0.972	
Vocab x/trill-d/	0.08	[−0.33, 0.49]	0.210	0.374	0.709	
Overall Fit	*R^2 = 0.669 p = < 0.001****					

*B = unstandardized regression weight. CI = confidence interval. **p < 0.01, ***p < 0.001.*

The multiple regression analyses for /tap-trill/ ($F(5, 24) = 4.79$, $p = 0.004$), /tap-d/ ($F(5, 24) = 5.449$, $p = 0.002$), and /trill-d/ ($F(5, 24) = 4.908$, $p = 0.003$) were significant (shown in **Tables 4–6**, respectively), while the regression for /f-p/ was not ($F(5, 24) = 1.858$, $p = 0.140$). This indicates that for all contrasts except /f-p/, lexical performance is explained in part by a combination of the other factors. We now consider each contrast in turn. As seen in **Table 4**, PSTM and vocabulary size were significant predictors of lexical decision scores in the /tap-trill/ condition. They each accounted for a similar amount of variance in lexical decision scores (ΔR^2), approximately 26% for PSTM and 22% for vocabulary size.

Table 6 displays the summary of the /tap-d/ analysis, showing that only vocabulary scores were a significant predictor of performance on the lexical decision task in this condition, explaining approximately 15% of the variance in scores.

Table 5 shows the results of the /trill-d/ analysis. Similar to the /tap-d/ analysis, only vocabulary size was a significant predictor of lexical decision performance in this condition, although it explained a larger portion of the variance in this analysis, around 50%.

In sum, vocabulary size was a significant predictor of lexical encoding accuracy in the overall analysis and for three of the four contrasts investigated, specifically /tap-trill/, /tap-d/, and /trill-d/, while PSTM was only significant for the individual /tap-trill/ analysis. Learners' scores in the oddity, flanker, and inhibition tasks were not significant predictors of lexical decision performance for any contrast.

DISCUSSION

While perception was predicted to have a large effect on lexical encoding accuracy, surprisingly there was no effect for any of the analyses. Most models of L2 speech acquisition implicitly or explicitly propose a direct link between perception ability and the accuracy of phonological representations in the lexicon ([SLM] Flege, 1995; [PAM-L2] Best and Tyler, 2007; [L2LP] van Leussen and Escudero, 2015)[2]. The lack of an effect for perception ability in the current study seems to contradict this assumption. Instead, the factor with the largest impact on L2 lexical encoding for most contrasts was revealed to be L2 vocabulary size. These results support the premise of a lexicon-first model like NLM-e, which proposes that learning phonological neighbors aids in the formation of phonetic categories, which in turn leads to refinement in the phonetic detail of existing phonolexical representations (Kuhl et al., 2008), as has been found for young children learning their L1 (see Stoel-Gammon, 2011, for a review). This idea is also touched on by Best and Tyler (2007) in their discussion of PAM-L2, in which they assert that the learning of many minimal pairs would exert pressure on learners' phonological system to begin to distinguish those sounds. This suggests that the acquisition of more and more phonologically similar words forces learners' phonological system to create more detailed representations in order for them to be differentiated.

[2]For an in-depth review of the major L2 phonological models and their connection to lexical encoding, see Daidone (2020).

TABLE 4 | Summary of regression analysis for lexical decision, /tap-trill/ condition.

Predictor	B	B 95% CI	Std Error B	*t*-value	ΔR^2	ΔR^2 95% CI	*p*	
(Intercept)	0.46	[0.15, 0.78]	0.163	2.795	NA	NA	0.010	*
Oddity	0.15	[−0.29, 0.56]	0.206	0.712	0.01	[0.00, 0.11]	0.483	
PSTM	0.62	[0.26, 0.94]	0.176	3.513	0.26	[0.03, 0.45]	0.002	**
Inhibition	−0.02	[−0.46, 0.26]	0.190	−0.085	0.00	[0.00, 0.08]	0.933	
Flanker	0.38	[−0.14, 0.77]	0.205	1.880	0.07	[0.00, 0.24]	0.072	
Vocab	0.61	[0.19, 1.12]	0.189	3.244	0.22	[0.03, 0.43]	0.003	**
Overall Fit	*R^2 = 0.499, 95% CI[0.31,0.77], p = 0.004***							

*B = unstandardized regression weight. ΔR^2 = the change in R^2 when the variable is removed. Numbers in brackets indicate the lower and upper limits of a bootstrapped 95% confidence interval. * $p < 0.05$, **$p < 0.01$.*

TABLE 5 | Summary of regression analysis for lexical decision,/trill-d/ condition.

Predictor	B	B 95% CI	Std Error B	t-value	ΔR^2	ΔR^2 95% CI	*p*	
(Intercept)	1.86	[1.66, 2.07]	0.099	18.892	NA	NA	< 0.001	***
Oddity	0.08	[−0.18, 0.33]	0.132	0.590	0.01	[0.00, 0.10]	0.561	
PSTM	0.09	[−0.12, 0.30]	0.104	0.831	0.01	[0.00, 0.15]	0.414	
Inhibition	0.05	[−0.25, 0.24]	0.110	0.425	0.00	[0.00, 0.10]	0.675	
Flanker	0.13	[−0.19, 0.45]	0.117	1.106	0.03	[0.00, 0.24]	0.280	
Vocab	0.54	[0.31, 0.75]	0.109	4.922	0.50	[0.16, 0.69]	< 0.001	***
Overall Fit	*R^2 = 0.506, 95% CI[0.32,0.82], p = 0.003***							

*B = unstandardized regression weight. ΔR^2 = the change in R^2 when the variable is removed. Numbers in brackets indicate the lower and upper limits of a bootstrapped 95% confidence interval. **$p < 0.01$, ***$p < 0.001$.*

TABLE 6 | Summary of regression analysis for lexical decision,/tap-d/ condition.

Predictor	B	B 95% CI	Std Error B	*t*-value	ΔR^2	ΔR^2 95% CI	*p*	
(Intercept)	0.97	[0.70, 1.20]	0.120	8.046	NA	NA	< 0.001	***
Oddity	0.15	[−0.22, 0.46]	0.157	0.988	0.02	[0.00, 0.15]	0.333	
PSTM	0.03	[−0.25, 0.30]	0.140	0.225	0.00	[0.00, 0.08]	0.824	
Inhibition	−0.20	[−0.44, 0.01]	0.136	−1.453	0.04	[0.00, 0.17]	0.159	
Flanker	0.00	[−0.22, 0.28]	0.144	0.017	0.00	[0.00, 0.05]	0.987	
Vocab	0.49	[0.16, 0.91]	0.176	2.783	0.15	[0.01, 0.41]	0.010	*
Overall Fit	*R^2 = 0.532, 95% CI[0.39,0.79], p = 0.002***							

*B = unstandardized regression weight. ΔR^2 = the change in R^2 when the variable is removed. Numbers in brackets indicate the lower and upper limits of a bootstrapped 95% confidence interval. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.*

Thus, the accuracy of learners' representations appears to stem more from properties of their lexicon over their perception abilities. However, it is also important to note that the Spanish contrasts examined in the current study were not particularly difficult for the English learners, with *d'* scores in the oddity perception task all averaging above 1.5 across conditions. It may be that beyond a certain threshold of accuracy, differences in perception ability no longer have an appreciable difference on lexical encoding accuracy, whereas they would be very important for more difficult contrasts. This hypothesis is in line with the results of Llompart (2021b), who found that vocabulary size but not perception ability was a significant predictor of lexical encoding ability for advanced German-speaking learners of English, who had more accurate perception abilities, while perception ability but not vocabulary size was significant for intermediate learners, who had less accurate perception abilities. Furthermore, vocabulary size can be thought of as a proxy for proficiency and has been used as such in previous research (e.g., Darcy et al., 2016). It is probable that those learners with a higher proficiency level have had more L2 input, leading to more detailed and delineated representations because their exemplars are based on more examples. However, more input on its own would likely not be sufficient unless perception, or perhaps more accurately attentional cue weighing, is at a stage where exemplars can be fine-grained in terms of L2-relevant phonetic details. Otherwise, their exemplars are likely to reflect heavy influence from the L1 phonology (Maye, 2007). Perhaps additional analyses divided by proficiency level would reveal more about the effects of perception ability versus vocabulary size.

PSTM was also a significant factor in the current study, but for only the /tap-trill/ contrast when individual contrasts were examined, for which it explained slightly more variance than vocabulary size. Perhaps PSTM is important solely for the /tap-trill/ contrast because this is the only contrast in the current study in which the L2 sounds would overwhelmingly be assimilated to the same L1 sound, in this case English /ɹ/ (see Rose, 2012). Therefore, it may be that differences in phonological short-term memory come into play when sounds are differentiated along a dimension not used phonologically in the L1, making it more important to be able to hold finely detailed representations in the phonological loop long enough so that these L2-relevant details can be transferred to long-term representations. Because those with lower PSTM cannot hold phonetic details in memory for very long, when it comes time to convert the L2 sounds stored in the phonological loop into long-term representations, the memory traces may have degraded into less specific representations, such that these low-PSTM learners no longer retain a difference between the Spanish rhotics. Further research on the importance of PSTM for different types of contrasts could shed light on this question.

None of the regression analyses found a significant effect of inhibitory control or attention control. One possibility is that rather than directly impacting L2 representations, the effect goes in the opposite direction, and these cognitive abilities are instead enhanced by learning an L2. A wealth of research on bilingualism has generally found that bilingual individuals have stronger cognitive abilities than monolinguals, including attention control and inhibitory control (e.g., Bialystok et al., 2005; Adesope et al., 2010; Long et al., 2020). For example, Long and colleagues found that the Gaelic level of L2 learners predicted their attention switching ability, and improvement in L2 Gaelic skills corresponded to gains in attention switching (Long et al., 2020). Another possible explanation is that there was a problem with the specific tasks used in the current study or the way they were scored, since some participants displayed unexpected reaction time tendencies across conditions in both tasks. In fact, Hedge et al. (2018) argue that these kinds of widely-used cognitive tasks do not produce reliable individual differences in general. They state that tasks such as the flanker task became popular because of their reliable and easily replicable results at the group level, but this translates into low between-subject variability that is not reliably replicated across sessions. They found that none of the cognitive tasks they examined, including the flanker task, had reliability metrics at 0.8 or above, which is the accepted standard for clinical uses. Thus, more work may be needed in order to create more reliable tasks or more reliable ways of calculating scores for existing tasks in order to conduct valid individual differences research.

Overall, this study shows that L2 lexical encoding is affected by factors beyond perception, specifically L2 vocabulary size and phonological short-term memory. This corroborates previous research showing that learners' phonolexical representations are fuzzy, above and beyond their ability to perceive the sounds within those words correctly. Additionally, this study reveals that the impact of individual differences depends on the contrast under examination, although acquiring a large vocabulary in the L2 appears to be the most important factor in mediating fuzzy lexical representations. Additional research is needed to determine if these results hold across other contrasts and language pairings, and to ascertain what other factors may be at play in the L2 lexical encoding of these contrasts, such as frequency, phonological neighborhood density, and phonetic variability in the input (see Llompart (2021a), under this Research Topic).

DATA AVAILABILITY STATEMENT

The datasets and analyses presented in this study can be found in the Open Science Framework repository at osf.io/w9mnr.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Human Research Protection Program (HRPP) at Indiana University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

As this article describes part of DD's dissertation research, she was the main contributor for all parts of this study and she alone performed the data collection. ID was the advisor for this work and aided in conceptualization and design of the experiment as well as the analyses and interpretation of the results. DD wrote the first draft of the manuscript, which ID revised and expanded. Both authors approved the final submitted version.

FUNDING

This research was funded by an Indiana University Grant-in-Aid of Doctoral Research to DD.

ACKNOWLEDGMENTS

This article is based on DD's dissertation, Daidone (2020), for which ID was the advisor. We are grateful for the helpful insights provided by the other dissertation committee members as well as the members of the IU Second Language Psycholinguistics Lab.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.688356/full#supplementary-material

REFERENCES

Adesope, O. O., Lavin, T., Thompson, T., and Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Rev. Educ. Res.* 80, 207–245. doi: 10.3102/0034654310368803

Algina, J., Keselman, H. J., and Penfield, R. D. (2008). Note on a confidence interval for the squared semipartial correlation coefficient. *Educ. Psychol. Meas.* 68, 734–741. doi: 10.1177/0013164407313371

Aliaga-García, C., Mora, J. C., and Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Stud. Contemp. Linguist.* 47:1. doi: 10.2478/psicl-2011-2012

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Appl. Psychol.* 37, 1221–1251. doi: 10.1017/S0142716415000557

Baddeley, A. (2003). Working memory and language: an overview. *J. Commun. Disord.* 36, 189–208. doi: 10.1016/S0021-9924(03)00019-14

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: in Honor of James Emil Flege*, eds O.-S. Bohn and M. J. Munro (Philadelphia, PA: John Benjamins), 13–34. doi: 10.1075/lllt.17.07bes

Bialystok, E., Martin, M. M., and Viswanathan, M. (2005). Bilingualism across the lifespan: the rise and fall of inhibitory control. *Int. J. Biling* 9, 103–119. doi: 10.1177/13670069050090010701

Bugg, J. M., and Crump, M. J. C. (2012). In support of a distinction between voluntary and stimulus-driven control: a review of the literature on proportion congruent effects. *Front. Psychol.* 3:367. doi: 10.3389/fpsyg.2012.00367

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., and Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Appl. Psychol.* 33, 643–664. doi: 10.1017/S0142716411000518

Bundgaard-Nielsen, R. L., Best, C. T., and Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Stud. Second Lang. Acquis* 33, 433–461. doi: 10.1017/S0272263111000040

Cerviño-Povedano, E., and Mora, J. C. (2015). Spanish EFL learners' categorization of /iː-ɪ/ and phonological short-term memory. *Procedia - Soc. Behav. Sci.* 173, 18–23. doi: 10.1016/j.sbspro.2015.02.024

Cook, S. V. (2012). *Phonological form in L2 Lexical Access: Friend or foe?*. dissertation, College Park (MD): University of Maryland.

Cook, S. V., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *Ment. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Daidone, D. (2020). *How Learners Remember Words in their Second Language: The Impact of Individual Differences in Perception, Cognitive Abilities, and Vocabulary Size*. dissertation, Bloomington (IN): Indiana University.

Daidone, D., and Darcy, I. (2014). "Quierro comprar una guitara: lexical encoding of the tap and trill by L2 learners of Spanish," in *Selected Proceedings of the 2012 Second Language Research Forum: Building Bridges between Disciplines*, eds R. T. Miller, K. I. Martin, C. M. Eddington, A. Henery, N. Marcos Miguel, A. M. Tseng, et al. (Somerville, MA: Cascadilla Proceedings Project), 39–50.

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Ment. Lex.* 8, 372–420. doi: 10.1075/ml.8.3.06dar

Darcy, I., and Mora, J. C. (2016). "Executive control and phonological processing in language acquisition: the role of early bilingual experience in learning an additional language," in *Cognitive Individual Differences in Second Language Processing and Acquisition*, eds G. Granena, D. O. Jackson, and Y. Yilmaz (Philadelphia, PA: John Benjamins), 249–277. doi: 10.1075/bpa.3.12dar

Darcy, I., Mora, J. C., and Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. *Concordia Work. Pap. Appl. Linguist* 5, 115–127. doi: 10.1007/s11786-010-0026-25

Darcy, I., Mora, J. C., and Daidone, D. (2016). The role of inhibitory control in second language phonological processing. *Lang. Learn.* 66, 741–773. doi: 10.1111/lang.12161

Darcy, I., Park, H., and Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: the relation between cognitive abilities and phonological processing. *Learn. Individ. Differ.* 40, 63–72. doi: 10.1016/j.lindif.2015.04.005

de Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* 47, 1–12. doi: 10.3758/s13428-014-0458-y

Detrixhe, K. A. (2015). *The Effect of Studying Abroad on the Acquisition of the Spanish Trill for Second Language Learners*. dissertation, Lawrence (KS): University of Kansas.

Diamond, A. (2013). Executive functions. *Annu. Rev. Psychol.* 64, 135–168. doi: 10.1146/annurev-psych-113011-143750

Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access. *Learn. Individ. Differ.* 22, 680–689. doi: 10.1016/j.lindif.2012.05.005

Elvin, J. D. (2016). *The Role of the Native Language in Non-native Perception and Spoken word Recognition: English vs. Spanish Learners of Portuguese*. dissertation, Penrith: Western Sydney University.

Eriksen, C. W. (1995). The flankers task and response competition: a useful tool for investigating a variety of cognitive problems. *Vis. Cogn.* 2, 101–118. doi: 10.1080/13506289508401726

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., and Posner, M. I. (2002). Testing the efficiency and independence of attentional networks. *J. Cogn. Neurosci.* 14, 340–347. doi: 10.1162/089892902317361886

Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, ed. W. Strange (Timonium, MD: York Press), 233–277.

Francis, A. L., Baldwin, K., and Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Percept. Psychophys.* 62, 1668–1680. doi: 10.3758/BF03212164

Friedman, N. P., and Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *J. Exp. Psychol. Gen.* 133, 101–135. doi: 10.1037/0096-3445.133.1.101

Ghaffarvand Mokari, P., and Werner, S. (2019). On the role of cognitive abilities in second language vowel learning. *Lang. Speech* 62, 260–280. doi: 10.1177/0023830918764517

Gökgöz-Kurt, B. (2016). *Attention Control and the Effects of Online Training in Improving Connected Speech Perception by Learners of English as a Second Language*. dissertation, Columbia (SC): University of South Carolina.

Guion, S. G., and Pederson, E. (2007). "Investigating the role of attention in phonetic learning," in *Language Experience in Second Language Speech Learning: in Honor of James Emil Flege*, eds O.-S. Bohn and M. Munro (Amsterdam: John Benjamins), 57–77. doi: 10.1075/lllt.17.09gui

Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980

Hedge, C., Powell, G., and Sumner, P. (2018). The reliability paradox: why robust cognitive tasks do not produce reliable individual differences. *Behav. Res. Methods* 50, 1166–1186. doi: 10.3758/s13428-017-0935-931

Heiberger, R. M., and Holland, B. (2004). *Statistical Analysis and Data Display: An Intermediate Course with Examples in S-PLUS, R, and SAS*. New York, NY: Springer.

Herd, W. (2011). *The Perceptual and Production Training of /d,ɾ,r/ in L2 Spanish: Behavioral, Psycholinguistic, and Neurolinguistic Evidence*. dissertation. Lawrence (KS): University of Kansas.

Kassambara, A. (2019). *rstatix: Pipe-friendly Framework for Basic Statistical Tests. R Package Version 0.3.1*. Available online at: https://cran.r-project.org/package=rstatix%0A (accessed December 3, 2020).

Kim, Y. H., and Hazan, V. (2010). "Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates," in *Proceedings of the New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, eds K. Dziubalska-Kolaczyk, M. Wrembel, and M. Kul (Poznań: School of English, Adam Mickiewicz University), 251–256.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Trans. R. Soc. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Lengeris, A., and Nicolaidis, K. (2014). "English consonant confusions by Greek listeners in quiet and noise and the role of phonological short-term memory," in *Proceedings of the 15th Annual Conference of the International Speech Communication Association (INTERSPEECH)*. (Singapore), 534–538.

Lev-Ari, S., and Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *J. Phon.* 41, 320–331. doi: 10.1016/j.wocn.2013.06.002

Lev-Ari, S., and Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *J. Phon.* 47, 36–46. doi: 10.1016/j.wocn.2014.09.001

Llompart, M. (2021a). Lexical and phonetic influences on the phonolexical encoding of difficult second-language contrasts: insights from nonword rejection. *Front. Psychol.* 12:659852. doi: 10.3389/fpsyg.2021.659852

Llompart, M. (2021b). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Biling. Lang. Cogn.* 24, 481–496. doi: 10.1017/S1366728920000656

Long, M. R., Vega-Mendoza, M., Rohde, H., Sorace, A., and Bak, T. H. (2020). Understudied factors contributing to variability in cognitive performance related to language learning. *Biling. Lang. Cogn.* 23, 801–811. doi: 10.1017/s1366728919000749

MacKay, I. R. A., Meador, D., and Flege, J. E. (2001). The identification of English consonants by native speakers of Italian. *Phonetica* 58, 103–125. doi: 10.1159/000028490

Mairano, P., and Santiago, F. (2020). What vocabulary size tells us about pronunciation skills: issues in assessing L2 learners. *J. French Lang. Stud.* 30, 141–160. doi: 10.1017/S0959269520000010

Majerus, S., Poncelet, M., Van der Linden, M., and 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

Maye, J. (2007). "Learning to overcome L1 phonological biases," in *Proceedings of the 16th International Congress of Phonetics Sciences*, eds J. Trouvain and W. J. Barry (Saarbrücken), 63–66.

Meara, P. M. (2005). *X_Lex: The Swansea Vocabulary Levels Test.* Swansea: Lognostics.

Metsala, J. L., and Walley, A. C. (1998). "Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability," in *Word Recognition in Beginning Literacy*, eds J. L. Metsala and L. C. Ehri (Mahwah, NJ: Lawrence Erlbaum), 89–120.

Miyake, A., and Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: four general conclusions. *Curr. Dir. Psychol. Sci.* 21, 8–14. doi: 10.1177/0963721411429458

Mora, J. C., and Darcy. (2016). "The relationship between cognitive control and pronunciation in a second language," in *Second Language Pronunciation Assessment: Interdisciplinary Perspectives*, eds T. Isaacs and P. Trofimovich (Bristol: Multilingual Matters), 95–120. doi: 10.21832/isaacs6848

Nagle, C. (2013). "A reexamination of ultimate attainment in L2 phonology: length of immersion, motivation, and phonological short-term memory," in *Selected Proceedings of the 2011 Second Language Research Forum*, eds E. Voss, S.-J. D. Tai, and Z. Li (Somerville, MA: Cascadilla Proceedings Project), 148–161.

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: views from cognitive and personality psychology and a working inhibition taxonomy. *Psychol. Bull.* 126, 220–246. doi: 10.1037/0033-2909.126.2.220

O'Brien, I., Segalowitz, N., Freed, B., and Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Stud. Second Lang. Acquis* 29, 557–581.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Core Team.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., and Dupoux, E. (2010). "A psycholinguistic perspective on the acquisition of phonology," in *Laboratory Phonology 10: Variation, Phonetic Detail and Phonological Representation*, eds C. Fougeron, B. Kühnert, M. d'Imperio, and N. Vallée (Berlin: Mouton de Gruyter), 311–340.

Rose, M. (2010). "Differences in discriminating L2 consonants: a comparison of Spanish taps and trills," in *Proceedings of the Selected Proceedings of the 2008 Second Language Research Forum*, ed. M. T. Prior (Somerville, MA: Cascadilla Proceedings Project), 181–196.

Rose, M. (2012). Cross-language identification of Spanish consonants in English. *Foreign Lang. Ann.* 45, 415–429. doi: 10.111/j.1944-9720.2012.01197.x

Safronova, E. (2016). *The Role of Cognitive Ability in the Acquisition of Second Language Perceptual Phonological Competence*. dissertation, Barcelona: University of Barcelona.

Safronova, E., and Mora, J. C. (2012). "Acoustic and phonological memory in L2 vowel perception," in *At a time of crisis: English and American studies in Spain*, eds S. Martín Alegre, M. Moyer, E. Pladevall, and S. Tubau (Barcelona: Departament de Filologia Anglesa i de Germanística, Universitat Autònoma de Barcelona/AEDEAN).

Sebastián-Gallés, N., and Baus, C. (2005). "On the relationship between perception and production in L2 categories," in *Twenty-first Century Psycholinguistics: Four Cornerstones*, ed. A. Cutler (Mahwah, NJ: Lawrence Erlbaum Associates), 279–292.

Simonchyk, A., and Darcy, I. (2017). "Lexical encoding and perception of palatalized consonants in L2 Russian," in *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference*, eds M. O'Brien and J. Levis (Ames, IA: Iowa State University), 121–132.

Speciale, G., Ellis, N. C., and Bywater, T. (2004). Phonological sequence learning and short-term store capacity determine second language vocabulary acquisition. *Appl. Psychol.* 25, 293–321. doi: 10.1017/s0142716404001146

Stanley, D. (2018). *apaTables: Create American Psychological Association (APA) Style Tables. R package Version 2.0.5*. Washington, DC: APA. Available online at: https://cran.r-project.org/package=apaTables (accessed February 1, 2020).

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *J. Child Lang.* 38, 1–34. doi: 10.1017/S0305000910000425

Strange, W., and Shafer, V. (2008). "Speech perception in second language learners: the re-education of selective perception," in *Phonology and Second Language Acquisition,* eds. J. G. Hansen Edwards and M. I. Zampini (Amsterdam: John Benjamins), 153–191.

van Leussen, J.-W., and Escudero, P. (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000

Vihman, M. M., and Croft, W. (2007). Phonological development: toward a "radical" templatic phonology. *Lingusitics* 45, 683–725. doi: 10.1515/LING.2007.021

Vitevitch, M. S., Stamer, M. K., and Kieweg, D. (2012). Short research note: the beginning Spanish lexicon: a web-based interface to calculate phonological similarity among Spanish words in adults learning Spanish as a foreign language. *Second Lang. Res.* 28, 103–112. doi: 10.1177/0267658311432199

Walley, A. C. (2007). "Speech learning, lexical reorganization, and the development of word recognition by native and non-native English speakers," in *Language Experience in Second Language Speech Learning: in Honour of James Emil Flege*, eds O.-S. Bohn and M. J. Munro (Amsterdam: John Benjamins), 315–330. doi: 10.1075/lllt.17.27wal

Zahler, S., and Lord, G. (in press). The role of phonological short-term memory in second language phonology: exploring vowel quality among English-speaking learners of Spanish. *Rev. Española Lingüística Apl. J. Appl. Linguist*.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 03 August 2021
doi: 10.3389/fpsyg.2021.686485

Emerging ASL Distinctions in Sign-Speech Bilinguals' Signs and Co-speech Gestures in Placement Descriptions

Anne Therese Frederiksen[1,2]*

[1] Department of Linguistics University of California, San Diego, La Jolla, CA, United States, [2] Department of Language Science, University of California, Irvine, Irvine, CA, United States

Previous work on placement expressions (e.g., "she put the cup on the table") has demonstrated cross-linguistic differences in the specificity of placement expressions in the native language (L1), with some languages preferring more general, widely applicable expressions and others preferring more specific expressions based on more fine-grained distinctions. Research on second language (L2) acquisition of an additional spoken language has shown that learning the appropriate L2 placement distinctions poses a challenge for adult learners whose L2 semantic representations can be non-target like and have fuzzy boundaries. Unknown is whether similar effects apply to learners acquiring a L2 in a different sensory-motor modality, e.g., hearing learners of a sign language. Placement verbs in signed languages tend to be highly iconic and to exhibit transparent semantic boundaries. This may facilitate acquisition of signed placement verbs. In addition, little is known about how exposure to different semantic boundaries in placement events in a typologically different language affects lexical semantic meaning in the L1. In this study, we examined placement event descriptions (in American Sign Language (ASL) and English) in hearing L2 learners of ASL who were native speakers of English. L2 signers' ASL placement descriptions looked similar to those of two Deaf, native ASL signer controls, suggesting that the iconicity and transparency of placement distinctions in the visual modality may facilitate L2 acquisition. Nevertheless, L2 signers used a wider range of handshapes in ASL and used them less appropriately, indicating that fuzzy semantic boundaries occur in cross-modal L2 acquisition as well. In addition, while the L2 signers' English verbal expressions were not different from those of a non-signing control group, placement distinctions expressed in co-speech gesture were marginally more ASL-like for L2 signers, suggesting that exposure to different semantic boundaries can cause changes to how placement is conceptualized in the L1 as well.

Keywords: speech-sign bilingualism, caused motion events, bidirectional language influences, sign language, co-speech gestures, iconicity, second language, fuzzy lexical representations

OPEN ACCESS

Edited by:
Andreas Opitz,
Leipzig University, Germany

Reviewed by:
Karen Emmorey,
San Diego State University, United States
Anne Wienholz,
University of Hamburg, Germany

****Correspondence:***
Anne Therese Frederiksen
a.t.frederiksen@gmail.com

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *26 March 2021*
Accepted: *23 June 2021*
Published: *03 August 2021*

Citation:
Frederiksen AT (2021) Emerging ASL Distinctions in Sign-Speech Bilinguals' Signs and Co-speech Gestures in Placement Descriptions. Front. Psychol. 12:686485. doi: 10.3389/fpsyg.2021.686485

INTRODUCTION

In learning how to say the equivalent of "the woman put the cup on the table" in a second language, many learners face the challenge of semantic reconstruction. A speaker whose first language (L1) is English and whose second language (L2) is Dutch must learn that Dutch does not have one verb that corresponds to "put" in English. Instead, when describing an event of putting in Dutch, the speaker must choose between different verbs. This choice requires attention to the shape and orientation of the object being placed. Thus, the learner must not only learn the appropriate vocabulary in the target language but may also need to reorganize their conceptualization of placement events. This is a challenge for many learners whose tendency to transfer semantic boundaries from the L1 onto the L2 can result in non-target like use of verbs of placement, indicating fuzzy placement semantics. Unknown is whether differences in semantic transparency in the target-language may help acquisition. Like spoken languages, sign languages use different verbal distinctions in descriptions of placement of different objects. Unlike spoken language verbs in many languages, however, placement verbs are often highly iconic in sign languages. They involve handshapes reflecting visual properties of their referents, and/or kinesthetic properties of how an entity is handled. Placement descriptions in sign languages therefore offer a transparent link between elements of the world and their linguistic encoding. It is unknown whether such transparency facilitates acquiring placement expressions for hearing second language learners of a sign language, or whether they experience the same difficulties in acquiring novel, semantic distinctions as do hearing second language learners of a spoken language.

Also poorly understood are the consequences for L1 placement semantics of learning a typologically different L2. The process of acquiring target-like semantic boundaries may require the learner to engage in semantic reconstruction. As the L2 is fully or partially acquired, this process may come to influence the L1, creating a system where the semantic boundaries of the L1 are (temporarily) fuzzy and unstable and consequently may differ from that of monolinguals and bilinguals with a different L2.

The present study aims to address these gaps in our knowledge by investigating placement descriptions in native English speakers learning American Sign Language (ASL) as an L2. Placement expressions are highly transparent in American Sign Language and at the same time, they exhibit some form overlap with co-speech gestures used in placement descriptions. We take advantage of these facts to ask (1) whether acquiring target-like placement verbs is challenging for different-modality L2 learners as has been shown for same-modality learners, or whether the transparency of ASL placement verbs decreases the difficulty of this task, and (2) whether the learners' English placement descriptions (speech and gesture) show evidence of influence from ASL.

Placement Events

Languages show considerable differences in the expression of placement events (Kopecka and Narasimhan, 2012). A placement event is a type of caused motion event, in which an agent moves something somewhere, e.g., putting a book on a bookshelf. Studies have shown that the descriptions of placement is a typologically quite diverse domain cross-linguistically, not least in terms of verb semantics (see Bohnemeyer and Pederson, 2011; Gullberg, 2011a; Slobin et al., 2011; Kopecka and Narasimhan, 2012). This is perhaps surprising. Given that speakers from different cultures share similar visual and motor experiences with respect to placing objects, we might expect them to describe those experiences in similar ways. However, studies from the last decades have shown that this is far from the case (Ameka and Levinson, 2007; Bohnemeyer and Pederson, 2011; Kopecka and Narasimhan, 2012). Narasimhan et al. (2012) note that languages such as Hungarian, Kalasha, Hindi, and Tamil use a semantically general verb for "put" (as do languages like English and French). This type of single-term or general placement verb language is in opposition to multi-term or specific placement verb languages, such as Tzeltal, which requires selection of one of numerous verb roots to describe a placement event. Languages such as Dutch, Swedish, Polish, and Yeli Dnye are a slightly different kind of multi-term languages. They select a verb from small set of so-called posture verbs, depending on several factors, including the orientation of the object being placed. For example, German, also a posture verb language, distinguishes between the verbs "*stellen*" "to put upright," "*setzen*" "to set," "*legen*" "to lay," "*stecken*" "to stick," and "*hängen*" "to hang" (De Knop, 2016).

Relevant semantic distinctions show up in the co-speech gestures of a language, as well as in speech (Hoetjes, 2008; Gullberg, 2009a, 2011a). Speakers frequently accompany their words with co-speech gesture (McNeill, 1992; Kendon, 2004). Many co-speech gestures are iconic representations of some part of the speech content, that is, they are handshapes that share form properties with the represented entity or action (Kendon, 1980; McNeill and Levy, 1982; McNeill, 1992). Because placement descriptions denote placement actions, which are similar across languages, we might expect speakers to accompany their descriptions with similar co-speech gestures irrespective of the variation in semantic distinctions in different languages. For example, gesturers across languages might use handshapes similar to the motor actions used to perform placement of different items, e.g., a "cup" handshape when talking about the placement of a cup or a glass, but a pincer handshape for small objects, such as beads or coins. However, research has shown that cross-linguistic differences in placement events extend beyond the verbal component of language, and that co-speech gestures instead exhibit patterns specific to language or language type (Hoetjes, 2008; Gullberg, 2009a, 2011a). In particular, speakers of general placement verb languages like French tend to use gestures that reflect the focus in their verbal expression on the act of moving something. This means that they gesture mainly about the direction or path of an object being moved. Conversely, speakers of multi-term, specific placement verb languages like Dutch who have to select verbs in part based on properties of the object being placed typically use gestures that represent form properties of the figure object (Gullberg, 2011a). **Figure 1** shows an example of a general placement verb system type path gesture compared to a specific placement verb system type

FIGURE 1 | Path gesture in description of placing a clothes hanger (left) and figure + path gesture in description of placing a speaker (right).

FIGURE 2 | ASL Sign CUP.

gesture including both information about the figure object and the path of the placement event.

Expression of Placement Events in English and American Sign Language

English tends to encode placement with the verb "put." Although English has specific placement verbs based on posture ("set" and "lay"), they are infrequent in placement descriptions (Pauwels, 2000) and English is categorized as a general placement verb language. Hoetjes (2008) examined English speakers' placement verbs and placement gestures. She found that English speakers as a group used "put" in 59% of their placement descriptions, and "place," another general placement verb, in 10% of their descriptions. In another experiment using the same stimuli, Gullberg (2009a) similarly found the mean proportion of "put" to be 61% for native English speakers. In both studies, the mean percentage of gestures incorporating information about the figure object was below 40% for native English speakers, and correspondingly, the proportion of path-only gestures was over 60%.

To date, no studies have specifically examined placement verbs in a sign language. American Sign Language (ASL) is the primary language of most Deaf individuals in the U.S. and parts of Canada. ASL is produced with the hands and body and perceived with the eyes. Expressing language in the visual-manual modality appears to afford a high degree of iconicity (Perniss et al., 2010). For example, the ASL sign "CUP" involves a sideways C-handshape, similar to the handshape one would use to hold a cup (**Figure 2**). In this paper, we use capitalized letters to indicate sign glosses and letters/numbers to indicate handshapes (the relevant handshapes are pictured in the **Supplementary Images**). Following convention, ASL signs are glossed with English words, but note that sign and gloss are not always translation equivalents.

While no studies have looked specifically at how placement is expressed in ASL, some aspects of ASL and other sign languages that are relevant for understanding signed placement verbs have been investigated in previous work. Specifically, most sign languages have a system of *classifiers* (Aronoff et al., 2005; Zwitserlood, 2012) that play an important role in this domain. Classifiers are handshapes that represent something about the object being described, e.g., shape and size, semantic class, or how an agent would handle the object. There are two broad categories of classifiers (Zwitserlood, 2012): handling (or *handle*) classifiers, where the hand(s) represent(s) how the entity is held by an agent (**Figure 3**), and entity classifiers, where the hand(s) represent(s) the entity (**Figure 4**).

Although classifiers represent information about the figure object, it is not the case that there are unlimited gradient distinctions in classifiers; rather, there exists a set of conventionalized handshapes, where each is conventionally used for specific types of objects. For example, the C handshape handling classifier is used for tall cylinder-like objects like vases, cups, bottles, etc., and the flattened O handshape handling classifier is used for thin flat objects like books, papers, blankets, etc. (Zwitserlood, 2012). Both handling and entity classifiers can be incorporated into verbs of motion and location, sometimes called classifier verbs or classifier predicates (Supalla, 1982; Aronoff et al., 2003, 2005). Verbs like "PUT" and "MOVE" are examples of verbs that can be classifier verbs (e.g., Slobin et al., 2003; Slobin, 2013) and can be used in placement descriptions[1]. When describing how objects are used or manipulated, ASL signers tend to incorporate handling classifiers, rather than entity classifiers into verbs (Padden et al., 2015). Thus, ASL appears to prototypically express placement events with verbs like "MOVE" and "PUT" with incorporated handling classifiers. Despite being similar to languages like English and French in using general placement verbs such as "MOVE" and "PUT" as the basic verb, ASL is differentiated from these languages by the frequent incorporation into the verb of an additional morpheme (the classifier) which specifies shape and orientation of the figure object. Because of this, ASL can be considered a specific

[1]"MOVE" and "PUT" both use the same flattened O handshape in the citation form (see **Supplemental Images**). The two verbs differ in their movements: whereas "PUT" has a defined ending point only, "MOVE" has both a defined beginning and ending point.

FIGURE 3 | Handling classifier representing an agent holding a tablecloth.

FIGURE 4 | Entity classifier representing plates.

placement verb language[2]. We confirmed this with data from two Deaf, native ASL signers, which will be described in more detail below.

Placement Events in the Context of L2 Acquisition

Expressing placement in a second language (L2) not only requires learning placement verbs, locative expressions and appropriate syntax. It also requires learning the semantic boundaries of placement words, which can differ, even between words that are cognates across languages. Learning new semantic boundaries requires first detecting the relevant difference and then mentally rearranging concepts and shifting boundaries accordingly. Rearranging concepts resulting from the semantics of the first language (L1) poses a challenge for the learner (Ijaz, 1986; Kellerman, 1995). This is because native language learning habituates the individual to thinking in ways that are compatible with available means of expression, i.e., to what Slobin (1996) calls "thinking for speaking." To become target-like in placement descriptions, many L2 learners must therefore learn a new way of categorizing semantically. This can cause a variety of issues in the L2, including L1 transfer, that is, mapping semantic boundaries from the native language onto the L2, and fuzzy semantic boundaries. Non-native patterns can arise both when going from a more general to a more complex placement verb system and vice versa (Cadierno et al., 2016). When the L1 uses a general placement verb (e.g., "put" in English) and the L2 distinguishes between several specific placement verbs (e.g., "zetten" vs. "leggen" in Dutch), learners overuse one verb and do not maintain obligatory semantic distinctions (Viberg, 1998; Gullberg, 2009a). In cases where the L1 has several, specific placement verbs, and the L2 has one or more general placement verbs, learners' use of placement verbs in speech may show native-like overall distinctions relatively quickly (Gullberg, 2009a, 2011b; Lewandowski and Özçalişkan, 2021) but can nevertheless include non-native verb forms (Cadierno et al., 2016) and overuse of more peripheral verbs in an attempt to re-create placement distinctions from the L1 (Gullberg, 2009b).

Co-speech gestures have been used as a means to probe L2 speakers' underlying representations in the placement domain (Gullberg, 2009a,b, 2011a,b). While L2 learners of a more general placement verb system (e.g., Dutch L1-French L2) may be able to acquire the verbal system with relatively little difficulty, studying their gestures reveals a somewhat different picture. Using a French native-like pattern in speech does not necessarily mean that the learners have abandoned the conceptualization of placement events from their L1. Specifically, native French speakers primarily accompany their placement verbs with path-only gestures. In contrast, many L1 Dutch learners of L2 French use significantly more figure gestures, maintaining the distinctions from their native Dutch (Gullberg, 2009b), even as they are using a target-like system in their spoken French (see also the study by Özçalişkan, 2016 showing persistent L1 co-speech gesture in L2 expression of voluntary motion). Conversely, L1 speakers of a general placement verb language (e.g., English) learning a specific placement verb L2 (e.g., Dutch) produce mainly English-like path-only gestures in their L2, even when they begin to use the appropriate verb distinctions in speech (Gullberg, 2009a). Importantly, learners of a specific system do not gesture about figure-objects unless they apply the relevant distinctions in speech (Gullberg, 2009b). A general observation is that although it is difficult, and the progress is gradual, semantic reconstruction away from fuzziness and into alignment with the L2 seems possible (Gullberg, 2009b).

While evidence suggests that acquiring semantic distinctions in an L2 is challenging, it is important to note that this evidence is based exclusively on research with same-modality L2 learners, that is, individuals with a spoken L1 learning a spoken L2. However, not all second language learning happens within the same modality. Many hearing individuals acquire a sign language as an L2. Researchers are increasingly asking to what extent the principles of L2 acquisition apply when the source language is spoken and the target language is signed (Chen Pichler and Koulidobrova, 2016). While many challenges are similar for hearing learners of signed and spoken second languages, additional issues arise when acquiring a new language in a new modality (McKee and McKee, 1992; Wilcox and Wilcox, 1997), including learning to manage visual-manual phonology (Bochner et al., 2011; Chen Pichler, 2011; Ortega, 2013; Ortega and Ozyurek, 2013; Ortega and Morgan, 2015), multiple articulators (Gulamani et al., 2020), spatial grammar and depicting referents with the body (Bel et al., 2014; Ferrara and Nilsson, 2017;

[2]Although note that it is at present an empirical question whether classifier incorporation is generally obligatory in placement descriptions.

Frederiksen and Mayberry, 2019; Kurz et al., 2019; Gulamani et al., 2020), using the face to display grammatical information (McIntire and Reilly, 1988), and the high degree of iconicity in sign languages (Lieberth and Gamble, 1991; Campbell et al., 1992; Baus et al., 2013; Ortega and Morgan, 2015). At the same time, it is possible that hearing learners' experience with co-speech gesture in their L1 affects their acquisition of a sign language (McIntire and Reilly, 1988; Taub et al., 2008; Chen Pichler and Koulidobrova, 2016). Hearing individuals produce spontaneous co-speech gestures when they speak. However, it is as of yet unclear whether co-speech gesture in fact helps or hinders sign acquisition, and previous work suggests that the answer to this question may vary by linguistic domain (Schembri et al., 2005; Ortega and Morgan, 2010; Chen Pichler, 2011; Ortega, 2013; Marshall and Morgan, 2015; Janke and Marshall, 2017; Kurz et al., 2019).

To date, it is unknown whether the acquisition of placement expressions is similarly difficult when acquiring a signed compared to a spoken L2. Many researchers have noted similarities between the classifier handshapes used by Deaf signers and the handshapes used in co-speech gesture and pantomime by hearing individuals in (Singleton et al., 1993; Schembri et al., 2005; Sevcikova, 2014; Marshall and Morgan, 2015; Quinto-Pozos and Parrill, 2015), despite obvious differences such as signers tapping into a much more conventionalized system and gesturers employing these handshapes on the fly. Other studies show similarities in how the signers and gesturers alternate between different (classifier) handshapes in their descriptions of objects and humans handling them (Brentari et al., 2012; Padden et al., 2015; Masson-Carro et al., 2016; Hwang et al., 2017; van Nispen, 2017; Ortega, 2020). Thus, it is possible that English speakers can build on their use of gestural distinctions to acquire ASL semantic boundaries in placement verbs relatively quickly. Moreover, the high degree of transparency in ASL placement distinctions may also be an advantage for L2 learners, decreasing the proficiency level required to use target-like placement distinctions in ASL compared to learners acquiring a less transparent system.

Bidirectional Language Influences

Research has shown that language influence can happen in both directions, from L1 to L2 but also from L2 to L1. In bilinguals, the two languages are activated at the same time and compete for selection (e.g., Jared and Kroll, 2001; Marian and Spivey, 2003; Dijkstra, 2005; Kroll et al., 2008). This not only results in influence from the first on the second language; even at the very beginning stages, learning a second language affects the first language (see Kroll et al., 2015). Further, this effect is not only observed with respect to language processing, but also in how events are conceptualized, e.g., "conceptual transfer" (Bylund and Jarvis, 2011; Daller et al., 2011). Bi-directional effects have been observed in the context of cross-modality language learning as well. Work by Morford et al. has shown that ASL signs are activated during English print word recognition in highly proficient ASL-English bilinguals, irrespective of language dominance (Morford et al., 2011, 2014). Similar effects have been reported for DGS (Deutsche Gebärdensprache, *German Sign Language*)-German bimodal bilinguals (Kubus et al., 2015; Hosemann et al., 2020).

L2 influence on placement verb semantic boundaries in L1 has not been researched specifically for the case when the L1 is clearly dominant and the L2 is weaker. However, Alferink and Gullberg (2014) investigated placement verbs in individuals who grew up with and continued to use both French and Dutch in their daily lives. These early bilinguals showed evidence of having blurred obligatory placement verb distinctions in Dutch, effectively using the same distinctions for both Dutch and French despite the former being a specific, multi-term language and the latter being a general, single-term language.

Thus, there is reason to expect L2 learners' placement expressions in English to be influenced by ASL. Such an influence could be evident in either speech or in gesture. Co-speech gesture research has found evidence of gestural transfer from the L2 to the L1 (Brown and Gullberg, 2008, 2011; see also overview in Gullberg, 2009c). Specifically, in L1 descriptions of voluntary motion, some L2 learners show evidence of aligning with the L2 in gesture while maintaining L1 patterns in speech. Pertinent to the present study, there appears to be an additional effect on L1 co-speech gesture from learning a signed as opposed to a spoken L2. Iconic gesture rates increase with sign language proficiency, something that does not happen when learning spoken second languages, even in languages known for frequent gestures, such as French or Italian (Casey et al., 2012; Weisberg et al., 2020). For hearing English-ASL bilinguals, there is the additional factor that classifiers, and particularly classifiers of the *handling* type that are predominant in placement descriptions, have iconic properties reflecting visuo-spatial properties of their referents (Zwitserlood, 2012), which offers a visual correspondence between elements of the world and their linguistic encoding. It has been shown that, in some domains, co-speech gesturers and signers tend to use their hands in similar ways (Sevcikova, 2014; Quinto-Pozos and Parrill, 2015). As such, placement descriptions in sign languages offer a visual correspondence between elements of the world and their linguistic encoding which may be easy to adopt either because it already overlaps with the distinctions used by the learners in co-speech gesture, or because of the high degree of transparency in the distinctions that are being employed in ASL.

THE PRESENT STUDY

The present study investigates what semantic reorganization in the domain of placement events looks like when the source and target languages do not share a sensory-motor modality. We ask two major questions: First, whether second language learners of ASL face similar challenges with placement verbs as do same-modality L2 learners, especially in the light of the high degree of transparency in ASL placement verb distinctions. Second, we ask whether there is evidence that learning new semantic boundaries for placement events affects L1 semantics. If modality and transparency do not matter, then we would expect English L1-ASL L2 language users to use general placement verbs such as "PUT" and "MOVE," and specifically, L2 signers should use classifiers at a lower rate than native signers and exhibit fuzzy semantic boundaries for the classifiers they do use. Further, if L1 semantics is affected by learning a signed L2, then we would expect L2 signers' placement descriptions in English speech and/or co-speech gesture to include ASL-like distinctions that do not occur in non-signing English speakers. Specifically, L2 signers would be expected to use comparatively more verbs with a specific rather than general placement meaning, and/or to use more co-speech gestures reflecting properties of the figure object.

TABLE 1 | The L2 signers' background information.

Participant information	Mean	Range	SD
Age (years)	21	19–22	1
Age began ASL learning (years)	18	13–20	2
Time learning ASL (years)	2.88	1–8	2.30
Self-reported expressive ASL proficiency (of 10)	6.75	4–9	1.49

Methods

Participants

We recruited eight hearing L2 signers (five female) to take part in the task. These individuals were native speakers of English who learned ASL as young adults. All were intermediate learners who had completed at least 1 year of ASL instruction (six weekly contact hours). At the time of participation, seven of the learners used ASL daily, and one learner used ASL once a month. Seven of the L2 learners had exposure to either Spanish or French starting between the ages of nine and fourteen. **Table 1** summarizes their demographic information. We additionally tested eight non-signing English speakers (five female, mean age: 19; SD: 1), and two Deaf, native ASL signers (two female, ages 21 and 62), on the same task. Seven of the non-signers had exposure to a language other than English (Spanish, French, Farsi, German); three were exposed to the non-English language after the age of 11, two at the age of eight, and two were exposed to Spanish from birth [3]. All non-signing participants reported that they were English dominant.

Stimuli and Procedure

We used a director-matcher task (e.g., Clark and Wilkes-Gibbs, 1986) to elicit placement descriptions of our stimuli. The stimuli consisted of a video segment showing a man repositioning objects in a room. The video was split into six parts, each containing the placement of four or five stimulus objects (e.g., a cup, a lamp, plates, a scarf), for a total of 25 events (see Appendix in **Supplementary Material** for a full list of stimulus items). In our version of the director-matcher task, the participant was the director, and their task was to watch the video clips (**Figure 5a**) one at a time and explain to the matcher (a native language user confederate) what happened after each clip. The matcher in turn drew this information on a picture of the empty room **Figure 5b**), specifically where the objects being described were placed. The video clips were not visible to the matcher and the drawing was not visible to the director.

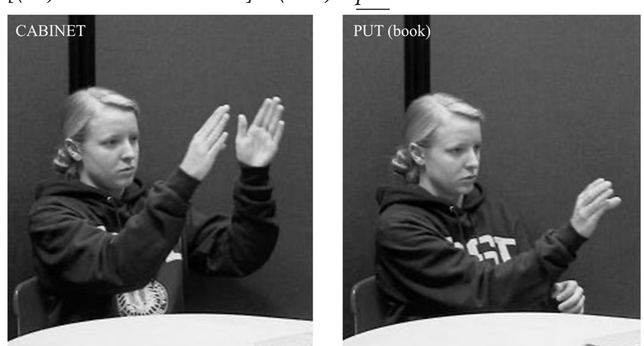
After providing written, informed consent, the director and matcher were seated across from each other, and the experimenter explained their tasks in the language of the experiment. As a memory aid, the director was given list of pictures corresponding to the objects-to-be-described[4]. The matcher was instructed to ask questions whenever clarification was needed. Two cameras captured the communication between the director and the matcher during the task. Only data from the director are analyzed in this study. The L2 signers completed the task twice, once in English and once in ASL with different interlocutors. We counterbalanced the order of languages, and participants performed another task before completing the task the second time. Participants were either paid a small amount or given credit in a college course for their participation.

Transcription and Coding

Speech and sign transcriptions were done in the ELAN software (Wittenburg et al., 2006) by transcribers who were native language users. For each of the stimulus objects, they identified the first complete, spontaneous, and minimal placement description. Such descriptions included mention of the figure object, generally in the form of a lexical noun phrase, the verb (or intransitive construction, e.g., "the cup is on the table"), and often the final location of the object as well. Repetitions, elaborations and answers to questions by the interlocutor were not included. From each of the included descriptions, the transcriber provided a verbatim transcription of the placement event specifically, shown in italics in 1) for English and 2) for ASL.

(1) [then he grabbed the paper towels] *and placed them on the kitchen counter*

(2) [BOOK TAKE-handshape: C] *CABINET PUT-handshape: C [(he) took the book] (and) put it in the cabinet*



Finally, it was noted which placement verb was used in each placement description [underlined in (1) and (2)]. For ASL trials, two trained coders, a Deaf native signer and a hearing proficient signer, noted whether the placement verb was used in the citation form or with an incorporated classifier. The coders were instructed to be conservative when encountering

[3] Spanish is similar to English in using general rather than specific placement verbs. Therefore, we do not expect the simultaneous speech-speech bilinguals to talk about placement differently than other English-speaking non-signers.

[4] Pilot work showed that this list was necessary in order for the participants to remember and describe every placement event. Unlike previous studies (e.g., Gullberg, 2009a), we used a list of pictures rather than written words in an effort to stay consistent between languages, as ASL has no conventionally recognized written form.

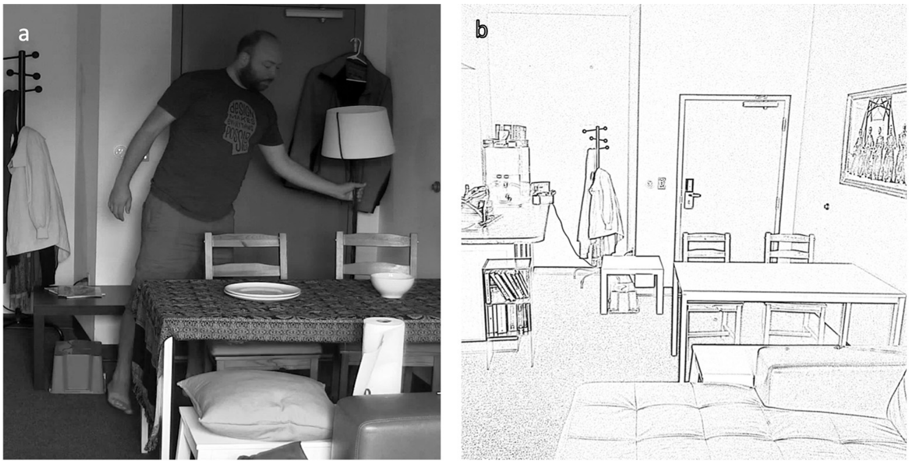


FIGURE 5 | Still picture from a stimulus clip **(a)** and picture for the matcher's drawing **(b)**.

handshapes that were ambiguous between citation form and incorporated classifier. For example, holding a thin flat object in a horizontal position would use the same flat O handshape as the citation form of "PUT." In such cases, the verb was coded as occurring in citation form unless the hand orientation or movement was different from the citation form. Flat and open/round versions of the same handshape were grouped together. Where applicable, the coders also noted the type of classifier (handling vs. entity) used in the placement verb. This resulted in three categorizations of verbs and intransitive constructions with possible classifier incorporation: (1) Handling classifier, (2) Entity classifier, and (3) Citation form. Across the entire data set, the two coders agreed on 92% of categorizations.

For gesture transcriptions, we focused on tokens occurring during the minimal placement descriptions identified for speech. We marked gesture strokes, that is, the most effortful and expressive part of the gesture, and post-stroke holds, that is, periods of maintaining the stroke handshape after the stroke itself (Kita et al., 1998; Kendon, 2004). For each gesture identified, two coders separately noted whether the handshape and hand orientation (a) expressed figure information by reflecting properties of the stimulus object in question, (b) did not reflect figure object properties but only indicated direction or path of movement, (c) reflected only properties of the ground object on which the figure object was placed, or (d) did not have any relationship to the placement event proper, that is, the gesture was a beat gesture serving to emphasize speech rhythm, a thinking gesture indicating word finding difficulty, or it was unclear what the gesture represented. This resulted in the following gesture categories: (1) Figure inclusion (a), (2) Path only (b), (3) Everything else (c and d). To minimize influence on gesture coding from the speech content, this coding was undertaking without access to the video's sound. Across the entire data set, the two coders agreed on 90% of form categorizations. In cases of disagreement, the judgement of the first coder was retained.

RESULTS

The analysis focuses on the target objects for which the placement description was complete. Some trials were skipped, most likely because participants simply forgot to mention a target object. Non-signing participants skipped nine target objects (5%), native ASL signers skipped one target object (2%), and the L2 ASL learners skipped ten (5%) target objects in English and 15 (7.5%) in ASL.

ASL

We first confirmed that the Deaf, native signer control data matched our expectations of ASL as using primarily general verb types like "PUT," and "MOVE" modified with classifiers to reflect properties of the figure object (see verb illustrations in the **Supplementary Images**). As shown in **Table 2**, the verb types "MOVE" and "PUT" together accounted for 88% of verb tokens used. "HANG" was the only additional verb used with any regularity. The native signers' mean proportion of verbs incorporating a classifier handshape was 66%. The vast majority of classifiers were handling classifiers (84%); only a small proportion were entity classifiers (16%). **Table 3** shows how often different verbs were produced with classifiers. On average, "MOVE" was used with a classifier more than half the time. 92% of those classifiers were handling classifiers. The verb "PUT" never occurred in citation form in these data; it was used with a handling classifier 94% of the time. Overall, seven different classifiers occurred: three handshapes occurred as handling classifiers (O, C, S) and four occurred as entity classifiers (Y, B, C, baby-C; see illustrations of classifier handshapes in the **Supplementary Images**). Thus, while a few non-specific placement verbs account for the vast majority of tokens in native ASL signers' placement descriptions, they are modified with classifiers more often than not, which creates a complex system involving multiple distinctions based on properties of the figure object.

TABLE 2 | Verb types by group.

	Native signers		L2 signers	
	Mean% (*N*)	**SD**	**Mean% (*N*)**	**SD**
MOVE	0.67 (33)	0.24	0.35 (64)	0.29
PUT	0.21 (10)	0.24	0.32 (59)	0.13
HANG	0.08 (4)	0.06	0.08 (15)	0.05
DRAPE	0.02 (1)	0.03	0.02 (3)	0.03
Intrans.	0.02 (1)	0.03	0.18 (33)	0.21
SET	0.00 (0)	-	0.06 (10)	0.09

TABLE 3 | Classifier incorporation by verb and group[5].

	Native signers		L2 signers	
	Mean% (*N*)	**SD**	**Mean% (*N*)**	**SD**
MOVE	0.62 (20)	0.07	0.65 (28)	0.45
PUT	1.00 (10)	0.00	0.87 (54)	0.23
HANG	0.17 (1)	0.24	0.92 (14)	0.20
DRAPE	1.00 (1)	-	1.00 (3)	0.00
Intrans.	NA	NA	0.09 (5)	0.18

We next asked whether the L2 signers similarly used general placement verbs modified with classifiers in their placement descriptions, which would suggest they that have reconstructed their conceptualization of placement events and are using semantic distinctions that are relevant in the target language.

An examination of how the L2 learners' use of broad ASL verb types compares with that of the two Deaf, native signers, irrespective of classifier use, shows that the L2 signers use "MOVE" and "PUT" most frequently, similarly to the native signers (**Table 2**). However, where the native signers use a greater proportion of "MOVE," L2 signers on average use the two verbs at a similar rate. The main difference between the groups' overall verb use is the L2 learners' use of a large proportion of intransitive descriptions, and their use of "SET," which are low or absent from the native data[6]. Both groups use comparable proportions of "HANG" and "DRAPE." Despite differences in proportion, an analysis of variance on the proportion of tokens (arcsine transformed) as a function of verb type and group showed no difference in the use of broad ASL verb types by the L2 learners as compared to the native signers [$F_{(6,48)} = 1.656, p = 0.153$].

We next asked whether L2 and native signers incorporated classifiers into their verbs at similar rates. Aggregating across verbs, the mean rate of classifier incorporation was 56% (SD = 27) for L2 signers, compared with 66% (SD = 13) for native signers. Nevertheless, a mixed-effect logistic regression (Jaeger, 2008) showed no significant effect of group ($ß = -0.329$, $p = 0.77$)[7]. **Table 3** shows how often classifiers were incorporated into the different verbs by native and L2 signers. Overall, the groups incorporated classifiers similarly. On average, L2 signers used a high rate of classifiers for both "MOVE" and "PUT," which is similar to the native signers. The main difference was in the verb "HANG," where L2 signers incorporated classifiers at a much higher rate than native signers. As for classifier type, the L2 signers were similar to the native signers in using mostly handling classifiers (79%), with fewer entity classifiers (21%). The learners used both "MOVE" and "PUT" primarily with handling classifiers. Overall, nine classifiers occurred in the L2 data: four handshapes were used as entity classifiers (C, B, baby-C, F) and five were used as handing classifiers (A, 5, C, O, S). The native signers used the handshapes C, B, and baby-C as well as Y as entity classifiers. The two groups overlapped in their use of the handling classifier handshapes C, O and S, and the L2 signers additionally used A and 5 handshapes.

We finally asked whether the classifiers used by the L2 signers were appropriate for the described objects, as compared with the native signers. We grouped the described objects into five categories based on shared characteristics that were expected to influence how placement of the object would be expressed: (1) Tall cylindrical objects with a thin handle or neck (wine bottle, water bottle, paper towel holder, lamp, and candle), (2) Objects with a functional base (glass, potted plant, plates, speaker, basket, and bowl), (3) Thin rectangular objects without clear functional base (picture frame, magazine, computer, and book), (4) Object made from fabric (jacket, table cloth, pillow, scarf, throw, bag, and hat), and (5) Other (cables, silverware, clothes hanger)[8]. We then examined which handshapes were preferred by native and L2 signers for objects in each of the five categories (**Table 4**). Visual inspection of **Table 4** shows that, excepting the Fabric category, the L2 signers used more verbs without classifiers than the native signers for all categories[9]. The most frequently occurring classifier was the same in both groups for the categories Tall Cylinder, Functional Base and Thin Rectangle. At the same time, the L2 signers used the preferred classifier in each group at a numerically lower rate than the native signers. Moreover, the L2 signers' use of classifiers was more variable than the native signers' across all categories.

We asked whether the L2 signers' classifier handshapes were appropriate for the figure objects in question. A Deaf, native ASL signer rated the appropriateness of classifier handshapes produced by each signer for each target object on a scale from 1 to 5. A rating of 1 corresponded to "very bad ASL," and a rating of

[5] "*SET*" is excluded from the table. Its citation form is made with an A-handshape that is sometimes considered a classifier and can be used to represent placement or location of objects, but this handshape does not alternate with other classifiers. We opted for a conservative analysis that does not treat instances of "*SET*" as containing a classifier that represents the figure object. However, an analysis that includes all instances of "*SET*" as having classifiers also does not result in any between-group differences.

[6] ASL has no copula. Intransitives are descriptions such as "PLATE ON TABLE," "the plates are on the table".

[7] Mixed effects logistic regression models were fit in R Core Team (2014) using the lme4 package (Bates et al., 2014). Models included random effects of subjects and items. Coefficient estimates, and *p*-values based on the Wald Z statistic are reported.

[8] Excepting the Fabric and Other categories, these categories largely correspond to the object shapes discussed by Zwitserlood (2012).

[9] For the Fabric category, the native signers used a large number of handshapes that were appropriate for showing how fabric items are handled but these handshapes were indistinguishable from citation forms and were consequently not counted as classifiers, as explained in the methods.

TABLE 4 | Classifier types by group and object type.

		Native signers % (N)	L2 signers % (N)
Tall cylinder	Handling		
	S	1.00 (10)	0.256 (10)
	C	0.00 (0)	0.103 (4)
	A	0.00 (0)	0.077 (3)
	O	0.00 (0)	0.026 (1)
	Entity		
	B	0.00 (0)	0.026 (1)
	F	0.00 (0)	0.026 (1)
	SET	0.00 (0)	0.179 (7)
	No classifier	0.00 (0)	0.308 (12)
Functional base	Handling		
	C	0.583 (7)	0.261 (12)
	S	0.00 (0)	0.043 (2)
	5	0.00 (0)	0.022 (1)
	O	0.00 (0)	0.022 (1)
	A	0.00 (0)	0.022 (1)
	Entity		
	Baby-C	0.083 (1)	0.13 (6)
	No classifier	0.333 (4)	0.50 (23)
Thin rectangle	Handling		
	C	0.25 (2)	0.167 (5)
	O	0.25 (2)	0.033 (1)
	A	0.00 (0)	0.067 (2)
	Entity		
	B	0.125 (1)	0.133 (4)
	Baby-C	0.00 (0)	0.033 (1)
	SET	0.00 (0)	0.100 (3)
	No classifier	0.375 (3)	0.467 (14)
Fabric	Handling		
	O	0.154 (2)	0.08 (4)
	S	0.077 (1)	0.08 (4)
	A	0.00 (0)	0.40 (20)
	C	0.00 (0)	0.08 (4)
	5	0.00 (0)	0.02 (1)
	Entity		
	Y	0.077 (1)	0.00 (0)
	C	0.077 (1)	0.06 (3)
	B	0.00 (0)	0.02 (1)
	No classifier	0.615 (8)	0.26 (13)
Other	Handling		
	O	0.167 (1)	0.00 (0)
	S	0.333 (2)	0.10 (2)
	A	0.00 (0)	0.35 (7)
	5	0.00 (0)	0.15 (3)
	Entity		
	Y	0.167 (1)	0.00 (0)
	No classifier	0.333 (2)	0.40 (8)

5 corresponded to "very good ASL." Submitting these ratings to an ordinal mixed effects model[10] revealed a significant difference between groups (L.R. $\chi^2 = 4.4711$, $df = 1$, $p < 0.05$), with native signers receiving higher ratings (M = 4.78, SD = 0.04) than the L2 signers (M = 3.19, SD = 1.05).

English

English Verbs

To assess whether the L2 signers experience influence from ASL on their L1, English, we first asked whether they use English placement verbs similarly to non-signers. As shown in **Table 5**, the most frequently used verb for English was the general verb "put" for both non-signers (69%) and L2 signers (63%). The general placement verbs, "move" and "place" were the also among the most frequently used, and two general verbs "be" and "bring" were used infrequently. Four specific verbs occurred: "hang," "drape," "spread" (used only by non-signers), and "stick" (used only by L2 signers). Two specific posture verbs, "set," and "lay" also occurred in the data. "Set" occurred in both groups and "lay" only occurred in the L2 signer group. Thus, the specificity in verbal expression is very similar across non-signers and L2 signers. Overall, the native English speakers (non-signers and L2 signers) exhibited a pattern of verb use that is congruent with previous research, namely by preferring non-specific placement verbs ("put," "move," and "place") at a mean rate of 85% for both the non-signers and L2 signers, and by specifically preferring "put." A mixed effects logistic regression analysis of verb type (general vs. specific) revealed no difference between groups ($\beta = 0.085$, $p = 0.866$).

English Co-speech Gesture

We next analyzed the co-speech gestures in the English data from the non-signers and the L2 signers. First, we asked whether exposure to ASL led the L2 signers to produce more iconic gestures (including gestures representing the Figure object, Path only, and Ground only, see Methods) in English compared to the non-signers. As a group, the non-signers produced a total of 185 gestures (M = 24.71, SD = 5.84) during their placement descriptions and the L2 signers produced a total of 121 gestures (M = 15.13, SD = 13.81). Thus, it was not the case that the L2 signers gestured more than the non-signers. A Poisson regression modeling number of gestures as a function of group showed no significant difference between the groups (L.R. $\chi^2 = 1.915$, $df = 1$, $p = 0.1881$)[11].

Next, we asked whether there is evidence of bidirectional transfer in the L2 signers' English co-speech gestures. It is possible that the L2 signers' semantic categories have realigned with ASL. This is because they have acquired from the ASL system the use of classifiers to make distinctions between different placement events (even if they do not always use the appropriate handshape). There was no evidence of realignment in speech. However, as discussed, English has only a few specific placement verbs. This, together with the fact that the specific placement verbs have low frequencies makes it unlikely that L2 signers would use them as a prominent part of their inventory of placement expressions. For this reason, speech data alone may

[10] Ordinal mixed effects models were fit in R Core Team (2014) using the packages Ordinal (Christensen, 2019), Car (Fox and Weisberg, 2019), and RVAdeMemoire (Hervé, 2021). Models included random effects of subjects and items.

[11] We used a generalized linear model for this regression. Because over-dispersion was indicated, we opted to use a quasi-poisson regression model.

TABLE 5 | Verb frequencies in English by non-signers and L2 signers.

	Verb type	Non-signers		L2 signers	
		Mean% (*N*)	SD	Mean% (*N*)	SD
Put	General	0.69 (129)	0.17	0.63 (120)	0.24
Move	General	0.09 (16)	0.04	0.14 (27)	0.18
Place	General	0.07 (13)	0.18	0.08 (15)	0.12
Hang	Specific	0.07 (14)	0.04	0.06 (11)	0.04
Set	Specific	0.06 (11)	0.09	0.04 (8)	0.08
Be	General	0.02 (3)	0.04	0.00 (0)	-
Drape	Specific	0.01 (2)	0.03	0.02 (3)	0.03
Bring	General	0.01(2)	0.02	0.00 (0)	-
Spread	Specific	0.01 (1)	0.01	0.00 (0)	-
Stick	Specific	0.00 (0)	-	0.02 (4)	0.04
Lay	Specific	0.00 (0)	-	0.01 (2)	0.02

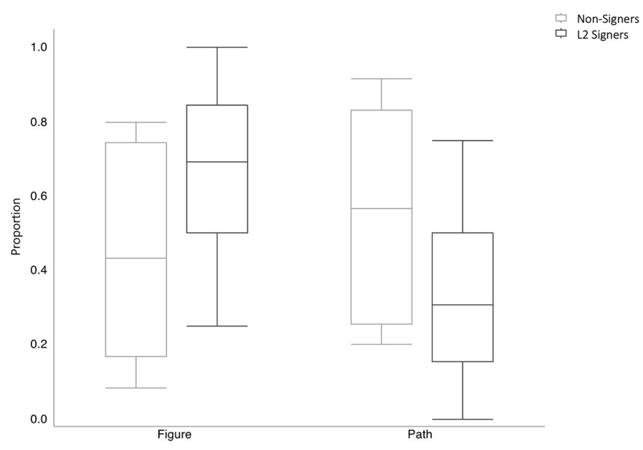


FIGURE 6 | Proportion gesture type in English by group.

not accurately reflect the state of the L2 signers' semantic organization, but English co-speech gesture data may.

If there is bidirectional transfer, then the co-speech gestures of L2 signers should look different than those of non-signing English speakers. Specifically, we expect L2 signers to produce more gestures incorporating figure object information and less path-only gestures compared to non-signers. To assess whether this was the case, we focused on gestures that expressed information about Path only vs. about Figure. One L2 signer produced no Figure or Path gestures and was excluded from this analysis. We compared non-signers' and L2 signers' use of these gesture types (**Figure 6**). A mixed effects logistic regression analysis of gesture type as a function of group revealed a marginally significant effect of group ($ß = -1.256$, $p = 0.065$), with L2 signers' producing numerically more gestures about figure than path-only gestures (67% vs. 33%, SD = 26) compared to the non-signers (46% vs. 54% SD = 28).

Given that the L2 signers frequently used classifiers in their ASL descriptions, it is possible that their distribution of gesture types is affected by ordering effects in the experiment. Specifically, half the L2 signers completed the task in ASL before completing it in English. Thus, the L2 signers' higher rate of figure gestures could be a result of priming from ASL. We compared the distribution of gesture types between the L2 signers who did the task in ASL first and those who did the task in English first. If the higher rate of figure gestures in the bilingual group as a whole is driven by priming effects in the ASL-first participants, then we should see higher rates of figure gestures in that group compared to the English-first group. This was not the case. The mean rate of figure gestures was 54% (SD = 24) for the ASL-first group ($N = 4$) and 85% (SD = 15) for the English-first group ($N = 3$). A mixed effects logistic regression analysis revealed no significant difference between those L2 signers who completed the task in English first and those who completed the task in ASL first ($ß = -1.012$, $p = 0.184$).

DISCUSSION

The present study examined placement descriptions by second language (L2) learners of American Sign Language (ASL) and asked two questions, namely (1) whether learning the semantics of placement descriptions in a typologically different L2 in a different modality presents a challenge similar to when it occurs within the same modality, and (2) whether cross-modal L2 learners show evidence of semantic reorganization in placement descriptions in their first language (L1).

We found that the hearing L2 ASL signers used verb types similarly to Deaf, native signers and even incorporated classifiers at a comparable, if still somewhat lower, rate. Based on previous research, we would predict that the L2 signers in the present study should have problems acquiring the ASL placement verb system, given its complexity compared to English. Specifically, we expected non-target like use of classifiers suggesting fuzzy semantic boundaries. It is then perhaps surprising to find that the hearing ASL learners were well on their way to acquiring native-like placement descriptions. This is not to say that L2 signers are fully target-like in their ASL use. Classifier handshape ratings from a Deaf, native signer were lower for the L2 signers than for native signers, suggesting that the L2 learners were less ASL appropriate in their handshape selection. Moreover, the L2 signers used a wider variety of handshapes than the native signers, which is in line with previous findings suggesting that L2 signers tend to struggle with selecting target-like handshapes for objects and make distinctions that are too fine-grained (Schembri et al., 2005; Brentari et al., 2012; Marshall and Morgan, 2015; Janke and Marshall, 2017). Finally, the present study focuses on the semantics of placement verbs. It is likely that an investigation of the syntactic constructions and pragmatic contexts in which placement verbs participate would reveal additional differences between L2 learners and native signers.

Nevertheless, the L2 signers in the present study performed unexpectedly similar to Deaf, native signers with respect to including figure information in placement verbs. This is so especially in light of the learners' limited ASL experience, which was <3 years on average (mean 2.88 years, range 1–8 years). By comparison, English learners of Dutch who were residing in the Netherlands and therefore immersed in the language and culture for several years on average (mean length of residence: 11 years, range: 4 months to 19 years) show substantial problems with Dutch placement verbs (Gullberg, 2009a). However, this difference is not necessarily attributable to the difference in modality *per se*. First, it is a limitation of the present study that we only had two Deaf signers in the control group for ASL. With additional signers, clearer patterns in classifier preference for placement of different object types could emerge and possibly a statistically significant difference in classifier rate between groups. The ASL L2 learners' relative success could also be due to the higher semantic transparency in ASL verbs compared to a language like Dutch. For example, in a study of placement verb acquisition in Tamil and Dutch speaking children, Narasimhan and Gullberg (2011) found that Tamil children acquire relatively infrequent placement verbs (specifically caused posture verbs) early. They attribute this to the semantic transparency of Tamil verbs, which consist of multi-morphemic units such as "make stand" that "individually label the causal and result subevents" (2011, p. 504). By comparison, Dutch caused posture verbs are highly frequent but are also monomorphemic and much less transparent. ASL placement verbs are highly transparent, consisting of a root verb movement (e.g., "MOVE") combined with a classifier representing either the handling or the shape and size of an object. It is possible that ASL learners are capitalizing on this transparency.

Another possibility is that L2 signers are benefitting from the fact that there is some overlap between the distinctions used in English co-speech gesture and in ASL. Specifically, even though native English speakers predominantly gesture about path in their English, they also gesture about the figure object at a non-negligible rate (46% in the present study, and 40% in Hoetjes, 2008). This overlap may help L2 signers reach the rate of 56% incorporation of classifiers expressing figure object information in ASL observed in the present study sooner than expected given their proficiency.

Regardless of the underlying reason, the results of the present study suggest that ASL learners successfully begin to reorganize their placement verbs semantics in the context of using ASL. It then becomes an important question whether this reorganization happens independent of their English semantics, that is whether their original system is still intact or whether the ability to express new and additional placement distinctions in ASL results in more far-reaching semantic changes. Previous work has found that same-modality learners can exhibit prolonged maintenance of their native placement distinctions, to the extent that L1-like patterns can be observed in L2 gesture, even when speech has become target-like (Gullberg, 2009a; Hoetjes, 2018). This suggests a persistence of L1 semantic organization for the purposes of speaking in the L2. In the case of ASL L2 acquisition, it is not possible to directly examine L2 gestures co-occurring with a main (verbal) expressive channel to assess whether there is evidence of maintenance of L1 semantics. However, the fact that L2 signers included figure information at a rate comparable to Deaf native signers along with the evidence of bidirectional transfer from the L2 to the L1 in the L2 signers' co-speech gesture pattern suggests limited persistence of L1 semantic patterns.

Unlike previous work (Weisberg et al., 2020), we did not find the cross-modal L2 learners in this study to use more iconic gestures than non-signers. This result is possibly an effect of proficiency and length of signing experience. In the present study, the L2 signers had around 3 years of sign experience on average. The L2 signers in the study by Weisberg and colleagues had 10 or more years of exposure to ASL, suggesting that increased gesturing may occur with increased exposure to and proficiency in ASL. In support of this hypothesis, Casey et al. (2012) found no clear increase in co-speech gesture frequency in L2 learners after 1 year of academic ASL instruction. While they found a numerical increase in gesture rate in the different-modality learners that was absent from the control group of same-modality language learners, the learner groups did not differ statistically from each other in terms of gesture rate. In the present study, however, we found the non-signers to use numerically (although not statistically significantly) more iconic gestures than the L2 learners. This difference could be due to the focus in the present paper on analyzing only the subset of participants' utterances that pertained to how an object was placed. Such utterances occur in specific discourse contexts, where the figure object is typically known information. Previous work has shown discourse context to affect gesture rate in non-signers (Debreslioska and Gullberg, 2020). We leave it for future research to determine whether L2 signers' and non-signers' gesture rates vary in a similar or different manner as a function of discourse context.

However, gesture rate was not the only domain we examined for bidirectional language transfer. In assessing the type of gestures used, we found that L2 signers gestured about the figure object (numerically) more than native English speaking non-signers, although statistically, the difference was only a trend. Nevertheless, the results suggest that on average the L2 signers have begun diverging from the monolingual pattern of gesturing nearly equally about Figure and Path and are instead favoring gestures about the figure object. The results here suggest that L2 learners' original system is not intact. This is remarkable given their profile as adult L2 learners of intermediate proficiency, with an average of <3 years of experience. This finding therefore raises questions about how gesture patterns compare in the native language of monolinguals and bilinguals when L2 learning happens within modalities. While previous work on placement descriptions has mainly focused on performance in the L2, one analysis of L2 learners' gesture patterns in the L1 replicated results for monolinguals in the same language (compare Hoetjes, 2008; Gullberg, 2009a). While more work is needed to confirm these patterns, comparing the L1 results of the current study to those of previous work suggests that semantic boundaries may be differentially affected by the L2 early on in acquisition in same- vs. different modality learning. In this context, it will be especially important to compare the present findings to situations in which both the L1 and the L2 are spoken and the L2 has specific placement verbs with transparent semantics in order to assess the role of a high degree of semantic transparency vs. the modality of the second language.

CONCLUSION

This study examined object placement event descriptions by English L1-ASL L2 language users, asking whether cross-modal L2 learners face similar challenges to same modality L2 learners in learning to talk about placement. Placement verbs in signed languages such as ASL tend to be highly iconic and to exhibit transparent semantic boundaries which could facilitate their acquisition. We also asked how exposure to a typologically different language affects different semantic boundaries in placement events in the L1. Overall, L2 signers used ASL placement descriptions that looked similar to the Deaf, native signers', despite using a wider range of classifier handshapes and using them less appropriately, indicating somewhat fuzzy and less target-like boundaries in their placement semantics. Moreover, the L2 signers' English co-speech gesture patterns suggest that learning ASL may affect conceptualization of placement in the L1. Specifically, the placement distinctions expressed in co-speech gesture by the L2 signers were marginally more ASL-like compared to non-signers' gestures. Taken together, these results suggest that the iconicity and transparency of placement distinctions in the visual modality may facilitate semantic reconstruction in the placement domain, leading to increased target-like use of placement distinctions in the L2 as well as L1 placement distinctions that may differ from those of non-signers with the same first language.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Research Protections Program, UC San Diego. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AF conceived and designed the study, conducted analyses, and wrote the manuscript. Data collection and coding was carried out with the help of research assistants.

ACKNOWLEDGMENTS

This manuscript would not have been possible without the signers and speakers who took part. Thank you to research assistants DeAnna Suitt, Samantha Moreno, Kalvin Morales, and Matthew Sampson for help with running the experiment, annotation, and coding. Thank you to Monica Keller, a native signer and linguist, for help with ASL judgments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.686485/full#supplementary-material

1. ASL verbs and intransitive constructions

Supplementary Image 1 | MOVE.

Supplementary Image 2 | PUT.

Supplementary Image 3 | HANG.

Supplementary Image 4 | DRAPE.

Supplementary Image 5 | Intransitive construction without classifier: the bag is on the door.

Supplementary Image 6 | Intransitive construction with classifier: the magazine is on the table.

Supplementary Image 7 | SET.

2. Classifiers

Supplementary Image 8 | 5 handshape (handling).

Supplementary Image 9 | A handshape (handling).

Supplementary Image 10 | C handshape (handling).

Supplementary Image 11 | C handshape, flat (handling).

Supplementary Image 12 | O handshape (handling).

Supplementary Image 13 | S handshape (handling).

Supplementary Image 14 | B handshape (entity).

Supplementary Image 15 | Y handshape (entity).

Supplementary Image 16 | F Handshape (entity).

Supplementary Image 17 | C handshape, flat (entity).

Supplementary Image 18 | baby-C handshape (entity).

REFERENCES

Alferink, I., and Gullberg, M. (2014). French–Dutch bilinguals do not maintain obligatory semantic distinctions: evidence from placement verbs. *Bilingual. Langu. Cogn*. 17, 22–37. doi: 10.1017/S136672891300028X

Ameka, F. K., and Levinson, S. C. (2007). Introduction: The typology and semantics of locative predicates: posturals, positionals, and other beasts. *Linguistics* 45, 847–872. doi: 10.1515/LING.2007.025

Aronoff, M., Meir, I., Padden, C., and Sandler, W. (2003). "Classifier constructions and morphology in two sign languages," in *Perspectives on Classifier*

Constructions in Sign Languages, ed K. Emmorey (Mahwah, NJ: Psychology Press), 53–84.

Aronoff, M., Meir, I., and Sandler, W. (2005). The paradox of sign language morphology. *Language*. 81:301. doi: 10.1353/LAN.2005.0043

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4.* (R package version 1.0-6). Available online at: http://CRAN.R-project.org/package=lme4 [Computer software].

Baus, C., Carreiras, M., and Emmorey, K. (2013). When does iconicity in sign language matter? *Langu. Cogn. Proc*. 28, 261–271. doi: 10.1080/01690965.2011.620374

Bel, A., Ortells, M., and Morgan, G. (2014). Reference control in the narratives of adult sign language learners. *Int. J. Bilingu*. 19, 608–624. doi: 10.1177/1367006914527186

Bochner, J. H., Christie, K., Hauser, P. C., and Searls, J. M. (2011). When is a difference really different? Learners' discrimination of linguistic contrasts in american sign language. *Lang. Learn.* 26, 1302–1327. doi: 10.1111/j.1467-9922.2011.00671.x

Bohnemeyer, J., and Pederson, E. (Eds.). (2011). *Event Representation in Language and Cognition*. Cambridge: Cambridge University Press.

Brentari, D., Coppola, M., Mazzoni, L., and Goldin-Meadow, S. (2012). When does a system become phonological? Handshape production in gesturers, signers, and homesigners. *Natur. Lang. Lingu. Theory*. 30, 1–31. doi: 10.1007/s11049-011-9145-1

Brown, A., and Gullberg, M. (2008). Bidirectional crosslinguistic influence in L1-L2 encoding of manner in speech and gesture: a study of Japanese speakers of English. *Studies Sec. Lang. Acqu*. 30, 225–251. doi: 10.1017/S0272263108080327

Brown, A., and Gullberg, M. (2011). Bidirectional cross-linguistic influence in event conceptualization? Expressions of Path among Japanese learners of English*. *Bilingualism* 14, 79–94. doi: 10.1017/S1366728910000064

Bylund, E., and Jarvis, S. (2011). L2 effects on L1 event conceptualization*. *Bilingualism* 14, 47–59. doi: 10.1017/S1366728910000180

Cadierno, T., Ibarretxe-Antuñano, I., and Hijazo-Gascón, A. (2016). Semantic categorization of placement verbs in L1 and L2 Danish and Spanish: placement verbs in L1 and L2 Danish. *Lang. Learn*. 66, 191–223. doi: 10.1111/lang.12153

Campbell, R., Martin, P., and White, T. (1992). Forced choice recognition of sign in novice learners of British Sign Language. *Appl. Lingu*. 13, 185–201.

Casey, S., Emmorey, K., and Larrabee, H. (2012). The effects of learning American Sign Language on co-speech gesture. *Bilingualism* 15, 677–686. doi: 10.1017/S1366728911000575

Chen Pichler, D. (2011). "Sources of handshape error in first-time signers of ASL," in *Deaf Around the World: The Impact of Language*, eds G. Mathur and D. J. Napoli (Oxford: Oxford University Press), 96–121.

Chen Pichler, D., and Koulidobrova, E. (2016). "Acquisition of sign language as a second language (L2)," in *The Oxford Handbook of Deaf Studies in Language: Research, Policy, and Practice*, eds M. Marschark and P. E. Spencer (Oxford: Oxford University Press), 218–230.

Christensen, R. H. B. (2019). *ordinal-Regression Models for Ordinal Data*. R package version 2019.12-10. Available online at: https://CRAN.R-project.org/package=ordinal

Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7

Daller, M. H., Treffers-Daller, J., and Furman, R. (2011). Transfer of conceptualization patterns in bilinguals: the construal of motion events in Turkish and German*. *Bilingualism* 14, 95–119. doi: 10.1017/S1366728910000106

De Knop, S. (2016). German causative events with placement verbs. *Lege Artis*. 1, 75–115. doi: 10.1515/lart-2016-0002

Debreslioska, S., and Gullberg, M. (2020). What's New? Gestures accompany inferable rather than brand-new referents in discourse. *Front. Psych*. 11:1935. doi: 10.3389/fpsyg.2020.01935

Dijkstra, T. (2005). "Bilingual visual word recognition and lexical access," in *Handbook of Bilingualism: Psycholinguistic Approaches*, eds J. F. Kroll and A. M. B. de Groot (Oxford: Oxford University Press), 179–201.

Ferrara, L., and Nilsson, A.-L. (2017). Describing spatial layouts as an L2M2 signed language learner. *Sign Langu. Lingu*. 1, 1–26. doi: 10.1075/sll.20.1.01fer

Fox, J., and Weisberg, S. (2019). *An {R} Companion to Applied Regression, Third Edition*. Thousand Oaks CA: Sage. Available online at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Frederiksen, A. T., and Mayberry, R. I. (2019). Reference tracking in early stages of different modality L2 acquisition: Limited over-explicitness in novice ASL signers' referring expressions. *Sec. Lang. Res*. 35, 253–283. doi: 10.1177/0267658317750220

Gulamani, S., Marshall, C., and Morgan, G. (2020). The challenges of viewpoint-taking when learning a sign language: data from the 'frog story' in British Sign Language. *Sec. Langu. Res.* 1–33. doi: 10.1177/0267658320906855

Gullberg, M. (2009a). Reconstructing verb meaning in a second language: how English speakers of L2 Dutch talk and gesture about placement. *Annu. Rev. Cogn. Lingu*. 7, 221–244. doi: 10.1075/arcl.7.09gul

Gullberg, M. (2009b). Gestures and the development of semantic representations in first and second language acquisition. *Acquis. Int. En Lang. Étrang. Aile... Lia* 1, 117–139. doi: 10.4000/aile.4514

Gullberg, M. (2009c). "Why gestures are relevant to the bilingual lexicon," in *The Bilingual Mental Lexicon*, ed A. Pavlenko (Bristol; Buffalo, NY: Multilingual Matters), 161–184.

Gullberg, M. (2011a). "Language-specific encoding of placement events in gestures," in *Event Representation in Language and Cognition*, eds B. Jürgen and P. Eric (Bristol; Buffalo, NY), 166–188.

Gullberg, M. (2011b). "Thinking, speaking, and gesturing about motion in more than one language," in *Thinking and Speaking in Two Languages*, ed A. Pavlenko (Bristol; Buffalo, NY: Multilingual Matters), 143–169.

Hervé, M. (2021). *RVAideMemoire: Testing and Plotting Procedures for Biostatistics.* R package version 0.9-79. Available online at: https://CRAN.R-project.org/package=RVAideMemoire

Hoetjes, M. (2008). "The use of gestures in placement events," in *Leiden Working Papers in Linguistics 2.3, Vol*. 5 (Leiden), 24–35.

Hoetjes, M. (2018). "Transfer in gesture: l2 placement event descriptions," in *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (Madison, WI), 1791–1796.

Hosemann, J., Mani, N., Herrmann, A., Steinbach, M., and Altvater-Mackensen, N. (2020). Signs activate their written word translation in deaf adults: an ERP study on cross-modal co-activation in German Sign Language. *Glossa J. Gen. Lingu*. 5:57. doi: 10.5334/gjgl.1014

Hwang, S.-O., Tomita, N., Morgan, H., Ergin, R., ILkbaşaran, D., Seegers, S., et al. (2017). Of the body and the hands: patterned iconicity for semantic categories. *Langu. Cogn*. 9, 573–602. doi: 10.1017/langcog.2016.28

Ijaz, I. H. (1986). Linguistic and cognitive determinants of lexical acquisition in a second language. *Lang. Learn*. 36, 401–451. doi: 10.1111/j.1467-1770.1986.tb01034.x

Jaeger, T. F. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang*. 59, 434–446. doi: 10.1016/j.jml.2007.11.007

Janke, V., and Marshall, C. (2017). Using the hands to represent objects in space: gesture as a substrate for signed language acquisition. *Front. Psych*. 8:13. doi: 10.3389/fpsyg.2017.02007

Jared, D., and Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *J. Mem. Lang*. 44, 2–31. doi: 10.1006/jmla.2000.2747

Kellerman, E. (1995). Crosslinguistic influence: transfer to nowhere? *Ann. Rev. Appl. Lingu*. 15, 125–150. doi: 10.1017/S0267190500002658

Kendon, A. (1980). "Gesticulation and speech: Two aspects of the process of utterance," in *The Relationship of Verbal and Nonverbal Communication*, ed M. R. Key (The Hague; New York, NY: Mouton), 207–227.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kita, S., van Gijn, I., and van der Hulst, H. (1998). "Movement phases in signs and co-speech gestures, and their transcription by human coders," in *Gesture and Sign Language in Human-Computer Interaction*, eds I. Wachsmuth and M. Fröhlich (Vol. 1371) (Berlin; Heidelberg: Springer), 23–35. doi: 10.1007/BFb0052986

Kopecka, A., and Narasimhan, B. (2012). *Events of Putting and Taking: A Crosslinguistic Perspective*. John Benjamins Publishing.

Kroll, J. F., Bobb, S. C., Misra, M., and Guo, T. (2008). Language selection in bilingual speech: evidence for inhibitory processes. *Acta Psychol*. 128, 416–430. doi: 10.1016/j.actpsy.2008.02.001

Kroll, J. F., Dussias, P. E., Bice, K., and Perrotti, L. (2015). Bilingualism, mind, and brain. *Annu. Rev. Lingu*. 1, 377–394. doi: 10.1146/annurev-linguist-030514-124937

Kubus, O., Villwock, A., Morford, J. P., and Rathmann, C. (2015). Word recognition in deaf readers: cross-language activation of

German Sign Language and German. *Appl. Psychol*. 36, 831–854. doi: 10.1017/S0142716413000520

Kurz, K. B., Mullaney, K., and Occhino, C. (2019). Constructed Action in American sign language: a look at second language learners in a second modality. *Languages*. 4:90. doi: 10.3390/languages4040090

Lewandowski, W., and Özçalişkan, S. (2021). The specificity of event expression in first language influences expression of object placement events in second language. *Studies Sec. Lang. Acqu*. 1–32. doi: 10.1017/S0272263121000048

Lieberth, A. K., and Gamble, M. E. B. (1991). The role of iconicity in sign language learning by hearing adults. *J. Commun. Disord*. 24, 89–99.

Marian, V., and Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Appl. Psychol*. 24, 173–193. doi: 10.1017/S0142716403000092

Marshall, C. R., and Morgan, G. (2015). From gesture to sign language: conventionalization of classifier constructions by adult hearing learners of British Sign Language. *Top. Cogn. Sci*. 7, 61–80. doi: 10.1111/tops.12118

Masson-Carro, I., Goudbeek, M., and Krahmer, E. (2016). Can you handle this? The impact of object affordances on how co-speech gestures are produced. *Lang. Cogn. Neurosci.* 12, 430–440. doi: 10.1080/23273798.2015.1108448

McIntire, M. L., and Reilly, J. S. (1988). *Nonmanual Behaviors in L1 and L2 Learners of American Sign Language*. Sign Language Studies. 61, 351–375.

McKee, R. L., and McKee, D. (1992). What's So Hard about Learning ASL?: Students'& teachers' perceptions. *Sign Lang. Studies*. 75, 129–158.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought.* Chicago, IL: University of Chicago Press.

McNeill, D., and Levy, E. (1982). "Conceptual representations in language activity and gesture," in *Speech, Place, and Action*, ed R. J. Jarvella and W. Klein (New York, NY: Wiley). p. 271–295.

Morford, J. P., Kroll, J. F., Piñar, P., and Wilkinson, E. (2014). Bilingual word recognition in deaf and hearing signers: effects of proficiency and language dominance on cross-language activation. *Sec. Lang. Res*. 30, 251–271. doi: 10.1177/0267658313503467

Morford, J. P., Wilkinson, E., Villwock, A., Piñar, P., and Kroll, J. F. (2011). When deaf signers read English: do written words activate their sign translations? *Cognition* 118, 286–292. doi: 10.1016/j.cognition.2010.11.006

Narasimhan, B., and Gullberg, M. (2011). The role of input frequency and semantic transparency in the acquisition of verb meaning: evidence from placement verbs in Tamil and Dutch. *J. Child Lang*. 38, 504–532. doi: 10.1017/S0305000910000164

Narasimhan, B., Kopecka, A., Bowerman, M., Gullberg, M., and Majid, A. (2012). "Putting and taking events: a crosslinguistic perspective," in *Events of Putting and Taking: A Crosslinguistic Perspective*, ed A. Kopecka and B. Narasimhan (Amsterdan: John Benjamins Publishing). p. 1–18.

Ortega, G. (2013). *Acquisition of a signed phonological system by hearing adults: The Role of sign structure and iconicity* (Unpublished Doctoral Dissertation). University College London, United Kingdom.

Ortega, G. (2020). Systematic mappings between semantic categories and types of iconic representations in the manual modality: a normed database of silent gesture. *Behav. Res. Methods* 52, 51–67. doi: 10.3758/s13428-019-01204-6

Ortega, G., and Morgan, G. (2010). Comparing child and adult development of a visual phonological system. *Lang. Interact. Acquis*. 1, 67–81. doi: 10.1075/lia.1.1.05ort

Ortega, G., and Morgan, G. (2015). Input processing at first exposure to a sign language. *Sec. Lang. Res.* 21, 443–463. doi: 10.1177/0267658315576822

Ortega, G., and Ozyurek, A. (2013). "Gesture-sign interface in hearing non-signers' first exposure to sign," in *The Tilburg Gesture Research Meeting [TiGeR 2013].* (Tilburg).

Özçalişkan, S. (2016). Do gestures follow speech in bilinguals' description of motion?* *Bilingualism* 19, 644–653. doi: 10.1017/S1366728915000796

Padden, C., Hwang, S.-O., Lepic, R., and Seegers, S. (2015). Tools for language: patterned iconicity in sign language nouns and verbs. *Topics Cogn. Sci*. 7, 81–94. doi: 10.1111/tops.12121

Pauwels, P. (2000). P*ut, Set, Lay and Place: A Cognitive Linguistic Approach to Verbal Meaning*. München: Lincom Europa.

Perniss, P., Thompson, R. L., and Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Front. Psych.* 1:227. doi: 10.3389/fpsyg.2010.00227

Quinto-Pozos, D., and Parrill, F. (2015). Signers and co-speech gesturers adopt similar strategies for portraying viewpoint in narratives. *Topics Cogn. Sci*. 7, 12–35. doi: 10.1111/tops,0.12120

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available online at: http://www.R-project.org/

Schembri, A., Jones, C., and Burnham, D. (2005). Comparing action gestures and classifier verbs of motion: evidence from australian sign language, taiwan sign language, and nonsigners' gestures without speech. *J. Deaf Studies Deaf Educ*. 10, 272–290. doi: 10.1093/deafed/eni029

Sevcikova, Z. (2014). *Categorical versus gradient properties of handling handshapes in British Sign Language (BSL)* (Unpublished Doctoral Dissertation). University College London, United Kingdom.

Singleton, J. L., Morford, J. P., and Goldin-Meadow, S. (1993). Once is not enough: standards of well-formedness in manual communication created over three different timespans. *Language* 69, 683–715. doi: 10.2307/416883

Slobin, D. I. (1996). 'From "thought and language" to "thinking for speaking",' in *Rethinking Linguistic Relativity*, eds J. J. Gumpertz and S. C. Levinson (Cambridge: Cambridge University Press), 70–96.

Slobin, D. I. (2013). "Typology and channel of communication. Where do sign languages fit in?," in *Language Typology and Historical Contingency: In honor of Johanna Nichols*, eds B. Bickel, L. A. Grenoble, D. A. Peterson, and A. Timberlake (Amsterdan: John Benjamins Publishing Company). p. 48–68.

Slobin, D. I., Bowerman, M., Eisenbeiss, S., and Narasimhan. (2011). "Putting things in places: developmental consequences of linguistic typology," in *Event Representation in Language and Cognition*, eds J. Bohnemeyer & E. Pederson (Cambridge: Cambridge University Press), 134–165.

Slobin, D. I., Hoiting, N., Kuntze, M., Lindert, R., Weinberg, A., Pyers, J., et al. (2003). "A cognitive/functional perspective on the acquisition of 'classifiers'," in *Perspectives on Classifier Constructions in Sign Languages*, ed K. Emmorey (Mahwah, NJ: Lawrence Erlbaum Associates), 271–296.

Supalla, T. R. (1982). *Structure and Acquisition of Verbs of Motion and Location in American Sign Language*. San Diego: University of California.

Taub, S., Galvan, D., Piñar, P., and Mather, S. (2008). *Gesture and ASL L2 Acquisition. Sign Languages: Spinning and Unravelling the Past, Present and Future.* Petrópolis: Arara Azul.

van Nispen, K. (2017). Production and comprehension of pantomimes used to depict objects. *Front. Psych*. 8:16. doi: 10.3389/fpsyg.2017.01095

Viberg, A. (1998). "Crosslinguistic perspectives on lexical acquisition: the case of language-specific semantic differentiation," in *Perspectives on Lexical Acquisition in a Second Language,* eds K. Haastrup and A. Viberg (Lund: Lund Univ Press), 175–208.

Weisberg, J., Casey, S., Sevcikova Sehyr, Z., and Emmorey, K. (2020). Second language acquisition of American Sign Language influences co-speech gesture production. *Bilingu. Lang. Cogn*. 23, 473–482. doi: 10.1017/S1366728919000208

Wilcox, S., and Wilcox, P. P. (1997). *Learning To See: Teaching American Sign Language as a Second Language (1st edition)*. Washington, DC: Gallaudet University Press.

Wittenburg, P., Brugman, H., Russell, A., Klassmann, A., and Sloetjes, H. (2006). "ELAN: a professional framework for multimodality research," in *Proceedings of the 5th International Conference on Language Resources and Evaluation* (Genoa), 1556–1559.

Zwitserlood, I. (2012). "Classifiers," in *Sign Language: An International Handbook*, eds R. Pfau, M. Steinbach, and B. Woll (Berlin; Boston, MA: Mouton de Gruyter), 158–186.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 19 November 2021
doi: 10.3389/fpsyg.2021.732030

Fuzzy Lexical Representations in Adult Second Language Speakers

Kira Gor[1], Svetlana Cook[2], Denisa Bordag[3,4], Anna Chrabaszcz[5,6] and Andreas Opitz[3]*

[1]Graduate Program in Second Language Acquisition, School of Languages, Literatures, and Cultures, University of Maryland, College Park, MD, United States, [2]National Foreign Language Center, University of Maryland, College Park, MD, United States, [3]Herder Institute, University of Leipzig, Leipzig, Germany, [4]University of Haifa, Haifa, Israel, [5]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, United States, [6]Center for Language and Brain, HSE University, Moscow, Russia

We propose the fuzzy lexical representations (FLRs) hypothesis that regards fuzziness as a core property of nonnative (L2) lexical representations (LRs). Fuzziness refers to imprecise encoding at different levels of LRs and interacts with input frequency during lexical processing and learning in adult L2 speakers. The FLR hypothesis primarily focuses on the encoding of spoken L2 words. We discuss the causes of fuzzy encoding of phonological form and meaning as well as fuzzy form-meaning mappings and the consequences of fuzzy encoding for word storage and retrieval. A central factor contributing to the fuzziness of L2 LRs is the fact that the L2 lexicon is acquired when the L1 lexicon is already in place. There are two immediate consequences of such sequential learning. First, L2 phonological categorization difficulties lead to fuzzy phonological form encoding. Second, the acquisition of L2 word forms subsequently to their meanings, which had already been acquired together with the L1 word forms, leads to weak L2 form-meaning mappings. The FLR hypothesis accounts for a range of phenomena observed in L2 lexical processing, including lexical confusions, slow lexical access, retrieval of incorrect lexical entries, weak lexical competition, reliance on sublexical rather than lexical heuristics in word recognition, the precedence of word form over meaning, and the prominence of detailed, even if imprecisely encoded, information about LRs in episodic memory. The main claim of the FLR hypothesis – that the quality of lexical encoding is a product of a complex interplay between fuzziness and input frequency – can contribute to increasing the efficiency of the existing models of LRs and lexical access.

Keywords: L2, L1, fuzzy, lexical representation, word recognition, lexicon, word learning

OPEN ACCESS

Edited by:
Antonio Benítez-Burraco, Sevilla University, Spain

Reviewed by:
Kathy Conklin, University of Nottingham, United Kingdom
Gary Libben, Brock University, Canada

***Correspondence:**
Denisa Bordag
denisav@uni-leipzig.de

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *28 June 2021*
Accepted: *20 October 2021*
Published: *19 November 2021*

Citation:
Gor K, Cook S, Bordag D, Chrabaszcz A and Opitz A (2021) Fuzzy Lexical Representations in Adult Second Language Speakers. Front. Psychol. 12:732030. doi: 10.3389/fpsyg.2021.732030

INTRODUCTION

This article introduces the fuzzy lexical representations (FLRs) hypothesis with a focus on adult second language (L2) learners. It outlines the construct of the FLR that is characterized by imprecise, or fuzzy encoding of its form and/or meaning, and potentially, the mapping between them. Less distinct boundaries of FLRs result in their reduced differentiation from neighboring representations in the mental lexicon. Fuzziness is primarily a property of less familiar words that occur less frequently in the input, both in the native language (L1) and L2; however, two factors contribute to its much greater pervasiveness in L2 than L1. First, less familiar words are more numerous in the L2 mental lexicon than the L1 mental lexicon. And second, for the reasons discussed below, L2 speakers experience more difficulties with encoding the phonological form and meaning of L2 words, establishing strong mappings between them, and integrating new L2 lexical entries in the mental lexicon compared to L1 speakers. In particular, more LRs in L2 retain fuzzy phonological encoding even for more familiar words.

Accordingly, we will refer to the empirical evidence primarily on L2, but also on L1 lexical processing, as appropriate. While a lexical representation (LR) in literate L2 speakers[1] encodes both the sound and the written form of the word, we will treat auditory encoding as the core aspect of form encoding. Indeed, the spoken modality is the primary source of input for a majority of monolingual speakers and also bilinguals/monolinguals, who learn languages beyond a traditional foreign language classroom with a strong emphasis on written input. We believe that, in addition to being more ecologically valid, this approach helps to better address the L2-specific sources of fuzziness and to offer suggestions for the development of testable models of auditory L2 word recognition. The focus on auditory LRs makes it possible to rely on the existing literature on the topic when more research on the role of phonological encoding than orthographic encoding in FLRs is available.

In the following sections, we will discuss the properties of FLRs in L2 and explore the dynamics of their engagement in the process of word recognition. In particular, we will address the following questions associated with fuzziness in L2 LRs:

- Why is there a need for a FLR construct?
- Is the construct of FLR new, or does it rename existing constructs?
- What causes the fuzziness of L2 LRs?
- What consequences does fuzziness in LRs have for L2 word recognition and lexical processing?
- How do FLRs develop over time – is fuzziness reduced?
- How can the construct of FLRs contribute to increasing the efficiency of the existing models of LRs and lexical access?

The FLR hypothesis is based on the idea that imprecise and ambiguous linguistic encoding of any component of the LR (phonological, orthographic, and lexical-semantic)[2] has important consequences for several aspects of L2 word storage and retrieval. First, poor encoding at one or more levels leads to weak form-meaning connections and, sometimes, incorrect form-meaning mappings. L2 speakers tend to confuse L2 words that are less similar-sounding more than L1 speakers do, that is, there is a greater Levenshtein distance between confusable L2 words than confusable L1 words (see Levenshtein, 1966; Cook et al., 2016). Second, poor phonological encoding leads to a low level of lexical activation and competition of individual LRs in spoken word recognition, because FLRs are not clearly separated from their phonological neighbors and none of the competing LRs is treated as a clear "target." At the same time, a larger set of word candidates including phonologically more distant words, which would not normally be activated in L1, gets activated, albeit at lower activation levels. And third, fuzzy phonological encoding of LRs may persist in L2 and interact with input frequency in such a way that the quality of encoding may not necessarily improve with repeated encounters with the word in the input. Accordingly, input frequency, while being an important factor in shaping LRs, is not the sole determining factor in resolving fuzziness of LRs in L2. Together, the unfaithful encoding and the mapping problems produce nonnative patterns of lexical activation, competition, and selection in L2 word recognition.

The FLR hypothesis accounts for a range of phenomena observed in L2 lexical processing, including well-documented lexical confusions in comprehension and production, slow lexical access, retrieval of incorrect lexical entries, weak lexical competition, reliance on sublexical rather than lexical heuristics in word recognition, the precedence of word form over meaning, and the prominence of detailed, even if imprecisely encoded, information about LRs in episodic memory.

WHY IS THERE A NEED FOR A FLR CONSTRUCT?

A number of phenomena observed in L2 word processing and learning seem to be associated with one core property characterizing nonnative LRs. According to the FLR hypothesis, this property is unfaithful or fuzzy encoding. The FLR hypothesis seeks a principled account of fuzziness characterizing L2 LRs and explores the sources of fuzziness in L2 lexical processing.

L2 speakers are notorious for confounding words, with numerous examples of misunderstandings observed in L2 oral and written productions (Dušková, 1969; Hemchua and Schmitt, 2006). They are unsure whether and how *truck* is different from *trunk*, *helmet* from *hamlet*, *lie* from *lay*, or *accident* from *incident* (a form-related confusion). At beginning stages of L2 acquisition, L2 learners may experience uncertainty about whether *trunk* means "suitcase," a "tree part," or both, and how these meanings can help to make an informed guess about what the elephant's *trunk* is (a meaning-related confusion). Both meaning- and form-related aspects of the development of robust lexical representations in L2 are prone to difficulties.

When learning new vocabulary, L2 speakers struggle to establish a strong connection between the semantic representation for a novel word with its form, especially when inferring the meaning from sentence context or extended context, in which the word occurs (see Bordag et al., 2017a,b). Novel word meaning recall in incidental and even deliberate vocabulary learning is very low even after several encounters with the same word in a text (Lawson and Hogben, 1998; Webb, 2005, 2007; Elgort, 2011; Elgort and Warren, 2014). L2 speakers do not remember the meaning of these words well, apparently,

[1]We are considering mostly spoken word encoding and processing, thereby narrowing the scope of the discussion to hearing people.

[2]Sensitivity to the morphological structure of morphologically complex words is another potential source of fuzziness in the lexical encoding. This is a major and controversial aspect of word structure, and it will remain outside the scope of this publication.

because it is not robustly encoded, and a new LR with a fuzzy meaning resists consolidation and efficient storage in long-term memory (Qiao et al., 2009; Qiao and Forster, 2017). While L1 speakers, adults and children, also experience difficulties in novel word learning, L2 speakers deal with an additional set of difficulties. Thus, improper phonological encoding of LRs that we refer to as phonolexical encoding (in contrast with phonological encoding of individual word segments, see also, e.g., Llompart, 2021 for the use of this term) leads to nonnative patterns in the processing of similar-sounding words. More generally, improper phonolexical encoding influences the properties of nonnative lexical networks – irrespective of whether the LRs involve especially difficult L2 phonological contrasts (Pallier et al., 2001; Darcy et al., 2012; Mora and Darcy, 2013) or not (Gor and Cook, 2020). Given that the fuzziness in L2 lexical encoding of particular words may never be resolved for individual L2 speakers even when they reach high L2 proficiency levels, it is useful to incorporate the degree of fuzziness in lexical encoding as a property of numerous L2 words that interacts with the well-attested input frequency effects. Accounting for fuzziness in form and meaning encoding as well as in the mapping between them will contribute to a more efficient and accurate modeling of the learning trajectories for different types of lexical units, L2 learner profiles, and learning conditions, as well as explain the differences between L1 and L2 lexical processing [see below about the developmental trajectories for L2 LRs within the framework of the Ontogenetic Model (Bordag et al., 2021a,b)].

IS THE CONSTRUCT OF FLR NEW, OR DOES IT RENAME THE EXISTING CONSTRUCTS?

The FLR hypothesis builds upon existing research, primarily on L2 phonolexical encoding. The term *fuzzy lexical representations* has been previously used in the SLA literature, albeit in a limited sense – to refer to the specific phonological difficulties that L2 learners encounter when encoding problematic L2 phonological contrasts in LRs (Darcy et al., 2013; Llompart and Reinisch, 2019). The FLR hypothesis also draws on research that does not necessarily use the construct of fuzziness in phonological representations; however, it examines the consequences of nonnative phonological encoding, such as lexical confusions (Pallier et al., 2001; Escudero et al., 2008, 2014; Hayes-Harb and Masuda, 2008; Escudero and Wanrooij, 2010; Chrabaszcz and Gor, 2014, 2017). This strand of research explores the bottom-up direction in the encoding of LRs: L2 speakers are inefficient at the processing of L2 phonetic cues, which leads to problems with the phonological categorization of word segments and the identification of the phonemic sequences corresponding to the spoken word input. The phonological categorization problems, in turn, contribute to the poor lexical encoding of the words with difficult L2 phonemes or phonological contrasts (e.g., "rock" and "lock" are confusable for Japanese learners of English; Ota et al., 2009).

Importantly, the construct of FLRs has also been extended to refer to low-resolution L2 lexical representations that do not necessarily involve particularly problematic L2 segments, but nevertheless lead to lexical confusions of similar-sounding words (Cook et al., 2016). Such poorly encoded FLRs contribute to a nonnative pattern of lexical competition in phonological priming experiments (Gor et al., 2010; Cook and Gor, 2015; Gor and Cook, 2020).

The FLR hypothesis expands the construct of fuzziness to all levels of L2 lexical encoding: – phonological and orthographic form, as well as meaning – and also to form-meaning and phonological form-orthographic form mappings. Within this approach, fuzzy encoding interacts with input-based factors, such as lexical frequency or context predictability, to shape the pattern of L2 spoken word recognition and other aspects of L2 lexical processing. Note that the FLR hypothesis shares its focus on the quality of lexical encoding with the lexical quality hypothesis developed for L1 reading (Perfetti and Hart, 2002; Perfetti, 2007), and the lexical entrenchment hypothesis developed for written word recognition (Diependaele et al., 2013; Brysbaert et al., 2017). However, in contrast to the lexical entrenchment hypothesis, the FLR hypothesis treats the quality of lexical encoding as a product of several interacting factors rather than solely an outcome of input frequency.

The FLR hypothesis seeks to build a bridge between the acquisitional aspects of SLA research and word recognition studies. It posits that an L2-specific set of difficulties in lexical encoding and word recognition arises from two major factors shaping adult SLA: age of onset and L1 transfer.

Late Age of Onset

The post-puberty age of onset of language acquisition is associated with lower learning outcomes for L2 (DeKeyser, 2012; Hartshorne et al., 2018; Bylund et al., 2021), with post-puberty learners failing to achieve native levels of proficiency on a battery of tests targeting different aspects of L2 linguistic knowledge, including L2 phonological sensitivity and control of idiomatic language (Abrahamsson and Hyltenstam, 2009; Bylund et al., 2021) and in lexical development (Bylund et al., 2019). Phonological acquisition is particularly vulnerable and shows early age effects (Granena and Long, 2013).

L1 Transfer

The L1 mental lexicon is already in place when L2 lexical learning starts and so is the L1 phonological system. L2 learners need to overcome the influence of L1 in developing the L2 phonological system and new form-meaning mappings for L2 LRs. The specific difficulties in L2 lexical encoding of both form and meaning can often be traced to a particular combination of L1 and L2 (Jarvis, 2000; Barrios and Hayes-Harb, 2020, 2021; Llompart, 2021). For example, an L1 German speaker may not encode the difference in the English words *cod* and *cot* due to final consonant devoicing in German, while encoding this difference will not present a problem to an L1 French speaker. An L1 French speaker will be confused with the meaning of the English word *library*, because *la librairie* in French is a bookstore.

While a late age of onset and L1 transfer are not independent factors – L1 transfer is to be expected at an older age when L1 is already in place – they contribute to FLRs in different ways. A late age of onset is associated with reduced network plasticity in general and parasitic reliance on the L1 lexical network (Hall and Ecke, 2003), both potentially increasing fuzziness in lexical connections. In contrast, L1 transfer depends on a particular combination of L1 and L2 and manifests itself in issues with individual aspects of lexical encoding that have their source in L1 phonology, orthography, or semantics.

Another factor that mitigates novel word learning is L2 proficiency. On the one hand, L2 proficiency encompasses different kinds of linguistic knowledge, with lexical knowledge being part of it, since the level of L2 proficiency is associated with the size and the degree of familiarity of L2 vocabulary (Laufer, 1997; Alavi and Akbarian, 2012). On the other hand, vocabulary size by itself is also a predictor of subsequent lexical learning. As in "the rich get richer," L2 learners with a larger vocabulary and a more elaborate lexical network are more efficient at the lexical encoding of novel words (Llompart, 2020; Daidone and Darcy, 2021). In this vein, the extent of fuzziness of individual LRs depends on their stage on each LR's acquisition trajectory. The actual shape of the developmental curve for different aspects of the lexical representation depends on a number of factors, and these aspects do not necessarily develop in parallel. In section "**How Can the Construct of FLRs Help to Improve the Existing Models of LRs and Lexical Access?**," we discuss how lexical encoding becomes more precise as L2 learners' proficiency increases within the framework of the Ontogenetic Model (Bordag et al., 2021a,b).

The effect of age of acquisition (AoA) for lexical learning can be separated from the effect of lexical frequency or the cumulative number of encounters with the word (with the latter also depending on AoA). The role of AoA for L1 words was demonstrated in a megastudy using crowdsourcing technology that explored self-reported AoA for over 30,000 English words and showed that the AoA ratings explained a substantial percentage of the variance in the lexical decision data of the English Lexicon Project, over and above the effects of log frequency, word length, and similarity to other words (Kuperman et al., 2012). This effect of AoA in L1 above and beyond lexical frequency (and the cumulative number of encounters with the word) suggests that at an older age, lexical encoding and retrieval becomes less efficient. It is to be expected that the quality of lexical encoding will be less efficient across-the-board in adult L2 learners.

Fuzziness can be also viewed as a property that characterizes the L2 lexicon or a lexicon of a nonproficient L2 speaker in general. In this sense, the approach is related to the cognitive theories that address the differences in representation and processing in novices and experts. One such theory is the Fuzzy Trace Theory (FTT) by Brainerd and Reyna (2002). FTT is a dual-process theory that assumes two types of representation of past events: meaning-based gist representations, which support fuzzy (yet advanced) intuition, and superficial verbatim representations of information, which support precise analysis (Reyna, 2012, p. 332). Both types of representations are encoded in parallel, can be retrieved independently from each other, and have different forgetting rates (with verbatim traces becoming inaccessible at a faster rate than gist traces). It is important to note that despite the relevance of the theory to the topics discussed in this paper, the term "fuzzy" is used differently in the FTT. In the FTT, it relates to the processing of experts, who rely on a broad and deep knowledge foundation, from which they can derive the "gist" which the authors of FTT refer to as fuzzy. The novices, on the other hand, do not have such a rich knowledge base at their disposal and are thus more focused on the surface, form-based representations (cf. form prominence in L2 in section "**Form Prominence in L2**" and fuzziness decreasing over time in section "**How Do FLRs Develop Over Time – Is Fuzziness Reduced?**").

The FLR hypothesis connects different strands of research on lexical encoding in L2 word recognition and vocabulary learning and frames the discussion of how to predict and measure fuzziness in lexical representations and incorporate it as an additional parameter in models of L2 lexical processing. The goal of the FLR approach is, on the one hand, to account for systematic patterns of fuzziness associated with specific encoding problems and, on the other, for random fuzziness also present in the LRs and lexical networks of adult L2 speakers. To summarize, the FLR hypothesis, while drawing on previous research, extends the construct of fuzziness to different aspects of lexical encoding of L2 words and, unlike other approaches, treats it as a property of L2 word encoding that interacts with other factors in vocabulary acquisition and processing.

WHAT CAUSES L2 LRS TO BE FUZZY?

Fuzziness in Form, Meaning, and Form-Meaning Mappings

When a learner encounters a new spoken word, the phonological form and the meaning of this word are encoded, and a connection between the form and the meaning is established. If the word is encountered only once and especially in noisy conditions – a property of naturalistic settings – its sound form may not get properly encoded. If the context in which the new word is encountered does not make it possible to unambiguously identify its meaning, it will also be encoded without proper specifications and details, maybe merely as a "place-holder" with broad semantic properties, such as a reference to a semantic field (e.g., "some kind of a gardening tool" and "a positive human character trait"). With more encounters with the word, its form- and meaning-related properties become better defined, and the encoding becomes more precise. In this respect, the word learning trajectory is similar in both L1 and L2 word learning.

At the same time, several factors specific to adult L2 learning contribute to increased fuzziness in L2 lexical representations. As we state above, these factors are globally defined by the late AoA of individual lexical items and the fact that L2 words are acquired when the L1 mental lexicon is already in place.

The existence of the L1 phonological system supported by a system of phonetic cues for the encoding of speech sounds as phonemes means that a lot of perceptual restructuring will be needed and new L2 phonological categorization routines will have to be established even for the L2 sounds that have correspondences in the L1 (Llompart and Reinisch, 2019; Llompart, 2020). According to the FLR hypothesis, two aspects of phonological processing in the L2 lead to less precise spoken word encoding: (i) problems with the phonological categorization of difficult L2 phonemes or contrasts (especially, in situations presenting an allophonic split problem, when two distinct L2 phonemes map onto a single L1 phoneme, i.e., in single-category assimilations – see Best and Tyler, 2007), and (ii) the overall imprecision of phonolexical encoding in L2 compared to L1 that may involve ambiguous word segments or their inexact sequence. Both aspects stem from the mismatches between the phonological systems of the L1 and the L2; however, the former has received more attention in the literature than the latter (see Llompart and Reinisch, 2019 regarding the role of phonetic flexibility in the robustness of L2 phonolexical encoding).

The first aspect manifests itself when adult L2 learners encounter a phonological contrast absent in their L1, such as the vowel /i/-/ɪ/ contrast in English that is absent in Spanish or French or the /y/-/u/ contrast in French that is absent in English. L2 learners' perceptual systems are not attuned to processing the phonetic cues differentiating these phonemes, and the L2 phonemes may not be properly represented and contrasted in the L2 phonological system. This absence of phonetic attunement and/or robust phonological categories in L2 has two implications for lexical learning. First, phonological encoding of the words differentiated by this contrast is fuzzy, because the contrasting phonemes are not properly categorized (see Pallier et al., 2001; Sebastian-Gallés et al., 2006; Hayes-Harb and Masuda, 2008; Broersma and Cutler, 2011; Darcy et al., 2012, 2013; Sebastián-Gallés and Díaz, 2012). And second, given that phonological categorization difficulties persist over time, phonolexical representations of the words with problematic L2 contrasts remain fuzzy even at higher levels of proficiency (Chrabaszcz and Gor, 2014, 2017).

The second aspect has to do with a more diffused perceptual categorization deficit in L2 (resulting from the phonetic differences between L1 and L2 sounds and also language-specific phonotactics, segmentation, and lexical prosody), which leads to "summative," less precise phonolexical encoding of novel spoken words (cf. coarse-grained orthographic representations that lack precise positional information, Grainger and Ziegler, 2011). This latter type of fuzziness in LRs is less systematic and more akin to white noise, as it makes the word encoding indistinct and leads to underdifferentiated LRs that are easily confusable not only with their phonological neighbors, but also with more distant similar-sounding words. The effects of such blurred phonolexical encoding in the absence of a particular difficult contrast leading to phonological confusion were reported in a study manipulating the Levenshtein distance (Levenshtein, 1966) between the matching Russian translation of an English word and its similar-sounding counterpart. It revealed that L1 Russian speakers and English-speaking L2 learners of Russian were differently sensitive to the "overall" phonological similarity between two L2 words. While Russian L1 speakers could be confused by two words with a Levenshtein difference of 1 (i.e., one phoneme substitution, addition, or deletion), L2 speakers were confused with the words with a Levenshtein difference of not only 1, but also 2 (Cook et al., 2016). Unlike the fuzziness in LRs resulting from particular resistant phonological difficulties, diffused spoken form fuzziness is reduced with more input. A similar pattern of lexical acquisition starting with low-resolution lexical representations that are improved during differentiation and sharpening was reported for L2 children (Baxter et al., 2021a,b). Meaning encoding in L2 is also characterized by L2-specific features. Thus, L1 speakers rely both on linguistic and on nonlinguistic context (i.e., on schemata, knowledge of the situation, or real-life knowledge) when they establish the meaning of novel words that they will encode in the LR. In contrast, lower-proficiency L2 speakers make inefficient use of the linguistic context because they do not understand it well and/or they fail to process it efficiently enough in real time. Multiple examples of the inability of L2 speakers to make use of the high close probability contexts to predict the upcoming word when the sentence is presented in noise is an illustration of the auditory processing constraints in L2, albeit in extreme conditions (see, e.g., Gor, 2014).

According to the FLR hypothesis, fuzziness in the encoding of form and/or meaning in L2 leads to fuzziness in form-meaning connections. As mentioned above, in L1, word forms and meanings are acquired together, which results in well-defined and strong form-meaning connections. While acquiring L1, children have to categorize both strings of sounds – to identify individual word forms – and portions of reality (objects, events, people, animals, etc.) that correspond to them; a child learning a new word, for example, "parrot," simultaneously encounters the bird (possibly, its picture, or a toy), to which it refers. The form and the semantic components of the L1 lexicon thus develop simultaneously, resulting in amalgamated lexical entries with strongly connected semantic and phonological representations (Perfetti, 2007). In contrast, in L2, word forms and meanings are often acquired separately, with word meanings initially borrowed from L1 (Jiang, 2000). An adult L2 learner acquiring the new English word "parrot" may have its meaning already represented through experience with L1, in which case a new L2 word is mapped onto the already existing semantics. However, this is not a unique possibility. First, novel L2 words may refer to objects or concepts that do not occur in L1 and have no corresponding word in L1 – these are often culturally-specific words (e.g., *bar/bat mitzva*, "the Jewish coming of age ritual," *Hanukkah* in Hebrew, *sutki* "24-h period" in Russian, *siesta* "an afternoon nap" in Spanish, and *bento* "a single-portion meal packed in a box" in Japanese). And second, the development of L2 semantic representations often starts by borrowing L1 semantic representations, which serve as a shortcut, although some restructuring and reconfiguration of the L2 semantic representations may be required at a later time, when more L2 input providing finer-grained information about the specific L2 meaning becomes available. For example, an L2 learner of English may first discover the loanword *kosher* in the general meaning of "legitimate" and later will discover that *kosher* has a very specific meaning referring to Jewish food that has been ritually prepared, or the discovery of the core and general-purpose meanings may occur in the inverse order. Complex relations often also exist between translation equivalents which are rarely completely equivalent in two languages. For example, an L1 speaker of Russian learning L2 English may initially map the concept of "a piece of furniture with a horizontal surface" to the L1 word *stol*, which is a semantic equivalent of the English *table*. However, with more experience with L2 English, the original mapping will have to be updated to reflect a difference in meaning for *table* vs. *desk* that Russian does not lexicalize.

Figure 1 is a schematic depiction of the lexical representation of the word "parrot" and a similar-sounding word "parent" in British English in the L1 and L2 lexicons. In British English, unlike many dialects of American English, these words are not phonological neighbors in strict terms, since their Levenshtein distance has a value of two (the difference in the /æ/-/ɛ/ vowels and the presence or absence of /n/, i.e., one substitution and one deletion/addition). In L1, the lexical representations are highly specified both at the phonological and semantic levels: Phonological form is encoded precisely allowing L1 speakers to efficiently constrain the word's phonological neighborhood and activate it quickly. The words *parrot* and *parent* are disambiguated at the initial syllable due to the accurate encoding of the first vowel. The form /pærət/ has a strong connection to the semantic representation of "parrot." The same is true for the word /pɛrənt/ that is strongly connected to "parent." "Parrot" is associated with the semantic field of "birds," while "parent" is associated with "family" and "children." In L2, the phonological representations are fuzzy, and at the initial stages of acquisition, they can be characterized by imprecise phonological sequences involving the inclusion of incorrect or additional phonemes (or their exclusion) or a scrambled order of phonemes. At this stage, the word *parrot* is encoded as /p?r?t/, which means that two segments are fuzzy. As a consequence, the distinction between similar-sounding L2 words is also fuzzy, or blurred, and the more familiar word *parent* may be accessed instead of the intended less familiar word *parrot*. Similarity of forms leads to the blending of two phonological neighborhoods in L2 and the activation of similar-sounding words that would not be activated in L1 lexical access. In this example, L2 semantic encoding is also fuzzy, and accordingly, *parrot* is associated with the broader semantic field of "birds" rather than "parrot," and moreover, if it erroneously accesses the semantics of "parent," it may activate the semantic field of "family."

The FLR hypothesis argues that the weak mapping links between L2 form and semantic representations are likely related to the specific starting conditions for their emergence in L2 acquisition. The focus of basic L2 word acquisition, especially at the initial stages, is on the word forms that need to be encoded, stored, and mapped onto the preexisting semantic representations borrowed from L1. Due to these developmental differences between L1 and L2 lexical acquisition, the mapping between phonological forms and semantic representations is weaker and fuzzier in L2. At the same time, since a substantial semantic

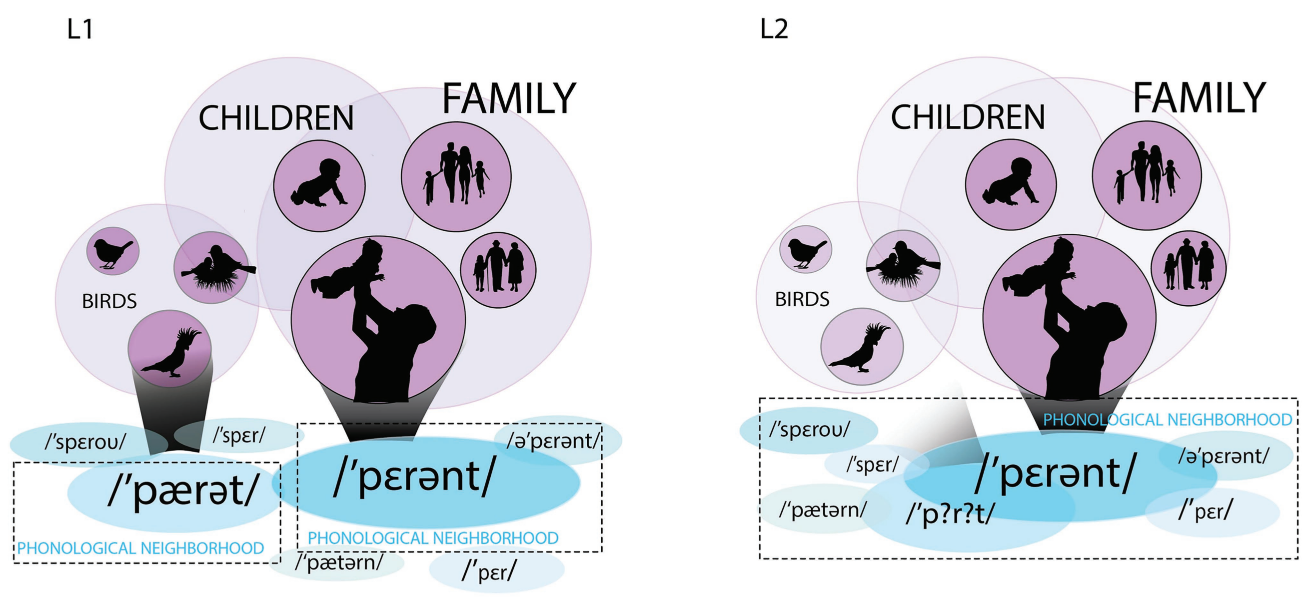


FIGURE 1 | L1 and L2 lexical representations. Panel on the left represents the word *parrot* and a similar-sounding word *parent* in L1 (in British English), and panel on the right – the same words in L2. The blue ellipses at the bottom represent the phonological neighbors and similar-sounding words. The size of the ellipses represents the lexical frequency of the words. The mauve circles represent the semantic representations and their semantic fields, while the grey cones represent the activation spreading from the form to meaning. In L1, the words *parrot* and *parent* are differentiated at the phonological level and belong to different neighborhoods. Each word activates its corresponding meaning. In L2, both the form and meaning of the word *parrot* are fuzzy. It is phonologically encoded as /p?r?t/ and is likely to be confused with /pɛrənt/, a high-frequency and more familiar word. Semantically, /p?r?t/ can activate "parent" and "family," but also "birds," rather than "parrot" because the exact semantic referent is unavailable given the fuzzy semantic encoding.

store has typically already been developed during L1 acquisition, fuzziness at the semantic level does not have to be initially as pervasive in L2 as it was during L1 acquisition. When new L2 forms can be mapped to the existing semantic representations, fuzziness may arise only later when the L2 learner discovers that there is no complete translation equivalence between two given words and that an adjustment to the semantic representation needs to be made. Importantly, new semantic representations also emerge in L2 – both temporal (e.g., when only imprecise meaning can be inferred from the context, e.g., that a "parrot" is some kind of a bird, without the specific knowledge about the species) and longer-lasting (when new meanings are acquired through L2, for which there is no preexisting semantic representation; Bordag et al., 2018). Such fuzzy representations undergo stages in a similar way as do emerging semantic representations in child L1 acquisition. It is for further research to establish whether the form-meaning links differ in their strength depending on whether the form and the semantic representation were acquired successively or simultaneously.

The studies on lexical learning recur to semantic priming to gauge the robustness of semantic representations of newly acquired words (Elgort, 2011; Elgort and Warren, 2014; Bordag et al., 2015, 2017a). The absence of semantic priming or semantic inhibition is interpreted as evidence of poor integration of the newly encoded lexical representation into the semantic network, which is associated with its fuzzy semantic encoding. In fact, recent research (Elgort, 2011; Bordag et al., 2015, 2017a, 2018) has already provided evidence that different semantic priming effects in L2 emerge depending on several properties of the newly learned words associated with the quality of semantic encoding. In semantic priming tasks testing the integration of new L2 LRs into the L2 semantic network, the primes that were known existing words (Bordag et al., 2015), new primes, for which the participants could recall the meaning (2017a), and novel words with existing meanings (2018) produced facilitation in the processing of the targets, semantically related real words. By contrast, the primes, for which new semantic representations were established (2015, 2018) or for which the participants could recall only the orthographic form but not the meaning (2017a), produced inhibition (cf. also Carr and Dagenbach, 1990; Dagenbach et al., 1990a,b for L1). Therefore, better encoded, that is, less fuzzy new LRs show evidence of integration into the semantic network (facilitation), while the LRs with fuzzy semantic encoding slow down lexical retrieval of the target.

The discussion above has focused on the problems with the linguistic encoding of form and meaning and weak form-meaning mappings – all contributing to FLRs. The next section is devoted to a major factor shaping word learning – lexical frequency.

FLRs, Lexical Frequency, and Lexical Entrenchment

Lexical frequency is estimated based on the frequency of word occurrence in a representative corpus. It is associated with word knowledge and the speed and accuracy of word recognition and retrieval from memory (see, e.g., Kuperman et al., 2012). Accordingly, one of the critical factors affecting L2 word learning, storage, and recognition is the reduced amount of L2 input leading to reduced subjective lexical frequencies (e.g., Ellis, 2002). A proposed explanation for the frequency effects evokes the notion of cognitive entrenchment – a cognitive consequence of increased exposure to a certain external stimulus. Every time a certain event occurs, its memory trace becomes more and more profound or entrenched. Since entrenchment is a function of repetition of cognitive events, units are variably entrenched depending on the frequency of their occurrence (Langacker, 1987; Tomasello, 2003). Higher levels of entrenchment are associated with a greater processing advantage. Conversely, a lack of entrenchment can lead to processing costs, which have varying implications for lexical access at different stages.

Cognitive entrenchment constitutes the core of the lexical entrenchment hypothesis (Diependaele et al., 2013; Brysbaert et al., 2017), which argues for the critical role of input frequency in determining the level of entrenchment of lexical entries in the L2 mental lexicon and focuses on the written modality. The lexical entrenchment hypothesis builds on the lexical quality hypothesis developed for reading in L1 (Perfetti and Hart, 2002; Perfetti, 2007) and treats lexical quality, or the quality of lexical encoding, as a direct product of the number of encounters with the word. It assumes that every new exposure to the word strengthens the form-meaning connections and contributes to stronger lexical entrenchment. Remarkably, Brysbaert et al. (2017) report that lexical information build-up is slower in L2 than in L1 and conclude that entrenchment in L2 may be qualitatively different from L1. This position, if further developed and substantiated by empirical evidence, may go in the direction of the acknowledgment of the increased role of fuzzy encoding in L2 compared to L1.

SLA research makes a distinction between input and intake in L2 (Corder, 1967; Gass, 1997), because for L2 learners, input processing in real time, and especially, processing auditory input, is effortful and error-prone. Depending on the L2 proficiency level, more or less auditory input is actually processed, that is, becomes intake. For spoken word recognition and learning, this means that only some L2 words in the input are noticed (Schmidt, 1990; Gass, 1997), understood, and lexically encoded. Accordingly, in contrast to the lexical entrenchment hypothesis and the computational models based on it, such as BIA+ and Multilink, the FLR hypothesis is built on the understanding that the quality of linguistic encoding in L2 is not determined solely by input frequency, but rather by a set of linguistic and cognitive factors, in addition to input frequency.

The factors that contribute to the processing of novel spoken words include the availability of the meaning for a new L2 lexical item and the relative ease or difficulty of phonological categorization and encoding given the combination of L1 and L2. According to the FLR hypothesis, the quality of lexical encoding interacts with input frequency, rather than automatically improves with more input. This interaction of the inherent difficulty of encoding, and in particular, phonological encoding of problematic L2 segments that is specific for different lexical entries with input frequency ultimately configures the properties of the LR in L2. Initial problems with encoding may be persistent, with some FLRs resisting lexical consolidation, in contrast to other LRs that are more amenable to robust encoding with sufficient input.

The quality of phonetic encoding has been shown to improve with increased input in young monolingual children (Swingley and Aslin, 2000; Garlock et al., 2001; White and Morgan, 2008) and adults (White et al., 2013). In a novel vocabulary learning experiment, monolingual English-speaking adults were trained and tested on nonword-nonobject picture pairings, with their eye movements monitored. Novel words were presented with various frequencies during training. In the testing phase, the participants showed no sensitivity to a one-feature phonetic mismatch in the presented test nonwords in their looking behavior. For higher frequency words, participants differentiated both one- and two-feature mispronunciations from correct pronunciations (White et al., 2013). This pattern can be partially extended to adult L2 learners; however, in L2 learners, input frequency does not solely determine the quality of phonological encoding of L2 words. As speakers of a particular L1, L2 learners experience difficulties in encoding particular phonemes and differentiating phonological contrasts, for example, *ship-sheep* is difficult to differentiate and properly encode for L1 Spanish speakers, and these difficulties persist even for high-frequency words. Thus, the quality of lexical encoding in L2 reflects the actual knowledge of the word, including its form and meaning, and depends on the word's frequency in the input. Importantly, it is also a product of the word's potential of being properly encoded given the particular combination of the L1 and L2, the linguistic properties of the word, such as the number of phonological neighbors, form salience, and imageability, the contexts in which it appears, and the proficiency level of the L2 learner.

WHAT CONSEQUENCES DOES FUZZINESS IN LRS HAVE FOR LEXICAL PROCESSING?

This section will address a range of issues observed in L2 word recognition and processing that the FLR hypothesis attributes to the fuzzy lexical encoding and form-meaning mappings. It will discuss the role of fuzziness in the nonnative patterns of lexical competition and lexical confusions, either transient or permanent. It will also connect fuzziness in LRs with the observations of form prominence in L2 and an increased reliance on recently engaged episodic representations in L2.

Spoken Word Recognition in L2: The Effect of FLRs on Lexical Competition in L2

According to the FLR hypothesis, lexical activation, competition, and selection for L2 words are affected by fuzziness in form and/or meaning encoding that leads to weak or incorrect form-meaning mappings. Recall that the FLR hypothesis identifies different sources of fuzziness, with fuzzy form or meaning encoding leading to fuzzy form-meaning mappings and lexical confusions, and asynchronous acquisition of form and meaning leading to weak (but not necessarily incorrect) form-meaning mappings. First, FLRs are weak competitors; consequently, lexical competition in L2 spoken word recognition is weak despite the fact that irrelevant competitors may be activated by mistake (Gor and Cook, 2020). And second, as a consequence of weak lexical competition, L2 speakers over-rely on sublexical processing in resolving lexical competition. Below, we elaborate on these points.

The models of auditory speech perception agree that before the correct word is identified, several potential word candidates are considered (TRACE, McClelland and Elman, 1986; Cohort theory, Marslen-Wilson, 1987; NAM, Luce and Pisoni, 1998). The selected word candidate will have the highest activation level among the competitors. According to one point of view, L2 speakers show greater processing costs in accessing L2 words because they activate a larger number of words than L1 speakers, that is, they have to deal with larger competitor sets (e.g., Van Wijnendaele and Brysbaert, 2002; Duyck et al., 2008; Gollan et al., 2008, 2011; Schmidtke, 2014).

The fact that L2 speakers engage larger competitor sets in L2, which are presumably a source of increased lexical competition in L2, is typically attributed to a perceptual deficit associated with nonnative phonology, such as inaccurate representations of L2 phonemes (Pallier et al., 2001; Cutler and Otake, 2004; Cutler et al., 2006; Darcy et al., 2012; Díaz et al., 2012). For example, Broersma and Cutler (2008, 2011) proposed that as a consequence of reduced sensitivity to nonnative phonological contrasts, L2 competitor sets include words that typically would not compete for selection during L1 lexical access. The competitor set, therefore, is expanded by these "phantom" activations, for example, a near-word DAF activates "deaf" in "DAFfodil," or words that arise at a word juncture (e.g., the near-word LEMP activates "lamp" in the phrase "eviL EMPire;" Broersma and Cutler, 2008, 2011). The L2 processing costs emerge from a greater competition due to spurious activation of irrelevant competitors. In contrast, Cook (2012) argued that phonological "foes," that is, phonological neighbors that compete for selection and slow down lexical access in L1 may turn into "friends" in L2 because of their phonological underspecification, and speed up lexical access in L2.[3]

Several issues need to be considered with regard to a larger competitor set in L2 in the light of the FLR hypothesis. First, the FLR hypothesis also predicts that fuzziness in the encoding will lead to spurious activation of irrelevant competitors thereby potentially increasing the competitor set. However, it predicts weaker competition in L2 spoken word

[3]Phonological neighbors with underspecification of phonological form in L2 may include homophones that are not differentiated in form, while they have distinct meanings (see Pallier et al., 2001). At the same time, most studies exploring phonological underspecification leading to lexical underspecification and to lexical confusions show an asymmetry associated with the degree of phonetic overlap between the L2 and L1 phonological units. In the case of *sheep-ship* in L1 Spanish speakers, *sheep* will be selected over *ship*, but not the other way round (cf. Barrios et al., 2016), suggesting that the two words are not completely homophonous.

recognition. This is because due to the uncertainty associated with the fuzzy encoding and weak form-meaning mappings, the words in the competitor set will have low resting level of activation, and the activation of a larger set of words will be low and diffused. Indeed, L1 neighborhood research shows that high-frequency competitors are likely to negatively impact the speed of identification of a lower-frequency target, while low-frequency competitors produce no sizeable effect (Luce et al., 1990). Since L2 speakers are less exposed to L2 than L1 speakers are to L1, their L2 mental lexicon consists of the words that display characteristics of low-frequency words in the L1 lexicon (Gollan et al., 2005). Therefore, there are no reasons to expect that the competition between the L2 words will be stronger than in L1.

Second, the L2 lexicon[4] of even high-proficiency speakers is generally smaller than the L1 lexicon, and consequently, the potential competitor set in L2 is more restricted than in L1. To illustrate the point, a well-educated native speaker knows about 20,000 word families, while highly educated nonnative speakers of English who are studying toward advanced degrees through the medium of English have a receptive English vocabulary size of around 8,000–9,000 word families (Nation, 2006). This comparison shows that the number of competing words in L2 should be considerably smaller than in L1, thus producing less competition overall. Thus, the effects of phantom activation of irrelevant words may be offset by a smaller competitor set in lower-proficiency L2 speakers.[5]

Third, the results of form priming experiments do not support strong lexical competition in L2. In phonological priming, onset overlap between the prime and the target leads to inhibition in L1, which is interpreted as an indication of strong lexical competition (e.g., Slowiaczek and Hamburger, 1992). Conversely, facilitation in form priming has been interpreted as a sign of weak lexical competition and an indication that sublexical facilitation dominates L2 processing of low-frequency words that are likely to have weak lexical representations in L2 (Gor and Cook, 2020; cf. Slowiaczek and Hamburger, 1992). For example, no significant inhibition was found in the group of L1 Dutch speakers for the L2 English minimal pairs, such as *flesh*-FLASH in cross-modal priming, although inhibition was observed in some individual participants (Broersma, 2012). In another study, L1 speakers of Russian consistently showed inhibition for phonological competitors with onset overlap in a phonological priming experiment, while L2 Russian speakers showed inhibition only for high-frequency word pairs, and facilitation for low-frequency word pairs (Gor and Cook, 2020).

To summarize, according to the FLR hypothesis, FLRs weaken lexical activation and competition in L2 lexical access and contribute to nonnative facilitation observed in phonological and orthographic priming tasks, with sublexical processes gaining more prominence in the L2. FLRs of newly acquired L2 words resist efficient consolidation and integration into lexical networks, which leads to the absence of the prime lexicality effect in L2 vocabulary training (Qiao and Forster, 2017) and to semantic inhibition when newly acquired lexical items in L2 with weak semantic representations serve as primes (Bordag et al., 2015, 2017a).

Fuzzy Form-Meaning Mappings Lead to Lexical Confusions

We claim that lexical confusions, a well-attested phenomenon in L2 (Laufer, 1990; Laufer, 1997; Hemchua and Schmitt, 2006; Cook and Gor, 2015; Cook et al., 2016), happen when the form-meaning connections are fuzzy due to the fuzzy form encoding of LRs. Fuzzy form-meaning mappings fall into two main categories, each with its own consequences for lexical processing. First, the FLR hypothesis identifies fuzzy form-meaning mappings that are weak and lead to unstable connections between word forms and meaning. Often, the source of such weak connections is the fact that L2 word forms and meanings are acquired at different times (with the meaning initially borrowed from L1), and form encoding is not deeply entrenched (Diependaele et al., 2013; Brysbaert et al., 2017). Form-meaning mappings may also be weak for newly acquired LRs with insufficient number of exposures and/or insufficient consolidation period. When form encoding of many LRs in the L2 mental lexicon is fuzzy, or approximative, forms and meanings of similar-sounding words are not robustly connected, and as a consequence, lexical activation is weak. The selection of the LR from the list of activated candidates becomes more effortful, leading to longer RTs in word recognition and the reversal of the phonological priming effect, as in Gor and Cook (2020), and also error-prone, leading to transient lexical confusions that are difficult to repair. Second, the form-meaning mappings may be incorrect, which leads to permanent lexical confusions and mistakes in meaning recognition (Laufer, 1990, 1997; Cook and Gor, 2015). This section will focus on the findings documenting both transient and definitive lexical confusions in L2 lexical processing.

In addition to weak form-meaning mappings in L2 – when forms and meanings of L2 words are acquired at different times due to developmental reasons and are loosely connected – fuzzy encoding of phonological forms may contribute to a different aspect of FLRs: incorrect form-meaning mappings leading to lexical confusions. The transient lexical confusion effect was demonstrated in a pseudo-semantic auditory priming experiment by Cook et al. (2016). In this experiment, the prime-target pairs were semantically related through a virtual competitor that had a phonological onset overlap with the

[4] This article does not address the issue of whether there are two separate lexicons corresponding to L1 and L2 or one lexicon encompassing both languages; at the same time, it supports the strong connection between the two lexicons.
[5] In lower-proficiency L2 speakers, L1 words may also be activated in the auditory modality. It is less obvious how strong the activation of L1 words is in highly proficient L2 speakers given that phonetic encoding serves as a strong cue, making it possible to tag the word as belonging to a particular language. With this in mind, the FLR hypothesis focuses on the L2 mental lexicon in the hope that more conclusive research on how phonetic tagging works (or does not work) in auditory perception will clarify the role of the L1 mental lexicon in L2 spoken word recognition.

target, as in *korova* ("*cow*")-*MOLOTOK* /malatok/ ("*hammer*"), with *molotok* "*hammer*" sharing its onset with *MOLOKO* / malako/ "*milk*." While L1 speakers showed the same RTs for pseudo-semantic primes as for completely unrelated primes, L2 speakers showed a significant delay in RTs in the pseudo-semantic priming condition. The authors argued that L2 learners temporarily considered the pair *korova-moloko* instead of *korova-molotok* that was presented to them and had difficulty with abandoning this association when the input became incompatible with a semantically related onset competitor, likely because the phonological representations of the target word and/or its competitor were not sufficiently robust. Note that the visual world eye-tracking studies on English monolinguals that use a similar design to capture the transient activation of the semantic network of the phonological onset competitor point to phonological associations as the locus of the activation that engages the semantic network (Yee and Sedivy, 2006) (e.g., if the participants heard the word *logs*, they fixated on *key* because of the partial activation of *lock* absent from the visual display). In L2 word recognition, phonological forms of LRs can be associated based on the similarity of their sublexical features (see section "**Spoken Word Recognition in L2: The Effect of FLRs on Lexical Competition in L2**" above).

What is specific to L2 word recognition and what emerges from both eye-tracking and priming studies focusing on lexical competition is the pervasiveness of L2 speakers' difficulty in resolving lexical competition. They do not efficiently and confidently identify the target word and abandon implausible competitors (Weber and Cutler, 2004; Cutler et al., 2006; Cook et al., 2016). According to the FLR hypothesis, the additional processing costs observed in L2 lexical processing are associated with the selection stage at which the LR is identified. Fuzzy encoding of form leads to fuzzy phonology-meaning mappings for individual LRs and impacts the functioning of FLRs in tasks involving word recognition. Thus, both visual world eye-tracking and pseudo-semantic priming experiments point to the same locus where transient lexical confusions that are difficult for L2 speakers to resolve originate – fuzzy phonological form encoding and fuzzy form-meaning mappings.

Further evidence for fuzzy mappings between word forms and meanings in L2 comes from experiments, which address the encoding of difficult phonological contrasts in orthographic representations of words, without involving any spoken input. While the main tenets of the FLR primarily concern spoken word encoding, fuzzy form-meaning mappings were also reported in experiments that involved no auditory input. A visual semantic-relatedness decision task (Ota et al., 2009) and a visual semantic categorization task (Ota et al., 2010) showed the effects of fuzzy phonological encoding which led to uncertainty regarding orthographic encoding of L2 words on the processing of L2 word meanings. The observation that incorrect semantic associations for English words, such as *key* and *rock,* emerged in the responses of L1 Japanese speakers who experience encoding problems with the /r/-/l/ contrast lends support to the idea that the LRs of *rock* and *lock* were fuzzy and not sufficiently separated in the mental lexicon. Crucially, while fuzzy phonological encoding is a result of perceptual categorization problems, auditory perception during the task completion could not be directly responsible for the semantic confusions reported by Ota et al. (2009, 2010). Accordingly, the results of the study speak in favor of FLRs being responsible for the confusions.

Form Prominence in L2

This section will provide a quick review of the findings regarding form prominence in different populations of speakers that points to the same source – fuzziness in lexical encoding leading to weak form-meaning connections for less familiar words. Indeed, L2 speakers show a stronger preference for form-based associations for a number of reasons, all traceable to fuzzy encoding. The idea that the L2 lexicon is qualitatively different from the L1 lexicon was originally proposed by Meara (1978, 1983, 1984) based on a word association (WA) study with monolingual speakers and L2 learners of French. According to Meara, phonological links between words tend to play a much more prominent organizing role in the L2 mental lexicon than in the L1 mental lexicon. Several strands of research have reported since then that word form, whether spoken or written, has a greater prominence in L2 than in L1. Evidence in favor of form prominence in L2 mainly comes from WA studies (Jiang and Zhang, 2021); however, additional insights can be gained from form-based facilitation observed in morphological (Heyer and Clahsen, 2015; Li et al., 2017) and phonological priming experiments (Gor and Cook, 2020) and also from the comparison of memory for surface linguistic detail in L1 and L2 in longer texts (Bordag et al., 2021c). It should be noted that the majority of WA responses both by L1 and L2 speakers are still semantic in nature, indicating the importance of semantic networks both in L1 and L2 (Jiang and Zhang, 2021).

In WA studies, participants are typically asked to respond with one or more words to a given stimulus word. The type (and sometimes the number) of participants' responses in WA tasks has been in the focus of L2 research for many decades (for an overview of WA in L2 research, see Fitzpatrick and Thwaites, 2020). Although the evidence accrued in this line of research, both with respect to L1 and L2, sometimes yields contradicting results, some patterns have been consistent. For instance, whether participants are more likely to respond to a given word either with a clang response (*mouse* – *mouth*), a syntagmatic response (*sit* – *chair*), or a paradigmatic response (*eagle* – *bird*) seems to be influenced by several factors. The probability of clang, or orthographic/phonological responses is increased in younger participants (Namei, 2004), if the cue word is relatively unfamiliar or newly acquired (Söderman, 1993; Wolter, 2001), or if the task is performed in L2 (Wolter, 2001; Fitzpatrick, 2006; Norrby and Håkansson, 2007; Jiang and Zhang, 2021). All these factors indicate that form-based responses are more likely in case of incomplete or unstable, that is, fuzzy representations. Syntagmatic or position-based responses are more likely to occur if participants respond in their L2 (Norrby and Håkansson, 2007; Zareva, 2007; Håkansson and Norrby, 2010) and have low L2 proficiency (Zareva and Wolter, 2012; Khazaeenezhad and Alibabaee, 2013). Paradigmatic, or meaning-based responses, including synonym responses, are more commonly observed for participants using their L1 (Fitzpatrick, 2006; Fitzpatrick and Izura, 2011) or in more proficient L2 speakers (Khazaeenezhad and Alibabaee, 2013), especially if they know the cue word well enough to use it in a sentence (Wolter, 2001), if they are expert users (L1 or advanced L2) of the language (Zareva, 2007; Jiang and Zhang, 2021), or if they are older [Namei, 2004, e.g., when they are adults as opposed to children (Cremer et al., 2011)].

Some of the mentioned factors are shared across L1 and L2, for instance, the observation that the better the word is known, or the older the speaker is, the more paradigmatic responses can be expected. However, some factors that change the proportion of responses are specific to L2: heritage L2 speakers (Kim, 2013) and more proficient speakers (Söderman, 1993; Zareva and Wolter, 2012; Khazaeenezhad and Alibabaee, 2013) are more likely to produce paradigmatic responses, while the proportion of syntagmatic responses is increased for speakers learning their L2 outside the target language environment (Håkansson and Norrby, 2010) or as a foreign rather than a second language (Norrby and Håkansson, 2007).

While semantic relations are at the core of the organization of the lexicon (as evidenced by the fact that semantic/paradigmatic/meaning-based responses are most prevalent across all studies, cf. Fitzpatrick and Thwaites, 2020), form- or syntagmatic (position-based) associations are more frequent, especially at lower acquisition stages (in children more than in adults), with less familiar words, and, importantly, in L2 compared to L1. Form prominence, as reported in the WA studies, can thus be related to fuzziness in LRs. Jiang and Zhang (2021) conclude that form is a more relevant factor in organizing the lexicon in L2 than in L1.

Additional support for form prominence and further evidence of the special status of form-based associations in the L2 lexicon comes from morphological priming studies that reveal reliable, purely form-based, orthographic priming effects in L2, while these effects are typically much weaker or missing in L1. For instance, Heyer and Clahsen (2015) observed facilitation in masked priming for purely form-related items (*career-CAR*) only in L2, while the facilitatory priming effects of the same size were found both in L1 and in L2 for morphologically and semantically related items (*darkness-dark*; for similar effects for compounding, see Li et al., 2017; however, for contrary findings see Diependaele et al., 2011). Form-based facilitation was also reported in phonological priming with onset overlap between the prime and the target (Gor and Cook, 2020).

Several factors can contribute to form prominence in L2. First, L2 learners can rely on the already existing L1 lexical system; therefore, when new L2 word forms need to be added, the already existing network of semantic representations can be engaged. Accordingly, the focus of acquisition is on the word forms that need to be stored and mapped onto the existing semantic representations. Since the semantic and the word form systems do not develop in parallel, they are less tightly connected in L2. Second, spurious activation of additional irrelevant competitors in L2 due to fuzzy phonolexical representations leads to more distributed and weaker activation of form-meaning connections. As a result, semantic representations that are activated through L2 word forms are activated less strongly than when the same representations are activated through L1 forms. Consequently, word meanings are less activated when processing L2, which foregrounds the form system and contributes to its prominence. Form prominence thus arises in L2 because of the reduced engagement of the semantic network compared to L1. The source of form prominence in L2 can also be traced to the specific role of episodic memory in L2 lexical processing, and the effort of L2 speakers to temporarily store rich detailed linguistic information because of their inefficiency in encoding and consolidating it for long-term storage. In the next section, we discuss how FLRs relate to different memory accounts.

Fuzzy Lexical Representations, Episodic Memory, and the Complementary Learning Systems

FLRs are encoded and stored in memory, as are any LRs. However, fuzziness resists efficient memory consolidation and, therefore, may be responsible for the differences in how different memory systems subserve L1 and L2 lexicons. Several proposals underlying the differences in L1 and L2 lexical memory organization exist in the literature.

According to the episodic L2 hypothesis by Forster and colleagues (Jiang and Forster, 2001; Witzel and Forster, 2012), L2 words are represented in a different memory system than L1 words. In their studies, episodic recognition tasks elicit masked translation priming effects from L2 to L1 for "studied" L1 words but not for "unstudied" L1 words, whereas lexical decision tasks elicit asymmetrical effects in facilitatory priming from L1 to L2 but not from L2 to L1. Because L2 primes appear to be activated only in tasks requiring access to episodic memory, the authors conclude that all L2 words must be represented in the episodic memory system (or some other yet unspecified L2-specific memory system), whereas L1 words are stored in lexical memory. The effects, on which the episodic L2 hypothesis claims are based, turn out to be volatile: they have been reported for masked translation priming only under specific presentation conditions (Jiang and Forster, 2001; Witzel and Forster, 2012), but were absent in overt translation and semantic priming with two different SOAs and L1 Dutch-L2 English participants (Schoonbaert et al., 2009). Also, while no L2-L1 translation priming was observed for L1 Chinese speakers of English (Jiang and Forster, 2001; Witzel and Forster, 2012), an L2-L1 translation priming effect was observed for low-proficient L1 Korean learners of English (Lee et al., 2018). The limited evidence in support of the episodic L2 hypothesis seems insufficient to corroborate a major claim that L2 speakers rely on a different memory type in lexical processing compared to L1 speakers. More importantly, episodic memory is characterized by rapid decay over time and is only engaged in lexical processing within a short time span (Takashima et al., 2017) and, therefore, cannot replace long-term memory for the purpose of storing L2 LRs.

While episodic memory, by definition, cannot subserve long-term lexical storage, which challenges the viability of the episodic L2 hypothesis, a number of observations point to a greater reliance of L2 speakers on episodic memory in tasks engaging episodic representations of recently activated L2 words (Francis and Strobach, 2013; Bialystok et al., 2020). In recognition memory tasks, where participants have to recall recently studied stimulus words, L2 speakers perform better when recalling words in their L2, a less proficient language, and better than L1 speakers recalling words in their native language (Francis and Strobach, 2013). L2 speakers are also less likely to develop false memories of semantic lures (e.g., incorrectly recalling the word *needle* after studying the words *thread, pin, point*, and *sharp*) compared to monolinguals; however, they are more susceptible to phonological (form-based) memories (Bialystok et al., 2020), supporting the idea of L2 word form prominence discussed in Section "**Form Prominence in L2**".

The FLR hypothesis maintains that certain properties of L2 lexical representations and the way they are acquired make them more likely to benefit from the demands imposed by episodic tasks. For example, form prominence (Jiang and Zhang, 2021) in the L2 lexicon (see discussion in Section "**Form Prominence in L2**" above) may help explain the priming asymmetry observed in the lexical decision tasks and the episodic memory tasks. In the episodic recognition tasks, participants have to identify the "old" words that they have studied vs. "new" words that they have not seen in the training set (but that they know). The task can be accomplished based on form recognition alone, without necessarily accessing the meaning. Thus, this task, where form recognition is critical, may be driven by form-based connections. Furthermore, greater episodic distinctiveness of low-frequency L2 words may result from a stronger novelty effect because frequency differences are subjectively greater in L2 (Francis and Strobach, 2013). The most likely reason why L2 speakers hold on to detailed episodic LRs is that they are inefficient at rapid and compact linguistic encoding of fuzzy LRs, and consolidation takes a longer time in L2. This last argument is supported by the Complementary Learning Systems (CLS) account (McClelland et al., 1995; Norman and O'Reilly, 2003) discussed below.

Evidence from word learning studies using consolidation paradigms suggests that there might be some differences between how L1 and L2 words are initially encoded in memory. According to the CLS account (McClelland et al., 1995; Norman and O'Reilly, 2003), memory traces are initially formed in the hippocampal and medio-temporal lobe (MTL) systems, which encode novel experiences (e.g., new words) immediately and support episodic memories. Over a consolidation period, these experiences are transformed into more stable representations supported by neocortical regions (temporal lobes). In L1 word learning studies, lexical competition exerted by newly learned words (e.g., *banara*) on the recognition of existing words (e.g., *banana*) after a period of consolidation results in inhibition, which is usually taken as evidence that a new word has been integrated into the mental lexicon (Gaskell and Dumay, 2003; Dumay and Gaskell, 2007; for a similar account, see Leach and Samuel, 2007). In L2, similar word learning paradigms yield different results. For example, Qiao and Forster (2017) observed that L2 speakers failed to show an inhibitory prime lexicality effect in a masked priming experiment in contrast to L1 learners (Qiao et al., 2009; Qiao and Forster, 2017). Instead, facilitatory (*banara* primed BANANA) and not inhibitory priming effects were observed in L2, suggesting that new words are not lexicalized or integrated into the lexical network in the same way in L2 as in L1, likely, because of less efficient encoding, that is, fuzziness.

Importantly, learning in the context of the CLS model depends on prior knowledge, or schemas – networks of interconnected, already existing neocortical representations that affect how new information is organized (Palma and Titone, 2020). The length of time during which new knowledge remains reliant on the MTL structures may depend on how well it fits a preexisting schema (Lindsay and Gaskell, 2010). Since L2 word learning is, by definition, subsequent to L1 word learning, encoding and integration of L2 words into existing memory may be mediated by the already existing schemas established during learning of the L1. Havas et al. (2018) explored this idea by examining the impact of existing phonological and semantic schemas on consolidation effects for words with familiar vs. unfamiliar semantics and with L1- vs. L2-like phonology. The authors found that both phonological and semantic aspects of word learning were enhanced by similarities with the existing schemas. For example, L1-like words were remembered better than L2-like words on the day of training (cf. the results of the study by McKean et al., 2013, in which children were also more accurate at a fast-mapping task for words with the phonotactics similar to their native language). These two findings suggest that the rate of initial encoding and later consolidation may differ for L1 vs. L2 word learning. Moreover, the relative engagement of episodic and semantic memory networks differs depending on whether only the word form or both the word form and its meaning were learned (Takashima et al., 2017). The fact that the reliance on episodic memory is increased for novel nonnative phonology and semantics provides support for the increased role of episodic LRs in L2 observed in the episodic memory tasks discussed above and for the association of episodic memory engagement and imprecise, or fuzzy lexical encoding.

HOW DO FLRS DEVELOP OVER TIME – IS FUZZINESS REDUCED?

The degree of fuzziness of LRs is related to their acquisition stage. Recently established and/or infrequently used representations are fuzzier than well-established, frequently used representations. The present article leaves out a detailed discussion of the developmental aspects of lexical representations due to space limitations. The developmental trajectories of individual LRs that depend on the linguistic properties of the LR and the learning context are captured by the Ontogenesis Model of the Lexical Representation (OM, Bordag et al., 2021a,b), and we refer the reader to these publications. The OM describes the ontogenesis of the LR within its phonological, orthographic and semantic domains, the mapping between them and with respect to their engagement in their corresponding networks. The OM assumes that most L2 LRs are fuzzy and that the ontogenetic curve of their development does not reach the *optimum* (i.e., the ultimate stage of their attainment with optimal encoding) in one or more dimensions. As has been discussed above, depending on the source of fuzziness, L2 LRs will be more or less amenable to more robust encoding with more input. The most "resistant" FLRs involve difficult L2 phonological contrasts or segments, which depends on a given L1-L2 combination. Such FLRs may continue to show poor phonolexical encoding even after extensive exposure to the spoken word. The OM focuses on unique developmental trajectories of individual LRs, hypothesizes that there is a developmental curve for each of the domains of a LR, phonological, orthographic and semantic, and proposes individual ontogenetic scenarios depending on linguistic and contextual factors [see, especially, Figure 5A,B in Bordag et al., (2021a)]. It extends the FLR hypothesis to the developmental domain.

The FTT, which is concerned with differences in the representation and processing of memories and decision-making in novices and experts (Brainerd and Reyna, 2002), takes a different approach to the role of L2 proficiency, or the acquisitional stage, in linguistic encoding that can be also applied to lexical encoding. The ability to derive the "gist" of a linguistic message seems to include several components: The ability to process the message and efficiently encode it verbatim, and then to extract the core meaning of the message, or the summary of its key points, and encode it as a compact "take-away" message. For spoken speech, the processing takes place in real time, thereby creating a high processing load that lower-proficiency L2 speakers cannot handle. It appears that the "universal" strategy of L2 speakers is to keep in memory a rich episodic representation that includes detailed, albeit imprecisely encoded form representations rather than to quickly package the semantic content of the received message. Note that low-precision phonological encoding with uncategorized raw phonetic details resisting consolidation is also characteristic of FLRs. Recent research on L2 text processing supports these assumptions. In a cued sentence recall procedure, Sampaio and Konopka (2013) showed that L1 and L2 speakers recall better the verbatim phrasing of sentences with nonpreferred lexical items (e.g., STRUCK vs. HIT) and are thus more sensitive to synonymous lexical substitutions in such sentences. Bordag et al. (2021c) directly compared memory for surface linguistic detail in L1 and L2 in longer texts and showed that L2 learners outperform L1 speakers not only in memory for lexical detail (cf. Sampaio and Konopka, 2013), but also for structural information. These findings indicate that L2 learners as novices are fixated on surface linguistic information (i.e., form-related), probably because they have to rely on reduced or inefficient access to the knowledge available in the semantic store.

HOW CAN THE CONSTRUCT OF FLRS HELP TO IMPROVE THE EXISTING MODELS OF LRS AND LEXICAL ACCESS?

The FLR hypothesis proposes an extension to the existing models of L2 word recognition that are supported by network simulations, such as BIA+ and Multilink (Dijkstra and Van Heuven, 2002; Dijkstra et al., 2019) – an addition of the quality of encoding for different layers, as a parameter that interacts with input frequency rather than being its product. First, it should be noted that BIA+ and Multilink are developed for orthographic input, and thereby obviate one of the core issues in L2 LRs of spoken words – problems with phonological categorization. While the problem of proper orthographic encoding exists in languages with deep orthography, it is conceivably more amenable to training than phonological categorization that poses continuous problems for L2 learners, both for nonword segments (e.g., PAM-L2, Best and Tyler, 2007) and in phonolexical encoding (e.g., Darcy et al., 2012, 2013; Daidone and Darcy, 2021). Furthermore, these models represent an ideal L2 speaker, whereas in reality, even advanced L2 speakers may store inaccurate orthographic representations of words. Thus, neither BIA+ nor Multilink builds the quality of encoding into different levels of the lexical representation (different layers in the model) as an independent variable [e.g., contributing to the resting activation levels (Dijkstra and Van Heuven, 2002; Dijkstra et al., 2019)]. Rather, the quality of form encoding depends on the word frequency and its frequency ranking in the corpus (Dijkstra et al., 2019, p. 661). Multilink establishes the strength of the links between the levels of form and meaning in the lexical representation by taking into account the L2 proficiency level under the same assumption that L2 proficiency is associated with exposure to L2, that is, with input frequency.[6] The majority of the existing computational models of L2 word recognition are not concerned with modeling the quality of lexical representations in a developing L2 lexicon within the L1 neural environment depending on AoA. In contrast, Zhao and Li (2007, 2010) have implemented three variants of a self-organizing neural network model: with simultaneous, delayed, or late AoA of the L2. Their main finding was that when the AoA was early, then functionally distinct lexical representations could be established for both languages; however, if the AoA was late, the model was unable

[6]Also, while these models focus on the interactions between the L2 and L1 lexicons and lexical representations, the FLR hypothesis is mainly concerned with L2, even though it assumes that L2 and L1 mental lexicons are interdependent given that the L2 lexicon develops when the L1 lexicon is already in place. Accordingly, the mapping issues between the L1 and L2 lexical representations are relevant for the structure of L2 LRs because the L1 and L2 lexicons coexist in one mental space and interact.

to recruit sufficient resources to entirely remap the existing L1 lexical network. L2 phonological representations were forced into the spaces unoccupied by L1, where accurate access and retrieval was made difficult, and chances of confusion were high because of how densely the conceptual space was populated – LRs were imprecise, and the boundaries between them were fuzzy. These findings are in agreement with the FLRs hypothesis, as it focuses on adult (i.e., late) L2 lexical acquisition and processing. Furthermore, the FLR hypothesis proposes a complex set of assumptions – the inherent variability, ambiguity, and imprecision of LRs in L2, that is, their fuzziness should be accounted for not only in the connection weights representing the qualitative aspects of mappings, but also in the direction and the quantity of these mappings, as well as their dynamic nature (see also Duta and Plunkett, 2021 about building a dynamic mapping between phonological and semantic representations in a bottom-up fashion into a neural model of spoken word recognition). The FLR hypothesis argues that form-meaning mappings in L2 are subserved by diffused activation engaging a greater number of form-level nodes due to larger competitor sets, as well as nonnative-like patterns of activation drawing on fuzzy phonolexical encoding. One possibility to build fuzziness in L2 phonolexical representations into a computational model of L2 spoken word recognition is to add the intake layer (see section "**FLRs, Lexical Frequency, and Lexical Entrenchment**"), where words will have ambiguous or incorrect encoding specific for a particular L1-L2 combination, to the model above the input layer (cf. Bordag et al., 2021a,b). The assumptions of the FLRs hypothesis, and specifically, the role of encoding at different levels of the LR and the consequences of fuzzy encoding for establishing form-meaning connections, and more broadly, for word storage and retrieval could be used to further develop the existing models of bilingual LRs and word recognition and to model L2-specific features of LRs, such as fuzziness, or imprecise encoding and mappings between the levels of LRs.

FINAL REMARKS

We have reviewed a number of phenomena reported in L2 word processing that point to the same origin – problems with lexical encoding in L2. The quality of encoding, the core property of lexical representations, according to the FLRs hypothesis, has also been evoked in several influential approaches to written word recognition, such as the lexical entrenchment hypothesis (Diependaele et al., 2013; Brysbaert et al., 2017) and the lexical quality hypothesis (Perfetti and Hart, 2002; Perfetti, 2007). While both the FLR hypothesis and the lexical entrenchment hypothesis acknowledge the role of input frequency in the quality of lexical encoding, they diverge in that the FLRs hypothesis argues for a certain independence of the quality of lexical encoding from word frequency. Indeed, it is true that L2 speakers are exposed to reduced L2 input, a major source of lexical fuzziness, and it also comes at a later age, and with the lexical system of L1 already in place – for all these reasons, L2 LRs are expected to be imprecisely encoded, or fuzzy. At the same time, according to the FLRs hypothesis, whose primary focus is spoken word storage and retrieval, if the source of fuzziness is a problematic phoneme or phonological contrast that entails difficulties in L2 perception (and typically, production as well), increased exposure to the word may not improve the quality of its phonolexical encoding or will improve it at a much slower rate. Given that weak phonolexical encoding leads to fuzzy form-meaning mappings, it is to be expected that the LRs of such words will remain fuzzy with increased input frequency. Meaning encoding may also develop slower in L2 because of the complex relations between the senses of L2 words and their L1 counterparts. The existing L1 semantic mappings may resist remapping as a result of decline in brain plasticity or because lower-proficiency L2 speakers fail to process spoken input efficiently and to take advantage of the context to build complex semantic representations of L2 words.

There is an important caveat to the claims that FLRs are observed, or even observable, in spoken word recognition tasks where deficits in online perception can be responsible for the outcomes ascribed to the stored LRs. While it is indeed impossible to determine whether the tasks using spoken words as input show the effects of lexical encoding of stored LRs, or online processing difficulties, or both, several data sets point to the unique contribution of the properties of stored LRs to the observed effects. These data sets, discussed above, rely on cross-modal priming that is argued to engage central representations rather than access representations (e.g., Broersma, 2012) and on semantic relatedness and categorization tasks that use visually presented words with underlying confusable phonological contrasts (Ota et al., 2009, 2010).

To summarize, the FLR hypothesis maintains, the quality of lexical encoding is the core property of L2 LRs that deserves further study. For example, initial problems with difficult L2 phonological contrasts leading to fuzzy L2 phonolexical encoding may persist over time, and phonolexical encoding is not improved with additional input. While such phonolexical encoding problems have obvious consequences – weak or incorrect form-meaning mappings in the L2 mental lexicon – they also impact all aspects of lexical retrieval: lexical activation, competition, and selection. Crucially, the encoding at all levels of FLRs may undergo later remapping resulting in new sources of fuzziness, as in semantic reconfiguration when new meanings are added. By using this approach, in which the quality of lexical encoding is not a direct product of more encounters with the word, but rather a combination of the linguistically driven encoding difficulty with input frequency, the models of LR and lexical processing will make it possible to explore how the quality of encoding and input frequency interact for different lexical units in L2.

No one set of behavioral evidence can fully test the FLR hypothesis given that it is making inferences about lexical representations that are not directly open to observation based on the processing data. Several lines of research and kinds of evidence need to be considered, which calls for a comprehensive program rather than a single test. Many studies reviewed in the manuscript test the FLR hypothesis; however, it is only by looking at the pattern of findings across several studies that we can claim that the FLR hypothesis receives empirical support. There are several kinds of evidence in support of FLRs identified so far:

- Lexical confusions in auditory word recognition (Darcy et al., 2012, 2013; Cook and Gor, 2015);
- Accuracy and speed in word recognition that suggest nonnative patterns in lexical activation, competition, and selection:
    - o A reversal of the phonological priming effect for less frequent/less familiar words from inhibition to facilitation interpreted as evidence of weak lexical competition and strong reliance on sublexical processing (Gor and Cook, 2020);
    - o A reversal of the semantic priming effect from facilitation to inhibition in semantic priming for newly learned words (Bordag et al., 2015, 2017a) and in pseudo-semantic priming for less frequent/familiar words (Cook et al., 2016);
- Semantic confusions in visual word processing in semantic relatedness and categorization tasks (Ota et al., 2009, 2010);
- Lexical confusions and overreliance on the sentence context to accept context-mismatching phonological neighbors of the target words with fuzzy phonolexical representations (Chrabaszcz and Gor, 2014, 2017).

These observed effects point to two main sources of fuzziness in L2 LRs: phonological encoding problems and semantic encoding problems. The main challenge in testing the FLR hypothesis is to tease apart the loci of fuzziness – the representational level or the perceptual level – which are confounded in experiments that rely entirely on auditory input. One way to differentiate the component of online perceptual difficulties from the representational deficits is to use orthographic input instead. This has been done in semantic relatedness and categorization tasks that engaged visual word processing to show semantic confusions (Ota et al., 2009, 2010) and in a visual semantic priming task (Bordag et al., 2015, 2017a). Another possibility would be to use visual primes in a cross-modal or visual masked priming experiment rather than auditory primes. The use of orthographic input would be justified for highly controlled orthographic stimuli to avoid potential orthographic encoding difficulties.

New and more focused research will provide additional behavioral and neurolinguistic evidence supporting fuzziness at the encoding, that is, the representational level in addition to the perceptual/processing level. Lexical confusions associated with orthographic encoding problems and the phonology/orthography interface also need to be tested. An ERP study of N400 effects for phonologically confusable incongruent lexical substitutions and the role of different factors contributing to the lack of sensitivity of L2 speakers to such substitutions will test the role of FLRs in sentence processing. Additionally, different dimensions of fuzziness of LRs can be further explored by comparing the performance of multiple native language groups on multiple phonological contrasts (similar to the approach of Barrios and Hayes-Harb, 2021). Word training studies can manipulate the hypothesized degree of fuzziness for L2 words (e.g., based on phonological contrasts or L1 transfer predictions) to examine how FLRs change over time and what consequences fuzziness has for their long-term maintenance in memory.

In the future, we plan to broaden the claim regarding fuzzy L2 lexical representations to potentially involve less explored nonlinguistic extensions of lexical representations that are processed and encoded in the sensory-motor and emotional systems and rely on different sensory pathways. Sensory pathways and emotions appear to be coactivated in parallel with the lexical representation in L1 (Altmann et al., 2012; Kuperman et al., 2014) and to a significantly weaker degree, in L2 (Sulpizio et al., 2019; see also Conrad et al., 2011; see however, Ponari et al., 2015).

In Section "**How Do FLRs Develop Over Time – Is Fuzziness Reduced?**," we argued that L2 lexical representations and lexical processing seem to be more oriented toward the surface, form level and we related this observation to a more general difference between novices and experts, as described in cognitive theories, such as FTT. We maintained that this orientation could be due to the fact that L2 learners cannot access the information that is stored at the semantic level to the same extent as L1 speakers. In addition, studies on emotions and L2 report greater emotional and cognitive distance in L2 compared to L1 (e.g., Harris et al., 2003; Puntoni et al., 2009; Caldwell-Harris, 2014; Hadjichristidis et al., 2015; Hayakawa et al., 2017). These findings lend themselves to various interpretations. First, this increased emotional and cognitive distance in L2 could be a consequence of typically different acquisition contexts, in which L1 and L2 are acquired: L1 is acquired in emotionally varied and rich contexts, while L2 is often acquired in a more emotionally neutral classroom environment (Ivaz et al., 2016; Dylman and Bjärtå, 2018). Another, not mutually exclusive explanation is based on the claim that L2 processing is more taxing on cognitive resources compared to L1 processing (see, e.g., Morishima, 2013), which results in limited resources available for the processing of emotions (see Yates et al., 2010).

Rather than explaining emotional distance through the cognitive load and resource allocation in L2, the FLR hypothesis suggests that due to the fuzziness at the form level that results in diffused spreading of activation among not closely related word forms and weak form-meaning mappings, less activation reaches the semantic network. As a consequence, the sensory-motor features associated with the word semantics do not become sufficiently activated in L2. It is likely that emotionally relevant representations can be activated both within the lexical-semantic system (e.g., *darkness* – *fear*/*danger*) and nonlexical, sensory-motor, and emotional systems. It is a question for future research to explore the hypothesis that L2 lexical representations are more emotionally "flat," because they are only weakly connected to the sensory-motor and emotional systems and/or because less activation is available in the semantic and sensory-motor systems to reach the corresponding features due to the fuzziness effects on the form and form-meaning mapping levels.

DATA AVAILABILITY STATEMENT

No original data were reported in the article; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

All authors contributed to the article and approved the submitted version. All authors conceived the original idea and contributed with drafting separate subsections of the paper, editing and proofreading. KG developed the theory and took the lead in writing the manuscript. KG coordinated and directed the project. SC created supporting visual content. DB provided financial support.

FUNDING

This article was funded by the Publication Fund of the University of Leipzig. The work on this manuscript by KG was supported by funding from the Slavic and East European Language Resource Center (SEELRC) at Duke University. The contribution of Anna Chrabaszcz was funded by the Center for Language and Brain NRU Higher School of Economics, RF Government Grant, ag. No. 14.641.31.0004.

ACKNOWLEDGMENTS

The authors acknowledge support from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing.

REFERENCES

Abrahamsson, N., and Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: listener perception versus linguistic scrutiny. *Lang. Learn.* 59, 249–306. doi: 10.1111/j.1467-9922.2009.00507.x

Alavi, S. M., and Akbarian, I. H. (2012). The role of vocabulary size in predicting performance on TOEFL reading item types. *System* 40, 376–385. doi: 10.1016/j.system.2012.07.002

Altmann, U., Bohrn, I. C., Lubrich, O., Menninghaus, W., and Jacobs, A. M. (2012). The power of emotional valence—from cognitive to affective processes in reading. *Front. Hum. Neurosci.* 6:192. doi: 10.3389/fnhum.2012.00192

Barrios, S. L., and Hayes-Harb, R. (2020). Second language learning of phonological alternations with and without orthographic input: evidence from the acquisition of a German-like voicing alternation. *Appl. Psycholinguist.* 41, 517–545. doi: 10.1017/S0142716420000077

Barrios, S. L., and 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. *Front. Commun. Lang. Sci.* 6:144. doi: 10.3389/fcomm.2021.689470

Barrios, S., Jiang, N., and Idsardi, W. J. (2016). Similarity in L2 phonology: Evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Sec. Lang. Res.* 32, 367–395. doi: 10.1177/0267658316630784

Baxter, P., Bekkering, H., Dijkstra, T., Droop, M., van den Hurk, M., and Léoné, F. (2021a). Grounding second language vocabulary instruction in cognitive science. *Mind Brain Educ.* 15, 24–34. doi: 10.1111/mbe.12278

Baxter, P., Droop, M., van den Hurk, M., Bekkering, H., Dijkstra, T., and Léoné, F. (2021b). Contrasting similar words facilitates second language vocabulary learning in children by sharpening lexical representations. *Front. Psychol.* 12:2648. doi: 10.3389/fpsyg.2021.688160

Best, C. T., and Tyler, M. (2007). "Nonnative and second-language speech perception," in *Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege*. eds. O. S. Bohn and M. Munro (Amsterdam: John Benjamins), 13–34.

Bialystok, E., Dey, A., Sullivan, M. D., and Sommers, M. S. (2020). Using the DRM paradigm to assess language processing in monolinguals and bilinguals. *Mem. Cogn.* 48, 870–883. doi: 10.3758/s13421-020-01016-6

Bordag, D., Gor, K., and Opitz, A. (2021a). Ontogenesis model of the L2 lexical representation. *Biling. Lang. Cogn.* 1–17. doi: 10.1017/S1366728921000250 [Epub ahead of print]

Bordag, D., Gor, K., and Opitz, A. (2021b). Refining key concepts of the Ontogenesis Model of the L2 lexical representation. *Biling. Lang. Cogn.* 1–6. doi: 10.1017/S1366728921000894

Bordag, D., Kirschenbaum, A., Rogahn, M., Opitz, A., and Tschirner, E. (2017a). Semantic representation of newly learned L2 words and their integration in the L2 lexicon. *Stud. Second. Lang. Acquis.* 39, 197–212. doi: 10.1017/S0272263116000048

Bordag, D., Kirschenbaum, A., Rogahn, M., and Tschirner, E. (2017b). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Second. Lang. Res.* 33, 147–178. doi: 10.1177/0267658316665879

Bordag, D., Kirschenbaum, A., Tschirner, E., and Opitz, A. (2015). Incidental acquisition of new words during reading in L2: inference of meaning and its integration in the L2 mental lexicon. *Biling. Lang. Cogn.* 18, 372–390. doi: 10.1017/S1366728914000078

Bordag, D., Opitz, A., Polter, M., and Meng, M. (2021c). Non-native readers are more sensitive to changes in surface linguistic information than native readers. *Biling. Lang. Cogn*. 24, 599–611. doi: 10.1017/S1366728920000772

Bordag, D., Opitz, A., Rogahn, M., and Tschirner, E. (2018). The strength of meaning: representations of new words with different properties in the L2 semantic network. *Men. Lex.* 13, 74–104. doi: 10.1075/ml.17005.bor

Brainerd, C. J., and Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Curr. Dir. Psychol. Sci.* 11, 164–169. doi: 10.1111/1467-8721.00192

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Lang. Cogn. Process.* 27, 1205–1224. doi: 10.1080/01690965.2012.660170

Broersma, M., and Cutler, A. (2008). Phantom word activation in L2. *System* 36, 22–34. doi: 10.1016/j.system.2007.11.003

Broersma, M., and Cutler, A. (2011). Competition dynamics of second-language listening. *Q. J. Exp. Psychol.* 64, 74–95. doi: 10.1080/17470218.2010.499174

Brysbaert, M., Lagrou, E., and Stevens, M. (2017). Visual word recognition in a second language: a test of the lexical entrenchment hypothesis with lexical decision times. *Biling. Lang. Cogn.* 20, 530–548. doi: 10.1017/S1366728916000353

Bylund, E., Abrahamsson, N., Hyltenstam, K., and Norrman, G. (2019). Revisiting the bilingual lexical deficit: the impact of age of acquisition. *Cognition* 182, 45–49. doi: 10.1016/j.cognition.2018.08.020

Bylund, E., Hyltenstam, K., and Abrahamsson, N. (2021). Age of acquisition–not bilingualism–is the primary determinant of less than nativelike L2 ultimate attainment. *Biling. Lang. Cogn.* 24, 18–30. doi: 10.1017/S1366728920000188

Caldwell-Harris, C. L. (2014). Emotionality differences between a native and foreign language: theoretical implications. *Front. Psychol.* 5:1055. doi: 10.3389/fpsyg.2014.01055

Carr, T. H., and Dagenbach, D. (1990). Semantic priming and repetition priming from masked words: evidence for a center-surround attentional mechanism in perceptual recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 341–350.

Chrabaszcz, A., and Gor, K. (2014). Context effects in the processing of phonolexical ambiguity in L2. *Lang. Learn.* 64, 415–455. doi: 10.1111/lang.12063

Chrabaszcz, A., and Gor, K. (2017). Quantifying contextual effects in L2 processing of phonolexically ambiguous and unambiguous words. *Appl. Psycholinguist.* 38, 909–942. doi: 10.1017/S0142716416000497

Conrad, M., Recio, G., and Jacobs, A. M. (2011). The time course of emotion effects in first and second language processing: a cross cultural ERP study with German–Spanish bilinguals. *Front. Psychol.* 2:351. doi: 10.3389/fpsyg.2011.00351

Cook, S. V. (2012). Phonological form in L2 lexical access: friend or foe? Doctoral dissertation. University of Maryland.

Cook, S. V., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *Men. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Corder, S. P. (1967). "The significance of learner's errors," in *International Review of Applied Linguistics in Language Teaching, Vol. 4*. eds. B. Malmberg and G. Nickel (Heidelberg: Julius Groos Verlag).

Cremer, M., Dingshoff, D., de Beer, M., and Schoonen, R. (2011). Do word associations assess word knowledge? A comparison of L1 and L2, child and adult word associations. *Int. J. Biling.* 15, 187–204. doi: 10.1177/1367006910381189

Cutler, A., and Otake, T. (2004). Pseudo-homophony in non-native listening. *J. Acoust. Soc. Am.* 5:2392. doi: 10.1121/1.4780547

Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *J. Phon.* 34, 269–284. doi: 10.1016/j.wocn.2005.06.002

Dagenbach, D., Carr, T. H., and Barnhardt, T. M. (1990a). Inhibitory semantic priming of lexical decisions due to failure to retrieve weakly activated codes. *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 328–340. doi: 10.1037/0278-7393.16.2.328

Dagenbach, D., Horst, S., and Carr, T. H. (1990b). Adding new information to semantic memory: how much learning is enough to produce automatic priming? *J. Exp. Psychol. Learn. Mem. Cogn.* 16, 581–591. doi: 10.1037/0278-7393.16.4.581

Daidone, D., and Darcy, I. (2021). Vocabulary size is a key factor in predicting second-language lexical encoding accuracy. *Front. Psychol.* 12:688356. doi: 10.3389/fpsyg.2021.688356

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Men. Lex.* 8, 372–420. doi: 10.1075/ml.8.3.06dar

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., et al. (2012). Direct mapping of acoustics to phonology: on the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second. Lang. Res.* 28, 5–40. doi: 10.1177/0267658311423455

DeKeyser, R. M. (2012). "Age effects in second language learning," in *Handbook of Second Language Acquisition*. eds. S. Gass and A. Mackey (London: Routledge), 442–460.

Díaz, B., Mitterer, H., Broersma, M., and Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: from acoustic-phonetic analysis to lexical access. *Learn. Individ. Differ.* 22, 680–689. doi: 10.1016/j.lindif.2012.05.005

Diependaele, K., Duñabeitia, J. A., Morris, J., and Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *J. Mem. Lang.* 64, 344–358. doi: 10.1016/j.jml.2011.01.003

Diependaele, K., Lemhöfer, K., and Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: a lexical entrenchment account. *Q. J. Exp. Psychol.* 66, 843–863. doi: 10.1080/17470218.2012.720994

Dijkstra, A. F. J., and Van Heuven, W. J. (2002). The architecture of the bilingual word recognition system: from identification to decision. *Biling. Lang. Cogn.* 5, 175–197. doi: 10.1017/S1366728902003012

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., et al. (2019). Multilink: a computational model for bilingual word recognition and word translation. *Biling. Lang. Cogn.* 22, 657–679. doi: 10.1017/S1366728918000287

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

Dušková, L. (1969). On sources of errors in foreign language learning. *Int. Rev. Appl. Linguist.* 7, 11–36. doi: 10.1515/iral.1969.7.1.11

Duta, M., and Plunkett, K. (2021). A neural network model of lexical-semantic competition during spoken word recognition, PsyArXiv. doi:10.31234/osf.Io/64nht

Duyck, W., Vanderelst, D., Desmet, T., and Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychon. Bull. Rev.* 15, 850–855. doi: 10.3758/PBR.15.4.850

Dylman, A. S., and Bjärtå, A. (2018). When your heart is in your mouth: the effect of second language use on negative emotions. *Cognit. Emot.* 33, 1284–1290. doi: 10.1080/02699931.2018.1540403

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Lang. Learn.* 61, 367–413. doi: 10.1111/j.1467-9922.2010.00613.x

Elgort, I., and Warren, P. (2014). L2 vocabulary learning from reading: explicit and tacit lexical knowledge and the role of learner and item variables. *Lang. Learn.* 64, 365–414. doi: 10.1111/lang.12052

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second. Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024

Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *J. Phon.* 36, 345–360. doi: 10.1016/j.wocn.2007.11.002

Escudero, P., Simon, E., and Mulak, K. (2014). Learning words in a new language: orthography doesn't always help. *Biling. Lang. Cogn.* 17, 384–395. doi: 10.1017/S1366728913000436

Escudero, P., and Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Lang. Speech* 53, 343–365. doi: 10.1177/0023830910371447

Fitzpatrick, T. (2006). Habits and rabbits: word associations and the L2 lexicon. *EUROSLA Yearb.* 6, 121–145. doi: 10.1075/eurosla.6.09fit

Fitzpatrick, T., and Izura, C. (2011). Word association in L1 and L2: an exploratory study of response types, response times, and interlingual mediation. *Stud. Second. Lang. Acquis.* 33, 373–398. doi: 10.1017/S0272263111000027

Fitzpatrick, T., and Thwaites, P. (2020). Word association research and the L2 lexicon. *Lang. Teach.* 53, 237–274. doi: 10.1017/S0261444820000105

Francis, W. S., and Strobach, E. N. (2013). The bilingual L2 advantage in recognition memory. *Psychon. Bull. Rev.* 20, 1296–1303. doi: 10.3758/s13423-013-0427-y

Garlock, V. M., Walley, A. C., and Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *J. Mem. Lang.* 45, 468–492. doi: 10.1006/jmla.2000.2784

Gaskell, M., and Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition* 89, 105–132. doi: 10.1016/S0010-0277(03)00070-2

Gass, S. M. (1997). *Input, Interaction and the Second Language Learner*. Mahwah, NJ: Lawrence Erlbaum.

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. *J. Mem. Lang.* 58, 787–814. doi: 10.1016/j.jml.2007.07.001

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., and Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Mem. Cogn.* 33, 1220–1234. doi: 10.3758/BF03193224

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., and Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. *J. Exp. Psychol. Gen.* 140, 186–209. doi: 10.1037/a0022256

Gor, K. (2014). Raspberry, not a car: context predictability and a phonological advantage in early and late learners' processing of speech in noise. *Front. Psychol.* 5:1449. doi: 10.3389/fpsyg.2014.01449

Gor, K., and Cook, S. V. (2020). A mare in a pub? Nonnative facilitation in phonological priming. *Second. Lang. Res.* 36, 123–140. doi: 10.1177/0267658318769962

Gor, K., Cook, S., and Jackson, S. (2010). Lexical access in highly proficient late L2 learners: evidence from semantic and phonological auditory priming. *Second Language Research Forum (SLRF)*, University of Maryland.

Grainger, J., and Ziegler, J. (2011). A dual-route approach to orthographic processing. *Front. Psychol.* 2:54. doi: 10.3389/fpsyg.2011.00054

Granena, G., and Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second. Lang. Res.* 29, 311–343. doi: 10.1177/0267658312461497

Hadjichristidis, C., Geipel, J., and Savadori, L. (2015). The effect of foreign language in judgments of risk and benefit: the role of affect. *J. Exp. Psychol. Appl.* 21, 117–129. doi: 10.1037/xap0000044

Håkansson, G., and Norrby, C. (2010). Environmental influence on language acquisition: comparing second and foreign language acquisition of Swedish. *Lang. Learn.* 60, 628–650. doi: 10.1111/j.1467-9922.2010.00569.x

Hall, C. J., and Ecke, P. (2003). "Parasitism as a default mechanism in L3 vocabulary acquisition," in *The Multilingual Lexicon*. eds. J. Cenoz, B. Hufeisen and U. Jessner (Dordrecht: Kluwer Academic Publishers), 71–85.

Harris, C. L., Aycicegi, A., and Gleason, J. B. (2003). Taboo words and reprimands elicit greater autonomic reactivity in a first language than in a second language. *Appl. Psycholinguist.* 24, 561–579. doi: 10.1017/S0142716403000286

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: evidence from 2/3 million English speakers. *Cognition* 177, 263–277. doi: 10.1016/j.cognition.2018.04.007

Havas, V., Taylor, J. S. H., Vaquero, L., de Diego-Balaguer, R., Rodríguez-Fornells, A., and Davis, M. H. (2018). Semantic and phonological schema influence spoken word learning and overnight consolidation. *Q. J. Exp. Psychol.* 71, 1469–1481. doi: 10.1080/17470218.2017.1329325

Hayakawa, S., Tannenbaum, D., Costa, A., Corey, J. D., and Keysar, B. (2017). Thinking more or feeling less? Explaining the foreign-language effect on moral judgment. *Psychol. Sci.* 28, 1387–1397. doi: 10.1177/0956797617720944

Hayes-Harb, R., and Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second. Lang. Res.* 24, 5–33. doi: 10.1177/0267658307082980

Hemchua, S., and Schmitt, N. (2006). An analysis of lexical errors in the English compositions of Thai learners. *Prospect* 21, 3–25.

Heyer, V., and Clahsen, H. (2015). Late bilinguals see a scan in scanner AND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Biling. Lang. Cogn.* 18, 543–550. doi: 10.1017/S1366728914000662

Ivaz, L., Costa, A., and Duñabeitia, J. A. (2016). The emotional impact of being myself: emotions and foreign-language processing. *J. Exp. Psychol. Learn. Mem. Cogn.* 42, 489–496. doi: 10.1037/xlm0000179

Jarvis, S. (2000). Methodological rigor in the study of transfer: identifying L1 influence in them interlanguage lexicon. *Lang. Learn.* 50, 245–309. doi: 10.1111/0023-8333.00118

Jiang, N. (2000). Lexical representation and development in a second language. *Appl. Linguist.* 21, 47–77. doi:10.1093/applin/21.1.47

Jiang, N., and Forster, K. I. (2001). Cross-language priming asymmetries in lexical decision and episodic recognition. *J. Mem. Lang.* 44, 32–51. doi: 10.1006/jmla.2000.2737

Jiang, N., and Zhang, J. (2021). Form prominence in the L2 lexicon: further evidence from word association. *Second. Lang. Res.* 37, 69–90. doi: 10.1177/0267658319827053

Khazaeenezhad, B., and Alibabaee, A. (2013). Investigating the role of L2 language proficiency in word association behavior of L2 learners: a case of Iranian EFL learners. *Theory Pract. Lang. Stud.* 3, 108–115. doi: 10.4304/tpls.3.1.108-115

Kim, J. (2013). Developing conceptual understanding of sarcasm in a second language through concept-based instruction. Doctoral dissertation. Pennsylvania State University.

Kuperman, V., Estes, Z., Brysbaert, M., and Warriner, A. B. (2014). Emotion and language: valence and arousal affect word recognition. *J. Exp. Psychol. Gen.* 143, 1065–1081. doi: 10.1037/a0035669

Kuperman, V., Stadthagen-Gonzalez, H., and Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behav. Res. Methods* 44, 978–990. doi: 10.3758/s13428-012-0210-4

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites. Vol. 1*. Stanford, California: Stanford University Press.

Laufer, B. (1990). 'Sequence' and 'order' in the development of L2 lexis: Some evidence from lexical confusions. *Appl. Linguist*. 11, 281–296. doi:10.1093/applin/11.3.281

Laufer, B. (1997). "The lexical plight in second language reading: words you don't know, words you think you know and words you can't guess," in *Second Language Vocabulary Acquisition*. eds. J. Coady and T. Huckin (NY: Cambridge University Press), 2–34.

Lawson, M. J., and Hogben, D. (1998). Learning and recall of foreign-language vocabulary: effects of a keyword strategy for immediate and delayed recall. *Learn. Instr.* 8, 179–194. doi: 10.1016/S0959-4752(97)00016-9

Leach, L., and Samuel, A. G. (2007). Lexical configuration and lexical engagement: when adults learn new words. *Cogn. Psychol.* 55, 306–353. doi: 10.1016/j.cogpsych.2007.01.001

Lee, Y., Jang, E., and Choi, W. (2018). L2-L1 translation priming effects in a lexical decision task: evidence from low proficient Korean-English bilinguals. *Front. Psychol.* 9:267. doi: 10.3389/fpsyg.2018.00267

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady* 10, 707–710.

Li, M. A. N., Jiang, N., and Gor, K. (2017). L1 and L2 processing of compound words: evidence from masked priming experiments in English. *Biling. Lang. Cogn.* 20, 384–402. doi: 10.1017/S1366728915000681

Li, J., Taft, M., and Xu, J. (2017). The processing of English derived words by Chinese-English bilinguals. *Lang. Learn.* 67, 858–884. doi: 10.1111/lang.12247

Lindsay, S., and Gaskell, M. G. (2010). A complementary systems account of word learning in L1 and L2. *Lang. Learn.* 60, 45–63. doi: 10.1111/j.1467-9922.2010.00600.x

Llompart, M. (2020). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Biling. Lang. Cogn.* 24, 481–496. doi: 10.1017/S1366728920000656

Llompart, M. (2021). Lexical and phonetic influences on the phonolexical encoding of difficult second-language contrasts: insights from nonword rejection. *Front. Psychol.* 12:659852. doi: 10.3389/fpsyg.2021.659852

Llompart, M., and Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Biling. Lang. Cogn.* 22, 1085–1100. doi: 10.1017/S1366728918000925

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

Luce, P. A., Pisoni, D. B., and Goldinger, S. D. (1990). "Similarity neighborhoods of spoken words," in *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives*. ed. G. T. M. Altmann (Cambridge, MA, U.S.: The MIT Press), 122–147.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9

McClelland, J. L., and Elman, J. L. (1986). The TRACE model of speech perception. *Cogn. Psychol.* 18, 1–86. doi: 10.1016/0010-0285(86)90015-0

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex: insights from the successes and failures of connectionists models of learning and memory. *Psychol. Rev.* 102, 419–457. doi: 10.1037/0033-295X.102.3.419

McKean, C., Letts, C., and Howard, D. (2013). Functional reorganization in the developing lexicon: separable and changing influences of lexical and phonological variables on children's fast-mapping. *J. Child Lang.* 40, 307–335. doi: 10.1017/S0305000911000444

Meara, P. (1978). Learners' word associations in French. *Interlang. Stud. Bull.* 3, 192–211.

Meara, P. (1983). Word associations in a foreign language. *Nottingham Linguist. Circ.* 11, 29–38.

Meara, P. (1984). "The study of lexis in interlanguage," In *Interlanguage*. eds. A. Davies, C. Criper and A. Howatt (Edinburgh: University of Edinburgh Press), 225–235.

Mora, J. C., and Darcy, I. (2013). "Inhibition and phonological processing in a second language," in *EUROSLA 2013 Conference*; August 28–31, 2013. Amsterdam, the Netherlands.

Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Process.* 50, 577–597. doi: 10.1080/0163853X.2013.846964

Namei, S. (2004). Bilingual lexical development: a Persian–Swedish word association study. *Int. J. Appl. Linguist.* 14, 363–388. doi: 10.1111/j.1473-4192.2004.00070.x

Nation, I. (2006). How large a vocabulary is needed for reading and listening? *Can. Mod. Lang. Rev.* 63, 59–82. doi: 10.3138/cmlr.63.1.59

Norman, K. A., and O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: a complementary learning-systems approach. *Psychol. Rev.* 110, 611–646. doi: 10.1037/0033-295X.110.4.611

Norrby, C., and Håkansson, G. (2007). The interaction of complexity and grammatical processability: the case of Swedish as a foreign language. doi:10.1515/IRAL.2007.002

Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition* 111, 263–269. doi: 10.1016/j.cognition.2008.12.007

Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2010). Is a FAN always FUN? Phonological and orthographic effects in bilingual visual word recognition. *Lang. Speech* 53, 383–403. doi: 10.1177/0023830910371462

Pallier, C., Colomé, A., and Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: exemplar-based versus abstract lexical entries. *Psychol. Sci.* 12, 445–449. doi: 10.1111/1467-9280.00383

Palma, P., and Titone, D. (2020). Something old, something new: a review of the literature on sleep-related lexicalization of novel words in adults. *Psychon. Bull. Rev.* 28, 96–121. doi: 10.3758/s13423-020-01809-5

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730

Perfetti, C. A., and Hart, L. (2002). "The lexical quality hypothesis," in *Precursors of Functional Literacy*. eds. L. Vehoeven, C. Elbro and P. Reitsma (Amsterdam: John Benjamins), 189–213.

Ponari, M., Rodríguez-Cuadrado, S., Vinson, D., Fox, N., Costa, A., and Vigliocco, G. (2015). Processing advantage for emotional words in bilingual speakers. *Emotion* 15, 644–652. doi: 10.1037/emo0000061

Puntoni, S., De Langhe, B., and Van Osselaer, S. M. (2009). Bilingualism and the emotional intensity of advertising language. *J. Consum. Res.* 35, 1012–1025. doi: 10.1086/595022

Qiao, X., and Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition* 158, 147–152. doi: 10.1016/j.cognition.2016.10.026

Qiao, X., Forster, K. I., and Witzel, N. (2009). Is BANARA really a word? *Cognition* 113, 254–257. doi: 10.1016/j.cognition.2009.08.006

Reyna, V. F. (2012). A new intuitionism: meaning, memory, and development in fuzzy-trace theory. *Judgm. Decis. Mak.* 7, 332–359.

Sampaio, C., and Konopka, A. E. (2013). Memory for non-native language: the role of lexical processing in the retention of surface form. *Memory* 21, 537–544. doi: 10.1080/09658211.2012.746371

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Appl. Linguist.* 11, 129–158. doi: 10.1093/applin/11.2.129

Schmidtke, J. (2014). Second language experience modulates word retrieval effort in bilinguals: evidence from pupillometry. *Front. Psychol.* 5:137. doi: 10.3389/fpsyg.2014.00137

Schoonbaert, S., Duyck, W., Brysbaert, M., and Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: making sense of the findings. *Mem. Cogn.* 37, 569–586. doi: 10.3758/MC.37.5.569

Sebastián-Gallés, N., and Díaz, B. (2012). First and second language speech perception: graded learning. *Lang. Learn.* 62, 131–147. doi: 10.1111/j.1467-9922.2012.00709.x

Sebastian-Gallés, N., Rodríguez-Fornells, A., de Diego-Balaguer, R., and Díaz, B. (2006). First-and second-language phonological representations in the mental lexicon. *J. Cogn. Neurosci.* 18, 1277–1291. doi: 10.1162/jocn.2006.18.8.1277

Slowiaczek, L. M., and Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 18, 1239–1250. doi: 10.1037/0278-7393.18.6.1239

Söderman, T. (1993). "Word associations of foreign language learners and native speakers. The phenomenon of a shift in response type and its relevance for lexical development," in *Near-Native Proficiency in English*. ed. H. Ringbom (Finland: Åbo Akademi University), 91–122.

Sulpizio, S., Toti, M., Del Maschio, N., Costa, A., Fedeli, D., Job, R., et al. (2019). Are you really cursing? Neural processing of taboo words in native and foreign language. *Brain Lang.* 194, 84–92. doi: 10.1016/j.bandl.2019.05.003

Swingley, D., and Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition* 76, 147–166. doi: 10.1016/S0010-0277(00)00081-0

Takashima, A., Bakker, I., Van Hell, J. G., Janzen, G., and McQueen, J. M. (2017). Interaction between episodic and semantic memory networks in the acquisition and consolidation of novel spoken words. *Brain Lang.* 167, 44–60. doi: 10.1016/j.bandl.2016.05.009

Tomasello, M. (2003). "The key is social cognition," in *Language in Mind: Advances in the Study of Language and Thought*. eds. D. Gentner and S. Goldin-Meadow (Cambridge/London: The MIT Press), 47–57.

Van Wijnendaele, I., and Brysbaert, M. (2002). Visual word recognition in bilinguals: phonological priming from the second to the first language. *J. Exp. Psychol. Hum. Percept. Perform.* 28, 616–627. doi: 10.1037/0096-1523.28.3.616

Webb, S. (2005). Receptive and productive vocabulary learning: the effects of reading and writing on word knowledge. *Stud. Second. Lang. Acquis.* 27, 33–52. doi: 10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Appl. Linguist.* 28, 46–65. doi: 10.1093/applin/aml048

Weber, A., and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *J. Mem. Lang.* 50, 1–25. doi: 10.1016/S0749-596X(03)00105-0

White, K. S., and Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *J. Mem. Lang.* 59, 114–132. doi: 10.1016/j.jml.2008.03.001

White, K. S., Yee, E., Blumstein, S. E., and Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *J. Mem. Lang.* 68, 362–378. doi: 10.1016/j.jml.2013.01.003

Witzel, N. O., and Forster, K. I. (2012). How L2 words are stored: the episodic L2 hypothesis. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 1608–1621. doi: 10.1037/a0028072

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon: a depth of individual word knowledge model. *Stud. Second. Lang. Acquis.* 23, 41–69. doi: 10.1017/S0272263101001024

Yates, A., Ashwin, C., and Fox, E. (2010). Does emotion processing require attention? The effects of fear conditioning and perceptual load. *Emotion* 10, 822–830. doi: 10.1037/a0020325

Yee, E., and Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 32, 1–14. doi: 10.1037/0278-7393.32.1.1

Zareva, A. (2007). Structure of the second language mental lexicon: how does it compare to native speakers' lexical organization? *Second. Lang. Res.* 23, 123–153. doi: 10.1177/0267658307076543

Zareva, A., and Wolter, B. (2012). The 'promise' of three methods of word association analysis to L2 lexical research. *Second. Lang. Res.* 28, 41–67. doi: 10.1177/0267658311423452

Zhao, X., and Li, P. (2007). "Bilingual lexical representation in a self-organizing neural network model," in *Proceedings of the Annual Meeting of the Cognitive Science Society*; August 1–4, 2007 (Vol. 29, No. 29).

Zhao, X., and Li, P. (2010). Bilingual lexical interactions in an unsupervised neural network model. *Int. J. Biling. Educ. Biling.* 13, 505–524. doi: 10.1080/13670050.2010.488284

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 20 January 2022
doi: 10.3389/fpsyg.2021.689137

L2 Learners Do Not Ignore Verb's Subcategorization Information in Real-Time Syntactic Processing

Chie Nakamura[1], Manabu Arai[2], Yuki Hirose[3] and Suzanne Flynn[4]*

[1] Global Center for Science and Engineering, Waseda University, Tokyo, Japan, [2] Faculty of Economics, Seijo University, Tokyo, Japan, [3] The Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan, [4] Department of Linguistics, Massachusetts Institute of Technology, Cambridge, MA, United States

This study addressed the question of whether L2 learners are able to utilize verb's argument structure information in online structural analysis. Previous L2 research has shown that L2 learners have difficulty in using verb's intransitive information to guide online syntactic processing. This is true even though L2 learners have grammatical knowledge that is correct and similar to that of native speakers. In the present study, we contrasted three hypotheses, the initial inaccessibility account, the intransitivity overriding account, and the fuzzy subcategorization frame account, to investigate whether L2 learner's knowledge of intransitive verbs is in fact ignored in L2 online structural analysis. The initial inaccessibility account and the fuzzy subcategorization frame account predicted that L2 learners cannot access intransitivity information in building syntactic structures in any situation. The intransitivity overriding account predicted that intransitivity information is accessed in L2 parsing, but this process is overridden by the strong transitivity preference when a verb is followed by a noun phrase. Importantly, the intransitivity overriding account specifically predicted that L2 learners would be able to use intransitive information in online syntactic processing when a noun phrase does not appear immediately following a verb. We tested the three accounts in an eye-tracking reading experiment using filler-gap dependency structures. We manipulated verb's transitivity information and lexically based plausibility information and tested English native speakers as a control L1 group ($N = 29$) and Japanese-English L2 participants ($N = 32$). The results showed that L2 learners as well as native speakers processed sentences differently depending on the subcategorization information of the verb, and adopted transitive analysis only when the verb was optionally transitive, providing support for the intransitivity overriding. The results further demonstrated that L2 learners had strong expectations for the transitive structure, which is consistent with the view proposed by the hyper-active gap-filling hypothesis. In addition, the results showed that the semantic mismatch in the incorrect transitive analysis facilitated native speaker's processing but caused difficulty for L2 learners. Together, the current study provides evidence that L2 learners use intransitive information of the verbs to guide their structural analysis when there are no overriding constraints.

Keywords: eye-tracking in reading, filler-gap dependency, verb subcategorization information, online structural analysis, second language processing

Edited by:
Andreas Opitz,
Leipzig University, Germany

Reviewed by:
Brian Dillon,
University of Massachusetts Amherst, United States
Steven G. Luke,
Brigham Young University, United States

***Correspondence:**
Chie Nakamura
cnakamura@aoni.waseda.jp

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *31 March 2021*
Accepted: *17 December 2021*
Published: *20 January 2022*

Citation:
Nakamura C, Arai M, Hirose Y and Flynn S (2022) L2 Learners Do Not Ignore Verb's Subcategorization Information in Real-Time Syntactic Processing. Front. Psychol. 12:689137. doi: 10.3389/fpsyg.2021.689137

INTRODUCTION

Individual verbs contain information about which structure they can appear in. For example, the verb *listen* possesses information that it can occur in the intransitive structure but cannot occur in the transitive structure while the verb *hear* is the other way around. This is called a verb's argument structure or subcategorization frame information. It is often assumed that language users use this information during real-time language comprehension to analyze a sentence structure. It is, however, still not clear whether this holds for second language (L2) learners. Specifically, it is still under debate whether L2 learners possess the same lexically specific knowledge as that of native speakers. This is an important question as it relates to larger questions such as whether there is a qualitative difference between L1 and L2 processing, and to what extent native speakers and L2 learners share the same processing mechanisms beyond the difference in their general language proficiency. The current study addressed these questions by testing the effect of verb subcategorization information in the process of syntactic ambiguity resolution with native English speakers and Japanese speakers learning English.

In first language (L1) processing, several studies have shown that verb's structural frequency information is used in structural building operations, providing evidence that subcategorization frames and frequency information associated with each verb are used at the very early stage in sentence processing (MacDonald et al., 1994; McRae et al., 1998 among others). For example, Trueswell et al. (1993) used sentences such as example in (1) and compared verbs that typically take a direct object (1a, *forget*) to verbs that rarely take a direct object (1b, *hope*).

(1a) The student forgot the solution was in the back of the book.
(1b) The student hoped the solution was in the back of the book.

In a self-paced reading experiment, they observed longer reading times at the region following the point of the disambiguation (i.e., *in*) in (1a) than in (1b), demonstrating that comprehenders committed more strongly to the incorrect direct object analysis in (1a) than in (1b). These results, along with those from other studies (e.g., Trueswell, 1996; Garnsey et al., 1997), suggest that native speakers use verb bias information to resolve structural ambiguities during online comprehension.

However, there are also studies that failed to observe an immediate effect of verb information in online structural analyses with native speakers. While these studies also assume a major role for verb information, they argue that verb information does not guide the initial parsing operation. For example, Pickering et al. (2000) tested sentences such as in example (2), in which the verb *realize* was biased toward the sentence complement. In eye-tracking experiments, they observed longer reading times at the post-verbal noun phrase (NP) in (2b, *her exercises*) than in (2a, *her potential*).

(2a) The young athlete realized her potential 1 day might make her a world-class sprinter.
(2b) The young athlete realized her exercises 1 day might make her a world-class sprinter.

The results showed that comprehenders initially analyzed the post-verbal NP as the verb's direct object even though the verb's structural frequency information was biased against the analysis and they experienced processing difficulty when the interpretation for the direct object analysis was semantically implausible. This suggests that comprehenders did not consider verb frequency information at the initial stage of processing and adopted a direct object analysis (see also Mitchell, 1987; Ferreira and Henderson, 1990; Kennison, 2001; Pickering and Traxler, 2003; for similar results).

These studies appear to be at odds with the other studies which observed an immediate effect of the verb's structural information in online comprehension. However, there is one possible interpretation that can reconcile the two different patterns of results; the results that failed to show an immediate effect of verb information do not necessarily mean that comprehenders ignore verb information in early processing. It is possible that although lexical information is accessed immediately upon encountering the verb, it is overridden by the preference for the most frequent direct object analysis when an NP directly follows the verb. In fact, this possibility has been supported by Arai and Keller (2012), who investigated eye movements reflecting predictive structural analysis in sentence processing. In a visual world eye-tracking paradigm, they manipulated verb types with different subcategorization frames such as in example (3) and showed that on encountering the verb (e.g., *punished/disagreed*) participants immediately looked more at an object that can serve as the verb's direct object in the visual scene (e.g., artist) in (3a) than in (3b).

(3a) Surprisingly, the nun punished the artist.
(3b) Surprisingly, the nun disagreed with the artist.

Their results provide evidence that comprehenders made different predictions based on the verb's subcategorization information. In the same study, they also tested whether frequency information that a particular verb that is used more frequently in a past participle form or in a main verb form plays a role in structural prediction. The results showed that participants predicted the correct sentence structure based on the verb's frequency information. These results support the view that the verb's lexically specific information about subcategorization information and frequency information, as well as the distribution of morphological forms are immediately accessed at the earliest stages of processing during online comprehension.

In L2 processing, the evidence for the use of the verb's structural information is relatively scarce. For example, Dussias and Cramer Scaltz (2007) examined whether Spanish-English L2 learners used verb bias information in processing sentences such as in example (4). The verb was either biased toward a direct object as in (4a), or toward a sentence complement as in (4b).

(4a) The CIA director confirmed the rumor could mean a security leak.
(4b) The ticket agent admitted the mistake when he got caught.

Their results from their self-paced reading experiments showed that L2 learners experienced processing difficulty when the verb was followed by a constituent that was inconsistent with the verb's structural bias. They also found that for the verbs whose bias differed between Spanish and English, L2 learners processed them based on their L1 verb bias. (For similar results with Chinese-English L2 learners, see Juffs and Harrington, 1995; Juffs and Harrington, 1996, with French-English L2 learners, see Frenck-Mestre and Pynte, 1997). These results suggest that L2 learners can access verbal information but it remains unclear whether the information L2 learners accessed in processing L2 sentences was the lexical information of the L1 or that of the L2 (but see Lee et al., 2013 for the finding of an effect of the verb's structural frequency information with advanced Korean-English L2 learners).

Some studies are clearly inconsistent with the view that L2 learners access verb subcategorization information in sentence processing. Nakamura et al. (2013), for example, tested Japanese-English L2 participants in processing temporarily ambiguous sentences such as example (5), in which the verb was either optionally transitive (5a, *watch*) or obligatory intransitive (5b, *cry*).

(5a) When the audience watched the actor rested behind the curtain.
(5b) When the audience cried the actor rested behind the curtain.

Their results from self-paced reading studies showed that the L2 learners initially analyzed a post-verbal NP as the verb's direct object both in (5a) and (5b), demonstrating that L2 learners ignored the verb's subcategorization information, viz. the information that the verb *cry* cannot take a direct object. L2 learners always adopted a direct object analysis initially regardless of whether the analysis was licensed by the verb's subcategorization information or not.

Using the same early/late closure ambiguity, Nakamura et al. (2019) observed that L2 learners adopted the direct object analysis both with (5a) and (5b), replicating their earlier study. Furthermore, they found that patterns of a priming effect were different for (5a) and (5b). After reading (5a), the processing cost in reanalysis was reduced in reading the subsequent target sentence that had the same verb. No such learning effect was observed with (5b), in which the verb was obligatory intransitive. Their findings suggest that the reading patterns in the prime sentences were ostensibly the same between (5a) and (5b) but the structure the L2 learners activated in reading these sentences was different depending on the verb's subcategorization information, which influenced the processing of the subsequent target sentences. Importantly, they also confirmed in an off-line task that their L2 learners possessed the correct knowledge regarding subcategorization frames for the verbs used in their study.

In summary, although the results of previous studies that tested L2 learners' use of verb information might be attributed to various factors such as similarities between the learners' L1 and L2, learner's proficiency level, and learner's cognitive capacity limitations (e.g., Dekydtspotter et al., 2006; Hopp, 2010), past L2 studies largely agree that L2 learners cannot use verbal information as reliably as native speakers do and this holds true even with L2 learners at an advanced level of competence (e.g., Hoover and Dwivedi, 1998; Jiang, 2004, 2007) and regardless of the similarities between the learner's L1 and L2 (e.g., Papadopoulou and Clahsen, 2003; Marinis et al., 2005; Roberts and Felser, 2011).

The verbal structural frequency information, such that *accused* is frequently used as a participle form but *searched* is hardly ever used as a participle form, is based on statistical distributions that are learned through linguistic input (Francis and Kucera, 1982). It is, therefore, reasonable to think that the verb's structural frequency information in an L2 is difficult to master perfectly because the majority of L2 learners are exposed to far less linguistic input in the target language compared to native speakers. This might account for the results of some of the research that found similar effects in the use of verb frequency information between native English speakers and advanced L2 learners who were living in an English-speaking country at the time of testing (Lee et al., 2013). However, it does not explain why L2 learners also show difficulty in using syntactic restriction information about which structure a particular verb can appear in, viz. the information that intransitive verbs cannot take a direct object (*intransitivity information* henceforth), in online processing even though they have the correct subcategorization knowledge in the L2. To be more specific, the studies by Nakamura et al. (2013, 2019) suggest that L2 learners do possess the correct subcategorization frame information for obligatory intransitive verbs but it seems problematic for them to apply the knowledge in online structural analyses; they adopted the transitive analysis with the obligatory intransitive verbs. One possible explanation for this outcome comes from previous L1 research in English which suggests that intransitivity information is distributional information that is associated with a specific verb, and this information can be learned only through linguistic experience. In Van Gompel et al. (2012), they demonstrated that the information about whether a specific verb should be used in an intransitive structure or not is represented at a lexically specific level, whereas transitivity information is the default, category-general information (see also Van Gompel et al., 2006). This predicts that L2 learners of English are strongly influenced by the general transitive bias of the verbs and the impact of verb-specific intransitive bias remains small due to the overall shortfall of linguistic input.

One piece of evidence for L2 learners' strong preference for the transitive structure comes from work by Roberts and Felser (2011). They tested sentences such as example (6) with advanced Greek-English L2 learners.

(6a) While the band played the song pleased all the customers.
(6b) While the band played the beer pleased all the customers.

The results showed that L2 learners analyzed the post-verbal NP as the verb's direct object and experienced large processing difficulty when the direct object interpretation was semantically implausible in (6b, *played the beer*), unlike the control native speaker group who were able to quickly revise the sentence structure for the correct analysis in which the NP (the beer) is a main clause subject, when the initially adopted direct object interpretation was semantically implausible. This processing pattern most likely reflects the fact that L2 learners relied more on semantic information than on the verb's subcategorization information. Since L2 learners' knowledge about intransitive use of optionally transitive verbs such as *play* was weak or unreliable, they were not able to abandon the semantically implausible direct object analysis. This caused L2 learners' reanalysis process to be delayed or blocked by semantic information (see also Juffs and Harrington, 1996; Juffs, 1998; Dekydtspotter and Seo, 2017; for studies that tested the effect of intransitivity information in L2 processing). Their results suggest that L2 learners' syntactic processing ability, especially the ability to use information that a specific verb can appear in an intransitive structure, is reduced compared to native speakers.

The results of Roberts and Felser (2011), along with other studies that failed to observe a reliable effect of verb subcategorization information in L2 processing, support the view that L2 learners have a strong tendency to analyze an NP that immediately follows a verb as the verb's direct object. As a consequence, L2 learners tend to create a VP even when the verb's subcategorization information does not allow the analysis (Juffs and Harrington, 1996; Juffs, 1998; Nakamura et al., 2013, 2019). Again, it is important to note that L2 learner's grammaticality judgments were similar to the native speaker's judgments, suggesting that L2 learners had the correct knowledge about the subcategorization bias of the verbs in an off-line task (Juffs, 1998; Nakamura et al., 2019). If L2 learners have similar subcategorization knowledge to that of native speakers, why is this knowledge not reflected in their online processing?

One possible and probably the most straightforward interpretation would be that L2 learners possess the knowledge that particular verbs cannot take a direct object, but L2 learners have difficulty with the immediate use of this information in online processing. Findings from some previous studies suggest that even native speakers initially adopt the transitive analysis on encountering an intransitive verb and experience processing difficulty (cf. Mitchell, 1987). It is possible that the preference for the transitive analysis is even stronger for L2 learners and consequently, with an intransitive verb, L2 learners always initially attempt to adopt the transitive analysis before they can use intransitivity information and experience processing cost. We call this the *initial inaccessibility account*.

The second possibility is that intransitivity information is accessed in L2 learner's online structural analysis, but intransitivity information is overridden by a strong transitivity preference when L2 learners see an NP following the verb. More specifically, L2 learners can access subcategorization knowledge for intransitive verbs, but the presence of an NP directly following a verb overrides the intransitivity information so that L2 learners adopt the transitive analysis over the intransitive analysis. We call this the *intransitivity overriding account*. In this case, L2 learners are predicted to use intransitivity information in an online structural analysis where an NP does not appear directly following a verb.

The third possibility is that L2 learners' lexical representation of argument structure information is fuzzy in the sense that L2 information about a certain subcategorization frame is not stored rigidly as either possible or not. Instead, the lexical representation of L2 learners may allow some ambiguity in their structural specifications. As a result, L2 learners may tolerate an incorrect subcategorization frame (e.g., an obligatory intransitive verb to take a direct object) to a greater extent compared to native speakers. Due to the fuzzy structural representations, L2 learners permit subcategorization violation and make semantic interpretations out of the information they receive. We call this the *fuzzy subcategorization frame account*. Under this account it is predicted that L2 learners cannot use verb's subcategorization information to sort out which structures are possible or not due to the L2 learner's fuzzy structural representations. Instead, they would rely on the semantic relationship between the verb and an NP. This account is consistent with the view suggested in some previous L2 studies that L2 processing is strongly influenced by lexical-semantic cues but less so by syntactic information (cf. Clahsen and Felser, 2006).

In order to test these accounts, we examined L2 learner's processing using the unbounded dependency structure such as (7a). In processing this structure, a parser needs to associate the object NP (*the celebrity*), which is referred to as the *filler*, to the correct post-verbal thematic position, called the *gap*. Since the verb *interview* in (7a) is optionally transitive, readers would typically posit an incorrect gap immediately following the verb (i.e., *That's the celebrity that the writers interviewed___* [*about.*]), by analyzing the NP as a direct object of the verb (i.e., *The writer interviewed the celebrity*). However, this interpretation turns out to be inconsistent with the sentence continuation at the information *at the conference*, and readers are thus forced to reanalyze the structure for the correct intransitive interpretation (i.e., the writer did not interview the celebrity, but she/he interviewed about the celebrity.).

(7a) That's the celebrity that the writer interviewed about at the conference.
(7b) That's the letter that the writer interviewed about at the conference.
(7c) That's the letter that the writer smiled about at the conference.

Using this structure, we investigated the influence of verb-specific information on L2 learner's initial parsing processes by manipulating transitivity information of the verb in (7). The verb was either optionally transitive (7a, 7b; *interview*) or intransitive (7c; *smile*). In addition, we also manipulated semantic information for the incorrect direct object analysis in two optionally transitive verb conditions (7a, b). In L1 studies, it has been shown that semantically anomalous interpretation helps L1 speakers to quickly abandon the favored analysis and adopt the correct analysis before disambiguation (Pickering and Traxler, 1998). In contrast, there is evidence that L2 learners cannot move beyond the favored analysis even when the interpretation for the analysis is semantically anomalous (Roberts and Felser, 2011). In order to examine whether semantic information helps L2 learners to recover from the incorrect analysis in processing filler-gap dependency structures, we used different nouns for the filler NP. The incorrect gap-filling direct object analysis resulted in either semantically plausible (7a; *interviewed the celebrity*, Plausible transitive condition), or impossible (7b; *interviewed the letter,* Implausible transitive condition).

If the verb's subcategorization information does not have an influence on L2 processing, the preceding NP (*the celebrity/the letter*) would be always analyzed as the verb's direct object in all conditions. Importantly, the initial inaccessibility account and the fuzzy subcategorization frame account both assume that subcategorization information cannot be used during L2 structural analysis, but these two accounts predict different reading patterns. Under the initial inaccessibility account, L2 learners are predicted to initially attempt to analyze the preceding NP (*the celebrity/the letter*) as the verb's direct object regardless of the verb type. As a result, L2 learners would experience processing difficulty after they encounter the verb both in (7b) and (7c) compared to (7a). In (7b), processing cost would be observed due to the semantically implausible interpretation (*interviewed the letter*). In (7c), L2 learners initially adopt a direct object analysis and experience processing difficulty because the analysis violates the verb's subcategorization information (*smiled the letter*). Under the fuzzy subcategorization frame account, it is predicted that L2 learners are incapable of making structural judgment based on verb's intransitivity information, thus they cannot reject the sentence when an intransitive verb is used in a transitive structure. This suggests that L2 learners form an interpretation of the ungrammatical sentence using semantic information, such that they would interpret "smiled the letter" in (7c) meaning something like "smiled about the letter" or "smiled at the letter". If this were the case, L2 learners would show processing difficulty only in (7b) because the direct object analysis with the optionally transitive verb *interview* generates semantically anomalous interpretation. There would be no processing difficulty in (7c) because L2 learners would tolerate the incorrect transitive use with the intransitive verb and build semantically plausible interpretation (e.g., *the writer smiled at/about the letter*).

In contrast to the two accounts, the intransitivity overriding account predicts that subcategorization information is accessed and used in L2 processing as long as an NP does not appear immediately following a verb. Under this account, the preceding NP would be analyzed as the verb's potential direct object only when the verb is optionally transitive in (7a, 7b), but not when the verb is intransitive in (7c). Thus, this account predicts that processing difficulty at the verb would occur only in (7b) due to the semantically implausible direct object analysis but not in (7a) and (7c).

Using these sentences, we conducted an eye-tracking experiment with native speaker of English and Japanese L2 learners of English. In what follows, we will first describe how the experiments were conducted in "Materials and Methods" section. We will then explain how the analyses were conducted and report the results.

MATERIALS AND METHODS

Participants

Twenty-nine native speakers of English (L1 group) and 31 Japanese learners of English (L2 group) participated in the study. Participants of the L1 group were recruited in the Boston area and received a small remuneration for their voluntary participation. All of the participants had normal or corrected-to-normal vision. Participants of the L2 group were adult Japanese-L1 English-L2 speakers living in Japan. They were all undergraduate students at the University of Tokyo, who had at least 6 years of English education in junior high and high school before enrolling in the university. We obtained L2 participants' scores for the standardized English test in the National Center Test for University Admissions (*mean score* = 194.8 out of 200, *SD* = 7.30). Our L2 participants' scores corresponded to the proficiency level of B2 to C1 (Independent user level to Proficient user level) on the Common European Framework of Reference for Languages (CEFR) (Council of Europe [COE], 2021).

Materials and Design

We created 24 sets of experimental items in three conditions (Plausible transitive, Impossible transitive, and Intransitive) as shown again below in example (7). Regions were divided as indicated by the region numbers. These regions were used for the purpose of analysis and were not presented in the experiment. The complete set of target items used in the experiment is shown in **Supplementary Table 1**.

(7a) Plausible transitive condition

1	2	3	4	5	6	7
That's	the celebrity	that	the writer	interviewed	about	at the conference.

(7b) Impossible transitive condition

| That's | the letter| that | the writer | interviewed | about | at the conference. |

(7c) Intransitive condition

| That's | the letter | that | the writer | smiled | about | at the conference. |

Procedure

Three lists of items were created following a Latin square design. Each list included 48 fillers and was presented in a pseudo-random order. The filler sentences were structurally unrelated copular sentences. The eye-movements during reading were recorded using EyeLink 1000 (SR Research) for the L1 group, and Eye-Link II (SR Research) for the L2 group. In both experimental settings, a 21" LCD monitoring screen was placed approximately 55 cm away from participants and participants' eye-movements were recorded at the sampling rate of 500 Hz. A brief calibration set-up was conducted at the beginning of each experimental session. Before each trial, participants saw a square box in the position of the first character of a sentence, which triggered the presentation of sentences. They pressed the space bar when they had finished reading the whole sentence. Eighteen comprehension questions were included following filler sentences to keep participants focused. None of the questions concerned the understanding of the filler-gap dependency structure. The experiment session always started with four practice sentences along with two comprehension questions.

DATA ANALYSIS AND RESULTS

Methods of Analysis and Results

Prior to the analysis of eye-tracking data, we checked the participants' response accuracy rate for the comprehension questions. The average correct response rate was 96.8% (*SD* = 17.4) for the L1 group, and 92.6% (*SD* = 2.7) for the L2 group. None of the participants were excluded from the analysis. The eye-tracking data were analyzed in three eye-movement measures; *first-pass*, *right-bound*, and *second-pass* reading times. First pass time is the sum of durations of the fixations in a particular region following the first entry in the region until the first fixation outside the region (either to the left or the right). Right-bounded reading time is the sum of fixation durations in a particular region before the first fixation exiting the region to the right. Second-pass reading time is the sum of fixations made in a region after the region has already been exited to the right. These measures were selected to analyze the initial and second stages of processing. First pass and right-bound times in a given region reflect reading patterns before seeing any information following the region of interest, thus they are considered early measures in the sense that the behavior does not reflect the uptake of information in the following regions. Second pass times reflect reading patterns following the encounter of the region for the second time or later (i.e., re-reading after readers proceeded to the following regions), thus considered as a measure that reflects a later stage of processing such as structural reanalysis. The mean reading times for the three eye-movement measures from Region 2 to Region 7 in each condition for the two groups are shown in **Supplementary Table 2**.

For statistical analysis, we analyzed the reading times in the three measures in each region using Linear Mixed-Effects (LME) models (Baayen, 2008). In the model, Condition (*Plausible transitive, Impossible transitive, or Intransitive*), Group (L1 or L2), and interaction between the two were included as fixed effects. Participants and items were included as random effects. In Region 2 and Region 5 where different words were used across conditions, the number of characters (Word Length) and word frequency (Frequency) were included in the model as additional control factors. The frequency for the lexical items used in Region 2 and Region 5 was counted using the written part of the British National Corpus (data obtained from http://english-corpora.org/bnc). For the lexical items used in the verb region (Region 5), we only counted instances in which the word was used as a verb. The mean frequency counts for the words used in Region 2 in the Plausible transitive condition and those used in the Impossible transitive/Intransitive conditions were 6,313 (*SD* = 7,034) and 13,748 (*SD* = 12,958) respectively. The mean frequency counts for the verbs used in Region 5 in the Plausible transitive/Impossible transitive conditions and those used in the Intransitive condition were 6,587 (*SD* = 7,026) and 2,937 (*SD* = 2,473) respectively.[1]

We analyzed the three conditions using LME models with dummy coding by treating the Impossible transitive condition as a baseline against which the effects of the other two conditions were tested. The Impossible transitive condition was used as a baseline so that we can examine the effect of semantic plausibility by a comparison between the Impossible transitive condition and the Plausible condition [e.g., *interviewed the letter* vs. *interviewed the celebrity* in example (7)], as well as the effect of verb's subcategorization information by the comparison between the Impossible transitive condition and the Intransitive condition [e.g., *interviewed the letter* vs. *complained the letter* in example (7)]. In the report, a main effect of Plausible transitive condition reflects an effect of semantically plausible/impossible transitive analysis, and a main effect of Intransitive condition reflects an effect of verb's subcategorization information. Importantly, an interaction between Group and the two experimental conditions reflects the difference in the use of semantic plausibility information and that of verb's subcategorization information. The factor Group was also dummy coded in the model in which the L1 group was treated as a baseline. For results that showed an interaction between Group and experimental conditions, we conducted a simple effect analysis using the same model by dummy coding the L2 group as a baseline to explore the significance of the main effect with the L2 group. The initial model included a random slope of the fixed effect for both participant and item random effects. The best-fitting model was explored using a backward selection approach. We excluded data that exceeded two standard deviations above the absolute value of residuals from the best-fitting model (Baayen, 2008).

Table 1 shows the results of the analysis in each region in each measure. *P*-values were obtained using the R package lmerTest, which estimate the degree of freedom *via* the Satterthwaite approximation. Below, we discuss the results in each region.

Region 2 (The Celebrity/The Letter)

Right-bounded times in this region showed an interaction between Group and Intransitive condition. The interaction indicates that there was a difference between the two groups in the way that the sentences with the intransitive verb were processed compared to the processing of the baseline Impossible transitive condition. The simple effect analysis showed that an effect of Intransitive condition was observed only with the L1 group but not with the L2 group (p = 0.523). With the L1 group, the reading time was shorter in the Intransitive condition than in the Impossible transitive condition (Intransitive: 394 ms, Impossible transitive: 474 ms). Since the lexical information is consistent

[1]For the verbs used in Region 5, we conducted another analysis to see whether L2 learners' familiarity with the verbs was different between the two types of the verbs (transitive and intransitive). For this analysis, we used a database in which familiarity for 3,000 English words is rated by Japanese learners of English (Yokokawa, 2006). The average familiarity rating for the verbs used in the transitive conditions was 5.67 in the 7 point scale (*SD* = 0.66), and that for the verbs used in the intransitive condition was 5.87 (*SD* = 0.67). The result of an unpaired t-test showed that there was no difference in L2 learners' familiarity with the verbs used in the transitive and intransitive conditions (p = 0.161).

TABLE 1 | Results of linear mixed-effects models of the three reading time measures in each region.

	β	*SE*	*t*	*p*
Region 2 (*the celebrity/the letter*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	212.88	31.02		
Plausible transitive condition	–20.71	35.01	–0.59	0.554
Intransitive condition	–3.35	33.09	–0.10	0.920
Group	**340.35**	**36.27**	**9.38**	**<0.001**
Group × Plausible transitive condition	15.96	39.85	0.40	0.689
Group × Intransitive condition	5.48	38.56	0.14	0.887
• *Right–bounded reading time*				
Intercept (Baseline: L1 group)	482.49	40.75		
Plausible transitive condition	–44.11	42.38	–1.04	0.300
Intransitive condition	**–103.59**	**38.16**	**–2.72**	**0.007**
Group	**131.78**	**50.51**	**2.61**	**0.011**
Group × Plausible transitive condition	36.87	49.18	0.75	0.455
Group × Intransitive condition	**117.91**	**44.27**	**2.66**	**0.008**
• *Second pass reading time*				
Intercept (Baseline: L1group)	586.12	83.95		
Plausible transitive condition	–98.86	79.27	–1.25	0.213
Intransitive condition	–113.67	74.56	–1.53	0.128
Group	153.65	106.79	1.44	0.153
Group × Plausible transitive condition	22.27	90.18	0.25	0.805
Group × Intransitive condition	26.10	86.10	0.30	0.762
Region 3 (*that*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	154.20	12.99		
Plausible transitive condition	8.60	16.08	0.54	0.593
Intransitive condition	4.99	16.94	0.30	0.768
Group	**103.01**	**15.68**	**6.57**	**<0.001**
Group × Plausible transitive condition	0.60	18.13	0.03	0.973
Group × Intransitive condition	–2.40	18.97	–0.13	0.899
• *Right–bounded reading time*				
Intercept (Baseline: L1group)	191.14	27.14		
Plausible transitive condition	20.31	20.92	0.97	0.332
Intransitive condition	15.16	22.41	0.68	0.499
Group	**102.50**	**20.27**	**5.06**	**<0.001**
Group × Plausible transitive condition	–23.37	23.63	–0.99	0.323
Group × Intransitive condition	–19.74	25.05	–0.79	0.431
• *Second pass reading time*				
Intercept (Baseline: L1group)	198.80	34.68		
Plausible transitive condition	–47.11	41.23	–1.14	0.254
Intransitive condition	21.66	42.91	0.51	0.614
Group	34.98	42.07	0.83	0.407
Group × Plausible transitive condition	48.37	45.94	1.05	0.293
Group × Intransitive condition	–40.84	47.46	–0.86	0.390
Region 4 (*the writer*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	221.42	25.92		
Plausible transitive condition	–2.47	19.93	–0.12	0.901
Intransitive condition	2.60	19.78	0.13	0.900
Group	**307.34**	**29.75**	**10.33**	**<0.001**
Group × Plausible transitive condition	–10.71	26.51	–0.40	0.686
Group × Intransitive condition	–2.14	26.52	–0.08	0.936
• *Right–bounded reading time*				
Intercept (Baseline: L1group)	372.98	36.01		
Plausible transitive condition	3.86	22.60	0.17	0.864
Intransitive condition	27.47	22.52	1.22	0.223
Group	**251.11**	**40.15**	**6.26**	**<0.001**
Group × Plausible transitive condition	–18.00	30.00	–0.60	0.548
Group × Intransitive condition	–39.11	30.01	–1.30	0.193
• *Second pass reading time*				
Intercept (Baseline: L1group)	420.38	58.84		
Plausible transitive condition	–0.89	37.28	–0.02	0.981
Intransitive condition	–46.78	36.93	–1.27	0.206

(Continued)

TABLE 1 | (Continued)

	β	*SE*	*t*	*p*
Group	107.93	77.97	1.38	0.170
Group × Plausible transitive condition	15.01	49.67	0.30	0.762
Group × Intransitive condition	41.15	29.23	0.84	0.403
Region 5 (*interviewed/smiled*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	216.11	16.38		
Plausible transitive condition	–2.96	12.56	–0.24	0.814
Intransitive condition	–9.74	12.48	–0.78	0.436
Group	**195.71**	**22.56**	**8.68**	**<0.001**
Group × Plausible transitive condition	1.72	17.04	0.10	0.920
Group × Intransitive condition	**42.48**	**17.07**	**2.49**	**0.013**
• *Right–bounded reading time*				
Intercept (Baseline: L1group)	269.69	18.56		
Plausible transitive condition	–10.81	13.98	–0.77	0.440
Intransitive condition	–20.27	13.93	–1.46	0.146
Group	**186.91**	**25.05**	**7.46**	**<0.001**
Group × Plausible transitive condition	–8.86	18.98	–0.47	0.641
Group × Intransitive condition	**43.52**	**19.03**	**2.29**	**0.022**
• *Second pass reading time*				
Intercept (Baseline: L1group)	236.31	48.20		
Plausible transitive condition	40.03	34.00	1.18	0.239
Intransitive condition	–4.87	34.30	–0.14	0.887
Group	**268.71**	**63.76**	**4.21**	**<0.001**
Group × Plausible transitive condition	–69.26	45.55	–1.52	0.129
Group × Intransitive condition	–51.03	45.95	–1.11	0.267
Region 6 (*about*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	207.56	10.84		
Plausible transitive condition	4.98	10.20	0.49	0.625
Intransitive condition	–8.02	11.19	–0.72	0.474
Group	**67.25**	**14.21**	**4.73**	**<0.001**
Group × Plausible transitive condition	**–22.86**	**13.00**	**–1.76**	**0.079**
Group × Intransitive condition	–2.79	14.00	–0.20	0.842
• *Right–bounded reading time*				
Intercept (Baseline: L1group)	242.84	12.08		
Plausible transitive condition	–10.06	11.85	–0.85	0.400
Intransitive condition	**–41.16**	**13.11**	**–3.14**	**0.002**
Group	**56.36**	**15.88**	**3.55**	**<0.001**
Group × Plausible transitive condition	–7.35	15.13	–0.49	0.627
Group × Intransitive condition	23.55	16.42	1.43	0.152
• *Second pass reading time*				
Intercept (Baseline: L1group)	220.74	37.21		
Plausible transitive condition	21.16	34.79	0.61	0.545
Intransitive condition	**–75.11**	**35.92**	**–2.09**	**0.038**
Group	**118.06**	**48.67**	**2.43**	**0.018**
Group × Plausible transitive condition	–30.22	44.63	–0.68	0.500
Group × Intransitive condition	–36.61	44.77	–0.82	0.415
Region 7 (*at the conference*)				
• *First pass reading time*				
Intercept (Baseline: L1group)	399.85	38.69		
Plausible transitive condition	–11.89	25.37	–0.47	0.640
Intransitive condition	35.10	25.28	1.39	0.165
Group	**324.39**	**50.01**	**6.49**	**<0.001**
Group × Plausible transitive condition	–41.18	35.85	–1.15	0.251
Group × Intransitive condition	–12.15	35.63	–0.34	0.733
• *Right–bounded reading time*				
Intercept (Baseline: L1group)	822.01	95.28		
Plausible transitive condition	**104.85**	**42.79**	**2.45**	**0.014**
Intransitive condition	39.55	42.39	0.93	0.351
Group	**538.48**	**69.30**	**4.47**	**<0.001**
Group × Plausible transitive condition	**–152.13**	**60.52**	**–2.51**	**0.012**
Group × Intransitive condition	–34.69	59.95	–0.58	0.563

The results with p-values less than 0.05 are shown in bold.

between the Impossible transitive condition and the Intransitive condition up to Region 4 (e.g., ... *the letter that the writer*...), the effect of Intransitive condition in this region with the L1 group is most likely a parafoveal effect. When the upcoming verb was intransitive, the L1 group processed the sentence faster and continued on to the following region quickly compared to when the upcoming verb was optionally transitive. No such parafoveal effect of intransitive information was found with the L2 group.

First pass and right-bounded times in this region also showed a main effect of Group, showing that the reading times of this region in these measures were longer with the L2 group than the L1 group.

Region 3 and 4 (That the Writer)

A main effect of Group was found in first pass times and right-bounded times in Region 3 and 4. These results indicate that the L2 group was overall slower to process information in these regions compared to the L1 group.

Region 5 (Interviewed/Smiled)

First pass and right-bounded times in this region showed an interaction between Group and Intransitive condition. The interaction indicates that there was a difference between the two groups in the way that the sentences with the intransitive verb were processed compared to the processing of the baseline Impossible transitive condition. The simple effect analysis indicated an effect of Intransitive condition with the L2 group (β = 32.75, *SE* = 11.74, *t* = 2.79, *p* = 0.005) but there was no effect of Intransitive condition with the L1 group. This demonstrates that the reading time for the intransitive verbs was significantly longer than that for the optionally transitive verbs in the Impossible transitive condition only with the L2 group (Intransitive: 500 ms, Impossible transitive: 421 ms for L2, Intransitive: 212 ms, Impossible transitive: 232 ms for L1). Similarly, the interaction between Group and Intransitive condition in right-bounded reading time also showed a marginal simple effect of Intransitive condition with the L2 group (β = 23.25, *SE* = 13.15, *t* = 1.77, *p* = 0.077) but there was no effect of Intransitive condition with the L1 group. This demonstrates that the L2 group spent longer time to process the verb before they proceeded to the following regions in the Intransitive verb condition than in the Impossible transitive verb condition, but this effect was not observed with the L1 group (Intransitive: 550 ms, Impossible transitive: 461 ms for L2, Intransitive: 260 ms, Impossible transitive: 284 ms for L1). **Figure 1** illustrates the different reading patterns in first pass times between the two groups. As shown in the figure, the L2 group showed increased reading time for the Intransitive condition compared to other two conditions.

The longer reading time for the intransitive verbs in first pass and right-bounded times with the L2 group indicates that L2 learners experienced processing difficulty on encountering an intransitive verb compared to when they saw a transitive verb. The L2 group's results that they required longer reading time to process intransitive verbs most likely reflect that L2 learners had a strong expectation for an upcoming verb to be a transitive verb that can take the preceded NP (*the celebrity*/*the letter*) as a direct object. As a result, they experienced processing difficulty on encountering an intransitive verb that did not match with the expectations they had made. Importantly, the different processing patterns between the Intransitive condition and the Impossible transitive condition provide evidence for the use of verb's subcategorization information in L2 processing. If the increased reading times in the Intransitive condition were due to the semantically anomalous interpretation of the direct object analysis (e.g., *smiled the letter*), then the reading times in the Impossible transitive condition would also be similar to that in the Intransitive condition because the direct object analysis in this condition also causes semantically anomalous interpretation (e.g., *interviewed the letter*). Thus, the difference between the two conditions indicate that the L2 group distinguished obligatory intransitive verbs from optionally transitive verbs in online structure analysis. As was confirmed by the results of the simple effect of Intransitive condition, processing difficulty on encountering the intransitive verbs was observed only with the L2 group but not with the L1 group.

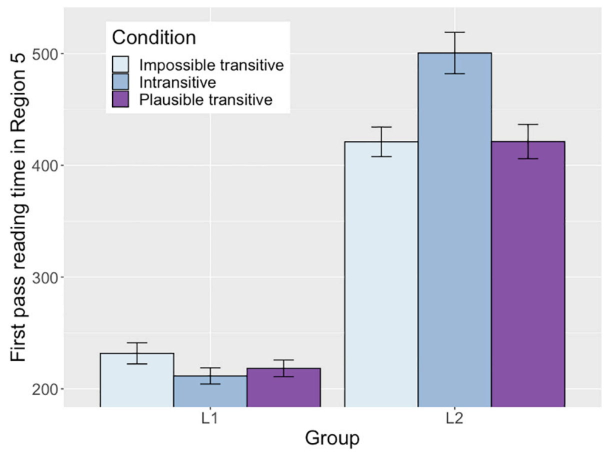


FIGURE 1 | First pass reading time in Region 5 in each condition in each group.

In addition, there was a main effect of Group in all measures, showing that the overall reading times with L2 group were longer than the L1 group in this region.

Region 6 (About)

Right-bounded times in this region showed a main effect of Intransitive condition, demonstrating that the time spent at this region, including re-reading of the earlier regions before participants proceeded to the following region was shorter in the Intransitive condition than in the Impossible transitive condition regardless of the group (Intransitive: 538 ms, Impossible transitive: 578 ms). The same effect of Intransitive condition was also observed in second pass times, showing that re-reading time in this region was shorter in the Intransitive condition compared to the Impossible transitive condition (Intransitive: 525 ms, Impossible transitive: 494 ms). These results most likely reflect a smaller cost for structural revision in the Intransitive condition as the verb's intransitivity information was helpful for both groups in reaching the correct analysis without structural reanalysis.

In addition, first pass reading time in this region showed a marginal interaction between Group and Plausible transitive condition. Although the interaction did not reach the level of significance, the main effect of Group in the first pass times shows that the L2 group's reading time was significantly longer compared to the L1 group's. This suggests the possibility that the marginally significant interaction between Group and Plausible transitive condition might indicate different reading patterns between the two groups. We explored this possibility by conducting a simple effect analysis with the L2 group. The results showed there was a marginal effect of Plausible transitive condition with the L2 group ($\beta = -17.41$, $SE = 9.40$, $t = -1.85$, $p = 0.064$), while there was no effect of Plausible transitive condition with the L1 group. The results demonstrates that the L2 group tended to spend longer time reading the post-verbal region in the Impossible transitive condition compared to the Plausible transitive condition, and this tendency was observed only with the L2 group (Impossible transitive: 292 ms, Plausible transitive: 275 ms for L2, Impossible transitive: 233 ms, Plausible transitive: 233 ms for L1). This most likely reflects that the L2 group showed a spill-over effect in this region when the verb information in the previous region resulted in semantically anomalous direct object interpretation in the Impossible transitive condition (e.g., *interviewed the letter*). The fact that the interaction failed to reach full significance might be, at least partly, due to low statistical power in our analysis. Our study could be considered as two sets of separate studies (L1 and L2 groups), with each having 29 and 31 participants. Although the number of participants is not small relative to other L2 studies, it may be possible that the sample size was not large enough for reliably observing an interaction in the unified analysis with a within-participants sentence condition (Impossible transitive vs. Plausible transitive) and a between-participants group condition (L1 and L2). At the same time, low statistical power could also increase the risk of Type I error so that we should be careful not to overinterpret this finding.

As in the other regions, a main effect of Group was observed in all reading measures in this region, again showing that the overall reading times with L2 group were longer than the L1 group.

Region 7 (At the Conference)

Right-bounded reading time in this region showed an effect of Plausible transitive condition as well as an interaction between Group and Plausible transitive condition. From the main effect of Plausible transitive condition, it was demonstrated that the L1 group's reading times in this region were longer in the Plausible transitive condition than in the Impossible transitive condition. The simple effect analysis with the L2 group showed this effect was not significant with the L2 group ($p = 0.269$). The longer right-bounded times in this region with the L1 group most likely reflect that the L1 group had adopted the direct object analysis in the Plausible transitive condition up to this region, and regressed to the earlier regions on encountering information that was not consistent with the analysis. This is because the direct object analysis in the Plausible transitive condition is possible up to the previous post-verbal region (e.g., *The writer interviewed the celebrity about*. . .[*her next concert*]), but this analysis becomes impossible in the current sentence-final region (e.g., **The writer interviewed the celebrity about at the conference*). **Figure 2** illustrates the different reading patterns in right-bounded times between the two groups. As shown in the figure, only the L1 group showed increased reading time for the Plausible transitive condition compared to the Impossible transitive condition.

The first pass and right-bounded times in this region showed a main effect of Group, reflecting that these reading times were longer with the L2 group compared to the L1 group.

Summary of the Results

The results revealed both similarities and differences between the L1 group and the L2 group in the use of subcategorization information during the processing of the filler-gap dependency structure.

First, the L2 group was surprised to encounter the intransitive verbs. The first pass and the right-bounded reading times at the verb region (Region 5) showed that only the L2 group showed longer reading times for the intransitive verbs compared to the transitive verbs. This demonstrates that the L2 group required extra time to process an intransitive verb, most likely reflecting their strong expectations for the upcoming verb to be a transitive verb that takes a preceded NP as the verb's direct object. Importantly, the results cannot be explained by the possibility that the L2 group adopted an direct object in the Intransitive condition by ignoring the verb's subcategorization information and experienced processing difficulty due to the semantically anomalous interpretation. If this were the case, the Impossible transitive condition should also show increased reading times for the same reason at the verb region. Thus, the difference between the Intransitive condition and the Impossible transitive condition provides evidence for the use of verb's subcategorization information in L2 processing. With the L1 group, a very early effect of Intransitive condition was observed as a parafoveal effect in Region 2.

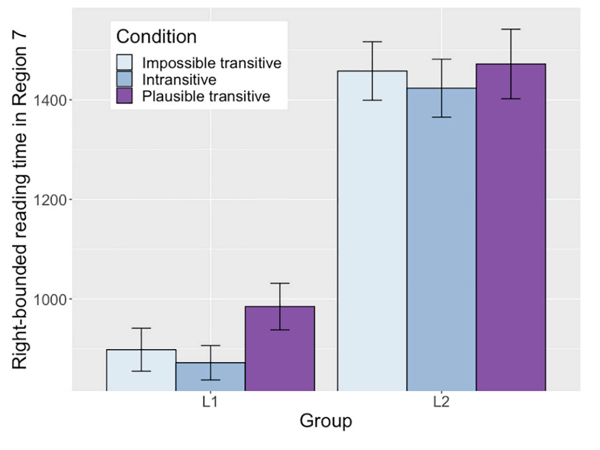


FIGURE 2 | Right-bounded reading time in Region 7 in each condition in each group.

Second, there was a suggestion for a semantic anomaly effect with the L2 group at the spill-over region. The first pass times at the post-verbal region (Region 6) showed that only the L2 group experienced processing cost in the Impossible transitive condition compared to the Plausible transitive condition. Although the effect was marginal, the L2 group showed a numerically longer reading time when the verb information in the previous verb region resulted in a semantically anomalous direct object analysis interpretation as in *the writer interviewed the letter* compared to when the direct object analysis was semantically plausible as in *the writer interviewed the celebrity*. The cost for semantic implausibility was not observed with the L1 group.

Third, both groups processed sentences differently depending on the verb type. At the post-verbal region (Region 6), right-bounded and second pass times showed that both L1 group and L2 group spent longer to read this region in the Plausible transitive condition than the Intransitive condition. This indicates that both groups experienced a larger reanalysis cost for the Plausible transitive condition, and that neither group was forced to perform structural revision in the Intransitive condition. The results provide evidence that both L1 and L2 groups used the subcategorization frame information of the verbs in structural analysis.

Finally, the L1 group showed processing difficulty in revising the initial direct object analysis in the Plausible transitive condition. The right-bounded times in the sentence-final region (Region 7) showed that L1 group regressed to the earlier regions when the direct object analysis became infeasible in the Plausible transitive condition. The L1 group initially analyzed the preceding NP as the verb's direct object in the Plausible transitive condition (e.g., *The writer interviewed the celebrity* . . .), but on encountering the sentence-final prepositional phrase (e.g., *at the conference*), they noticed that the analysis they had adopted was inconsistent with the sentence continuation. They thus revised the preposition as a part of a prepositional phrase (e.g., *smiled about* as in *the writer smiled about the letter*. . .).

GENERAL DISCUSSION

The current study addressed the question of whether L2 learners are able to exploit the verb's subcategorization information in online syntactic processing. We contrasted three possible accounts; the initial inaccessibility account, the intransitivity overriding account, and the fuzzy subcategorization frame account. The initial inaccessibility account that L2 learners possess the correct subcategorization knowledge, but intransitivity information cannot be immediately accessed during online structural analysis. The intransitivity overriding account that L2 learners can access subcategorization information but the use of intransitivity information is overridden by the presence of a postverbal noun due to the strong preference for analyzing the noun as a verb's direct object. Under this account, it is predicted that L2 learners would be able to use intransitivity information as long as an NP does not directly appear following an intransitive verb. The fuzzy subcategorization frame account that L2 learners' structural information is represented somewhat in a fuzzy way so that it allows certain ambiguities in the prescription of subcategorization frames. Under this account, L2 learners are expected to be unable to make structural judgment based on verb's subcategorization information. As a result, L2 learners would tolerate the incorrect transitive use for an obligatory intransitive verb and form a plausible interpretation, causing little or no processing difficulty. We tested these accounts by examining the processing of the locally ambiguous unbounded dependency structure with native speakers and Japanese-English L2 learners. The results from the eye-tracking experiment indicated that the filler-gap dependency structure was processed using verb's subcategorization information by both native speakers and L2 learners. The pattern of the results, however, was not identical between the two groups and we now discuss both the common findings and the differences including the implications of these results.

The results indicated that both groups processed sentences differently depending on the verb's subcategorization information. Both groups experienced reduced processing cost when the verb was intransitive as indexed by shorter reading times at the post-verbal region in the Intransitive condition than the Transitive conditions. This reflects that verb's intransitivity information helped readers to adopt the correct sentence structure; they adopted the direct object analysis only when the verb was optionally transitive but not when the verb was obligatory intransitive. It was also shown that L2 learners experienced an extra cost in processing intransitive verbs compared to optionally transitive verbs. This was indexed by a longer reading time at the verb region in the intransitive condition than in the transitive conditions. This reflects that L2 learners had a strong expectation for the upcoming verb to be transitive, and experienced difficulty on encountering an intransitive verb. These results provide evidence for the use of verb's subcategorization information in L2 processing, and are incompatible with the previous studies which found that intransitivity information was ignored during L2 comprehension. This is most likely because previous studies used a structure in which an NP appears directly following a verb, and this structure perhaps triggered an L2 learner's strong preference for a direct object analysis, which overrode intransitive subcategorization restrictions in L2 processing. The results of the current study thus provide support for the overriding account, showing that L2 learners can immediately use intransitivity information to guide their structural analysis when there are no overriding constraints. One possible drawback is that different verbs had to be used for the transitive and intransitive conditions, and any potential differences in the frequency or the number of characters between the two types of the verbs could have contributed to the difference in reading times between the conditions. This is perhaps most relevant for the finding of longer reading times at the verb region with the intransitive verbs than with the transitive verbs with L2 learners. We checked the familiarity of the two sets of the verbs using a L2 database (Yokokawa, 2006) and confirmed that there was no reliable difference between the transitive and intransitive verbs. This, however, is a null effect based on one database and we thus cannot reject the possibility for the influence of lexical factors completely.

The results also revealed that L2 learners experienced immediate processing disruption on encountering an intransitive verb. This processing pattern was unique to the L2 learners, and we interpret this finding as being consistent with the hyper-active gap filling hypothesis proposed by Omaki et al. (2015). Although the hypothesis was originally proposed for L1 processing, we think it is plausible that L2 learners show a stronger tendency for the hyper-active gap filling because L2 learners would rely more on category-general transitive knowledge in processing sentences in their L2 (see also Omaki and Schulz, 2011). As discussed in the mono-transitivity information as category-general hypothesis in the study of Van Gompel et al. (2012), transitive information applies to almost all verbs whereas the occurrence of an intransitive structure is much less frequent (Roland et al., 2007). Considering the limited exposure to L2, it is reasonable to think that L2 learners' experience with intransitive verbs is even more limited compared to the native speakers'. Thus, the lexically specific knowledge of whether a particular verb should be used in an intransitive structure is less solid with L2 learners, and this leads them to have strong expectations for a more general transitive structure. As a result, L2 learners were surprised to see an intransitive verb because it violates the expectation they generated prior to the verb.

The reading patterns at the post-verbal region also suggested that L2 learners experienced processing difficulty due to semantically anomalous direct object analysis in the Impossible transitive condition at the spill-over region with L2 learners. Native speakers did not show processing difficulty in the Impossible transitive condition, most likely reflecting that native speakers adopted the correct analysis using semantic information without processing cost. The finding that the cost for semantic mismatch was observed only with L2 learners but not with native speakers suggest that native speakers immediately revised the structure for the correct structure when the direct object analysis was semantically anomalous. In contrast, L2 learners adopted the direct object analysis even when the interpretation was semantically anomalous and experienced processing difficulty. The results are consistent with the previous research, which showed that L2 learners stick to the initially adopted semantically anomalous direct object analysis and this causes a delay in revising the sentence to the correct structure in L2 processing (Roberts and Felser, 2011).

With native speakers, processing cost due to structural reanalysis was observed at the sentence final region in the Plausible transitive condition. The direct object analysis in the Plausible transitive condition was plausible up to this region, and when native speakers encountered information that was inconsistent with the analysis in the sentence-final region, they spent longer time re-reading the earlier regions. With L2 learners, no evidence for structural reanalysis cost in the Plausible transitive condition was observed. This might suggest the possibility that L2 learners did not reach the correct structural interpretation in the Plausible transitive condition. Some previous studies have suggested that readers do not always engage in fully detailed analysis but instead use heuristics in processing sentences. They have shown that readers often preserve an initially adopted incorrect analysis even after the initial analysis turns out to be incorrect (Christianson et al., 2001; Van Gompel et al., 2006). There is also a finding that readers perform incomplete reanalysis more with complex sentence structure when the semantic information supports the initial incorrect analysis (Nakamura and Arai, 2015). The view that language users do not always build a complete sentence representation, known as the Good Enough approach, suggests the possibility that our participants in the present study did not ultimately reach the correct analysis. Given that there was no evidence for structural reanalysis cost in the Plausible transitive condition with L2 learners, L2 learner's structural reanalysis in the Plausible transitive condition may possibly have ended up incomplete, with the incorrect direct object analysis retained as the final interpretation of the sentences. In the experiment, we did not include questions about the final interpretation because they would draw participants' attention to the structural ambiguity (e.g., *Did the writer interviewed the celebrity?* for 7a), causing them to notice the purpose of our experiment. We therefore cannot know what final representation they constructed with these sentences and this issue is left to future research.

The initial inaccessibility account predicted that L2 learners cannot use verb's subcategorization information during online sentence processing even though they have the experience-based knowledge about possible argument structures. This account was dismissed because our results provided clear evidence that intransitive verbs and transitive verbs were processed differently by L2 learners. The fuzzy subcategorization frame representation account predicted that L2 learners' fuzzy representation of argument structure information would cause them not to exclude the argument structure that is not a part of the verb's subcategorization frame. As a result, L2 learner's tolerate the incorrect transitive use for an intransitive structure, and form an interpretation out of the ungrammatical sentence using semantic information. This account was also dismissed because our results showed that L2 learners experienced a facilitatory effect due to the semantically anomalous direct object analysis only in the Impossible transitive condition but not in the Intransitive condition. Instead, our results provided support for the intransitivity overriding account, which predicted that L2 learners are able to apply the knowledge as long as there is no overriding information, i.e., when the sentence structure does not have an NP directly following an intransitive verb.

We now explore the possibility of an alternative explanation that the use of verb subcategorization information by our L2 learners may be accounted for by L1 transfer. In Japanese, sentence structure is typically expressed by two grammatical features; case marking and verb inflection (Jacobsen, 1991). For example, to use the verb *break* (*kowasu*) is transitive in Japanese; the verb should be preceded by an NP with the accusative case particle *o* as in *kabin o kowasu* (vase-ACC break, "break a vase"). For the same verb to be used intransitively, the verb should occur with the suffix *-eru* as in *kabin ga kowareru* (vase-NOM break, "the vase broke") in which case the verb is preceded by the NP with the nominative case particle *ga*. However, the suffixes used to mark transitive and intransitive forms of a verb are not always consistent across all verbs and it is thus not always possible to tell the structure from the verb form (For example, the suffix *-eru* is used to mark an intransitive form for the verb *kowasu* "break" as in the example above, but the same suffix *-eru* is used to mark a transitive counterpart for the verb *aku* "open," as in *akeru*). Furthermore, Japanese allows arguments to be expressed implicitly as in *Kare-ga kowashita,* "He broke (something)." Thus, only when the NP with the accusative case marker is explicitly present in the sentence, can one rule out the intransitive structure. In the filler-gap dependency structure used in the current study, an accusative NP preceded the verb, which is in a way similar to the Japanese head-final construction. Thus, it may be possible that our Japanese participants saw some similarity between the cleft sentences used in the current study and Japanese transitive sentences even though the NP in the former does not contain case marker, which may have contributed to some extent to the prediction about the upcoming transitive verb. The degree of contribution of L1 transfer on our findings needs to be confirmed in future research by testing different populations of L2 learners.

In sum, the results of the current study together demonstrated that L2 learners are able to use lexically specific intransitivity information to guide their structural analyses when an NP was not present directly following a verb, thus providing support for the overriding account. The overriding account can also account for the previous studies that failed to observe an effect of subcategorization information in L2 processing; L2 learners possess subcategorization knowledge for individual verbs, but their intransitivity information often failed to guide the initial syntactic analysis. Our study, together with previous studies, suggested that the intransitivity information would be overridden when an NP appeared directly following the verb due to a strong bias toward the direct object analysis but can be used to process the sentence structure when the NP is dislocated from the post-verbal position and preceded the verb as in a filler-gap dependency structure.

CONCLUSION

This study investigated the effect of lexically specific verb information as well as semantic information in processing a temporarily ambiguous unbounded dependency structure with Japanese speakers learning English as an L2. Previous research that examined the use of the verb's subcategorization information in L2 processing showed that L2 learners ignore intransitivity information in online structural analyses. The current study examined the possibility that verbal intransitivity information is in fact accessed in L2 parsing but the information tends to get overridden by a strong preference to analyze an NP directly following a verb as the verb's direct object. The results of eye-tracking experiments in reading unbounded dependency structures showed that L2 learners treated intransitive verbs differently from transitive verbs in incremental structural analysis. It was also revealed that L2 learners had stronger expectations for a transitive structure than native speakers, and that L2 learners required a longer time to revise the sentence structure when the semantic information did not support the initially adopted analysis. To conclude, this study provided evidence for the use of verb subcategorization information as well as semantic information in L2 processing, demonstrating that verb's subcategorization information is not ignored in L2 processing and L2 learners can make use of verb information in situations where no overriding information is present.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Massachusetts Institute of Technology Committee on the Use of Humans as Experimental Subjects. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CN, MA, YH, and SF contributed to the conception and design of the study. CN organized the data collection and performed the statistical analysis. MA supported the statistical analysis process. YH and SF contributed to interpreting the results from the statistical analysis. CN and MA wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by the JSPS KAKENHI Grant-in-Aid for the Scientific Research (B) 19H01279 and JSPS KAKENHI Grant-in-Aid for the Promotion of Joint International Research (B) 21KK0006.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.689137/full#supplementary-material

REFERENCES

Arai, M., and Keller, F. (2012). The use of verb-specific information for prediction in sentence processing. *Lang. Cogn. Process.* 28, 525–560. doi: 10.1080/01690965.2012.658072

Baayen, R. (2008). *Analyzing Linguistic Data: A Practical Introduction To Statistics Using R*. Cambridge: Cambridge University Press.

Christianson, K., Hollingworth, A., Halliwell, J., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cogn. Psychol.* 42, 368–407. doi: 10.1006/cogp.2001.0752

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholinguist.* 27, 3–42. doi: 10.1017/s0142716406060024

Council of Europe [COE] (2021). *The CEFR Levels. Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Strasbourg: Council of Europe.

Dekydtspotter, L., Schwartz, B. D., and Sprouse, R. A. (2006). "The comparative fallacy in L2 processing research," in *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006): The Banff Conference*, eds M. Grantham O'Brien, C. Shea, and J. Archibald (Somerville: Cascadilla Press), 33–40.

Dekydtspotter, L., and Seo, H.-K. (2017). Transitivity in the processing of intransitive clauses: a category-based prediction in low-intermediate learners of English. *Stud. Second Lang. Acquis.* 39, 527–552.

Dussias, P. E., and Cramer Scaltz, T. R. (2007). Spanish-English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychol.* 128, 501–513. doi: 10.1016/j.actpsy.2007.09.004

Ferreira, F., and Henderson, J. M. (1990). Use of verb information in syntactic parsing: evidence from eye movements and word-by-word self-paced reading. *J. Exp. Psychol.* 16, 555–568. doi: 10.1037//0278-7393.16.4.555

Francis, W. N., and Kucera, H. (1982). *Frequency Analysis Of English Usage*. Boston: Houghton Mifflin.

Frenck-Mestre, C., and Pynte, J. (1997). Syntactic ambiguity resolution while reading in second and native languages. *Q. J. Exp. Psychol. A* 50A, 119–148.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., and Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *J. Mem. Lang.* 37, 58–93. doi: 10.1006/jmla.1997.2512

Hoover, M. L., and Dwivedi, V. D. (1998). Syntactic processing by skilled bilinguals. *Lang. Learn.* 48, 1–29. doi: 10.1111/1467-9922.00031

Hopp, H. (2010). Ultimate attainment in L2 inflection: performance similarities between non-native and native speakers. *Lingua* 120, 901–931.

Jacobsen, W. M. (1991). *The Transitive Structure of Events in Japanese, Studies in Japanese Linguistics*. Tokyo: Kuroshio Publishers.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Appl. Psycholinguist.* 25, 603–634. doi: 10.1016/0093-934x(89)90059-x

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Lang. Learn.* 57, 1–33. doi: 10.1017/s0140525x0600906x

Juffs, A. (1998). Some effects of first language argument structure and morphosyntax on second language sentence processing. *Second Lang. Res.* 14, 406–424. doi: 10.1191/026765898668800317

Juffs, A., and Harrington, M. (1995). Parsing effects in second language sentence processing: subject and object asymmetries in wh-extraction. *Stud. Second Lang. Acquis.* 17, 483–516.

Juffs, A., and Harrington, M. (1996). Garden path sentences and error data in second language sentence processing. *Lang. Learn.* 46, 283–323.

Kennison, S. M. (2001). Limitations on the use of verb information during sentence comprehension. *Psychon. Bull. Rev.* 8, 132–138. doi: 10.3758/bf03196149

Lee, E., Lu, D., and Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Biling. Lang. Cogn.* 17, 234–236.

MacDonald, M. C., Pearlmutter, N. J., and Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychol. Rev.* 101, 676–703. doi: 10.1037/0033-295x.101.4.676

Marinis, T., Roberts, L., Felser, C., and Clahsen, H. (2005). Gaps in second language processing. *Stud. Second Lang. Acquis.* 27, 53–78.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *J. Mem. Lang.* 38, 283–312. doi: 10.1006/jmla.1997.2543

Mitchell, D. C. (1987). "Lexical guidance in human parsing: locus and processing characteristics," in *Attention and Performance XII: The Psychology of Language*, ed. M. Coltheart (Hillsdale: Lawrence Erlbaum Associates), 601–618.

Nakamura, C., and Arai, M. (2015). Persistence of an initial misinterpretation without referential ambiguity. *Cogn. Sci.* 40, 909–940. doi: 10.1111/cogs.12266

Nakamura, C., Arai, M., and Harada, Y. (2013). The use of verb subcategorization information in processing garden-path sentences: a comparative study on native speakers and Japanese EFL learners. *Stud. Lang. Sci.* 12, 43–69.

Nakamura, C., Arai, M., and Harada, Y. (2019). Difference between L1 and L2 language processing in the use of subcategorization information: evidence from syntactic priming. *Gengokenkyu* 155, 1–33.

Omaki, A., Lau, E. F., White, D. I, Dakan, M. L., and Phillips, C. (2015). Hyper-active gap filling. *Front. Psychol.* 6:384. doi: 10.3389/fpsyg.2015.00384

Omaki, A., and Schulz, B. (2011). Filler-gap dependencies and island constraints in second-language sentence processing. *Stud. Second Lang. Acquis.* 33, 563–588.

Papadopoulou, D., and Clahsen, H. (2003). Parsing strategies in L1 and L2 sentence Processing: a study of relative clause attachment in Greek. *Stud. Second Lang. Acquis.* 25, 501–528. doi: 10.1017/s0272263103000214

Pickering, M. J., and Traxler, M. J. (1998). Plausibility and recovery from garden paths: an eye-tracking study. *J. Exp. Psychol.* 24, 940–961.

Pickering, M. J., and Traxler, M. J. (2003). Evidence against the use of subcategorization frequency in the processing of unbounded dependencies. *Lang. Cogn. Process.* 18, 469–503.

Pickering, M. J., Traxler, M. J., and Crocker, M. W. (2000). Ambiguity resolution in sentence processing: evidence against frequency-based accounts. *J. Mem. Lang.* 43, 447–475.

Roberts, L., and Felser, C. (2011). Plausibility and recovery from garden-paths in L2 sentence processing. *Appl. Psycholinguist.* 32, 299–331. doi: 10.1017/s0142716410000421

Roland, D., Dick, F., and Elman, J. L. (2007). Frequency of basic English grammatical structures: a corpus analysis. *J. Mem. Lang.* 57, 348–379. doi: 10.1016/j.jml.2007.03.002

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *J. Mem. Lang.* 35, 566–585. doi: 10.1006/jmla.1996.0030

Trueswell, J. C., Tanenhaus, M. K., and Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J. Exp. Psychol. Learn. Mem. Cogn.* 19, 528–553. doi: 10.1037//0278-7393.19.3.528

Van Gompel, R. P. G., Arai, M., and Pearson, J. (2012). The representation of mono- and intransitive structures. *J. Mem. Lang.* 66, 384–406.

Van Gompel, R. P. G., Pickering, M., Pearson, J., and Jacob, G. (2006). The activation of inappropriate analyses in garden-path sentences: evidence from structural priming. *J. Mem. Lang.* 55, 335–362. doi: 10.1016/j.jml.2006.06.004

Yokokawa, H. (2006). *Database for L2 Research & Pedagogy: English Word Familiarity for Japanese EFL Learners*. Tokyo: Kuroshio publisher.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Communication

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fcomm.2021.726443

Fuzzy or Clear? A Computational Approach Towards Dynamic L2 Lexical-Semantic Representation

Xiaowei Zhao[1] and Ping Li[2]**

[1]Department of Psychology and Neuroscience, Emmanuel College, Boston, MA, United States, [2]Department of Chinese and Bilingual Studies, Faculty of Humanities, The Hong Kong Polytechnic University, Hong Kong, China

In this paper, we present a computational approach to bilingual speakers' non-native (L2) lexical-semantic representations. Specifically, based on detailed analyses of the error patterns shown in our previous simulation results (Zhao and Li Int. J. Bilingual. Educ. Bilingual., 2010, 13, 505–524; Zhao and Li, Bilingualism, 2013, 16, 288–303), we aim at revealing the underlying learning factors that may affect the extent of fuzzy category boundaries within bilinguals' L2 representation. Here, we first review computational bilingual models in the literature that have focused on simulating L2 lexical representations, including the Developmental Lexicon II (DevLex-II) model (Zhao and Li, Int. J. Bilingual. Educ. Bilingual., 2010, 13, 505–524; Zhao and Li, Bilingualism, 2013, 16, 288–303), on which the current study is based. The DevLex-II modeling results indicate a strong age of acquisition (AoA) effect: When the learning of L2 is early relative to that of native language (L1), functionally distinct lexical representations may be established for both languages; when the learning of L2 is significantly delayed relative to that of L1, fuzzy L2 representations may occur due to the structural consolidation (or the *entrenchment*) of the L1 lexicon. Next, we explore the error patterns shown in both lexical comprehension and production in DevLex-II. A novel contribution of the current study is that we systematically compare the computational simulation results with empirical findings. Such model-based error analyses extend our previous findings by indicating, especially in the late L2 learning condition, that fuzzy L2 semantic representations emerge and lead to processing errors, including errors in unstable phonology-semantic and semantic-phonemic mappings. The DevLex-II model provides a computational account of the development of bilinguals' L2 representation with reference to the dynamic interaction and competition between the two lexicons. We point to future directions in which fuzzy L2 representations may be overcome, through a framework that highlights the social learning of L2 (SL2) and the embodied semantic representation of the lexicon in the new language (Li and Jeong, Npj Sci. Learn., 2020, 5, 1–9; Zhang, Yang, Wang and Li, Lang. Cogn. Neurosci., 2020, 35, 1223–1238).

Keywords: fuzzy lexical representation, bilingualism, computational modeling, emergentism, DevLex-II

Edited by:
Svetlana V. Cook,
University of MarylandUnited States

Reviewed by:
Junru Wu,
East China Normal University, China
Katsuo Tamaoka,
Hunan University, China

****Correspondence:***
Xiaowei Zhao
xiaoweizhao@gmail.com
Ping Li
pi2li@polyu.edu.hk

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Communication

Received: *16 June 2021*
Accepted: *13 December 2021*
Published: *21 January 2022*

Citation:
Zhao X and Li P (2022) Fuzzy or Clear? A Computational Approach Towards Dynamic L2 Lexical-Semantic Representation. Front. Commun. 6:726443. doi: 10.3389/fcomm.2021.726443

INTRODUCTION

It is widely recognized that at least half of the world's population can use more than one language in their daily lives (Grosjean and Li, 2013). Many of them are bilinguals who are fluent in both of their languages, but many more are individuals who are second language (L2) learners with varying levels of mastery of their L2 depending on various learning and learner factors, such as timing and history of learning L2 (Li, 2013), social interaction needs (Li and Jeong, 2020), context of usage (the Complementarity Principle; Grosjean, 2013), and cognitive abilities [such as working memory; see reviews by Kormos (2015), Wen et al. (2017)]. Behavioral evidence has shown that L2 learners often have comprehension and production problems in L2, particularly with decoding and producing the ambiguous sounds that they are unsure of (such as phonemes not available in their native language, see a review by Gor, 2015). Bilinguals, as compared with monolinguals, may have more difficulties in word recognition tasks (Lemhöfer et al., 2008) or generating fast and accurate names in picture naming or word naming tasks (Gollan et al., 2005; Gollan et al., 2008; Li et al., 2019; Peñaloza et al., 2019). They may also find it difficult to make accurate phonology-semantic or semantic-phonemic mappings in their L2 (Cook et al., 2016) and experience higher rate of tip-of-the-tongue state in L2 (Kreiner and Degani, 2015).

Many theoretical frameworks have been proposed to account for such patterns and difficulties in the L2, which include hypotheses on a fuzzy, weak, or less entrenched lexical-semantic representation/network of L2 (Hernandez and Li, 2007; Diependaele et al., 2013; MacWhinney, 2013; Cook and Gor, 2015). In their recent Ontogenesis Model (OM), Bordag et al. (2021, p.2) argued that a crucial property of L2 lexical representation is *fuzziness*, which "refers to inexact or ambiguous encoding of different components or dimensions of the lexical representation that can be caused by several linguistic, cognitive, and learning-induced factors." (see also Gor et al., 2021 for the Fuzzy Lexical Representations account in this special research topic). Such a view is highly consistent with the concept of "parasitism" proposed by Hernandez et al. (2005), according to which factors such age of acquisition (AoA), proficiency, and in particular competition/interaction between L1 and L2, are responsible for a L2 lexical-semantic representations that become parasitic (and usually fuzzy) to L1 representations (also see Hernandez and Li, 2007). It is thus important to study the dynamic interaction between L1 and L2 and how the L2 learning history can shape the bilinguals' L2 representations. Indeed, an issue of enduring interest in bilingualism research has been how the lexical-semantic system of L2, as a dynamic system, is represented and developed, and subsequently interacts with the L1 system in the bilingual's mind (Li, 2013).

Since the 1950s, there have been many models and hypotheses of bilinguals' lexical representations and processing (see, e.g., Jiang, 2015 for a historical review). These models often offer good explanatory power of bilingual language patterns and have made significant contributions to our understanding of bilingualism. However, most of the models so far are verbally descriptive in nature and have been designed to capture bilingual lexical processing for the mature adult bilingual speakers, rather than accounting for the developmental changes associated with the learning of a L2 or the processes underlying learning. In this paper, we advocate a dynamic approach of bilingual lexical-semantic representation, viewing it as constantly changing and evolving as learning progresses (e.g., Li, 2015). Rather than just taking a snapshot of the static situation (e.g., as end result or outcome), we focus on the underlying mechanisms that may affect the learning process that leads to fuzzy category boundaries within bilinguals' L2 representation. To reach this goal, we use computationally implemented models, which are particularly helpful in helping us to understand the L1-L2 interaction and the emergence of fuzzy representations in L2. Computational models allow the researchers to bring multiple variables and the complex interaction between the variables under systematic control, and test hypotheses about the roles of variables of interest in bilingual representation while holding other variables constant. Such a systematic control of variables is particularly important for studying bilingualism, due to the multitude of potentially confounding variables existed in the natural contexts of bilingual learning that would be otherwise difficult to manipulate in behavioral studies (see Li and Zhao, 2017; Li and Zhao, 2021 for discussions of the role of computational modeling in psycholinguistics).

In this paper, we focus on systematically analyzing the patterns of the inaccurate lexical comprehension and production, along with their underlying mechanisms in a computational bilingual model. Error analysis of second language learners has been an important topic in applied linguistics, especially in foreign language education (see Chapter 2. Ellis, 1994; Swan, 1997). However, it attracts less attention in computational studies, which often focus on what a model structure is capable/competent to achieve in term of empirical behaviors, with simulation errors commonly treated as byproducts of statistical fluctuations (but see the classic connectionist models of U-shaped behavior in monolingual past-tense learning, Rumelhart and McClelland, 1986). Here we advocate a systematic comparison of computational modeling results with a detailed analysis of error patterns that occur in behavioral data from real second language learners, which can help us to understand the role of the interactive mechanisms embodied in the model on the emergence of fuzzy L2 lexical-semantic representations.

In the following sections, we first review computational bilingual models that have previously focused on simulating L2 lexical processing and representations[1], including the Developmental Lexicon II (DevLex-II), an unsupervised connectionist model that includes three basic levels for the representation and organization of linguistic information (see

[1]It is worth noting that there have been many computational studies of other important issues in bilingualism (code switching as one example; see recent publications by Xu et al. (2021), Tsoukala et al. (2021). It is beyond the scope of this paper to review all computational models of bilingualism. Readers are referred to various resources such as the special issue in *Bilingualism: Language and Cognition* [edited by Li (2013), Shirai (2019) book, and the online bibliography of (Li and Zhao, 2020)].

Li et al., 2007; Zhao and Li, 2010; Zhao and Li, 2013). Our current computational simulations are based on the DevLex-II model. In the main part of this paper, we focus on analyzing the late L2 learning condition, in which fuzzy L2 representations (at both semantic and phonological levels) lead to processing errors or confusions, including errors in unstable phonology-semantic and semantic-phonemic mappings. The error patterns are further compared with behavioral data from previous literature and a Second Language Acquisition (SLA) corpus (COPA corpus on TalkBank, Zhang, 2009a). We argue that the DevLex-II model provides a computational account of the development of bilinguals' L2 representation with reference to the dynamic interaction and competition between the two lexicons. We further point to future directions, including studies of fuzzy L2 representations that may be overcome through an approach that highlights the social learning of L2 (SL2) and the embodied semantic representation of the lexicon in the new language (Li and Jeong, 2020; Zhang et al., 2020).

COMPUTATIONAL MODELS OF BILINGUAL LEXICAL REPRESENTATIONS

Interactive Activation Models

IA-based models have been mainly used in simulating patterns in bilingual language processing. The Bilingual Interactive Activation (BIA) model (Dijkstra and van Heuven, 1998) might be the best-known computational model of bilingual lexical processing so far. Similar to the famous monolingual IA model (McClelland and Rumelhart, 1981), BIA has three levels of nodes representing orthographic features, letters, and words. Unlike IA, however, the BIA has linguistic inputs from two languages, and is equipped with a language node level that provides top-down information regarding the language identity of perceived words.

As a successor of the BIA model, BIA+ (Dijkstra and van Heuven, 2002) incorporates semantic and phonological representations into its main component (i.e. the word identification system), along with a nonlinguistic task/decision system. The nonlinguistic task/decision system receives input from the identification system and computes processing steps and determines decision criteria for the simulation task, such as bilingual reading. Dijkstra et al. (2012) applied the BIA + structure to model lexical processing of Dutch-English bilinguals. In their simulations, proficiency and AoA were modeled by adjusting the relative frequency of the L2 words and the size of the lexicon, and the model predicted a gradual increase in processing speed in L2 for the late L2 learners.

Diependaele et al. (2013) used a bimodal interactive activation model to simulate the difference between the frequency effects in L1 and L2 word recognition. Specifically, they reduced the resting levels of the word nodes to simulate the "weaker" lexical memory representations in L2. They also reduced the level of word-word lexical inhibition in the model to simulate the increased competition between similar words caused by the less "precise" lexical representations in L2. Their simulation results were in line with the patterns shown in large-scale English word identification times from three bilingual populations. They concluded that L2 is less entrenched in late L2 learners' lexical system due to low L2 proficiency, and this lower entrenchment could explain the stronger frequency effect in L2 word recognition.

As a new computational model of bilingual representation along this direction, Multilink (Dijkstra et al., 2019) represents the latest efforts by the researchers to scale up the BIA modeling enterprise to a larger and more realistic lexicon (over 1,500 words from both lexicons). Multilink is an interactive based model integrating certain features of BIA+ (Dijkstra and van Heuven, 2002) and the Revised Hierarchical Model (RHM; Kroll and Stewart, 1994). By considering the role of multiple factors such as the frequency, length, orthographic similarity, and phonological neighborhood of words, it has been used to test and verify against empirical data from bilingual word recognition and translation.

It is worth noting that many IA-based models shown above lacked learning/development mechanisms. Their representations were often fixed and their parameters (e.g., resting level) manually adjusted to capture adult bilingual speakers' word processing (see Li and Grant, 2019 for a commentary on the Multilink model). In fact, a wide variety of computational developmental models with a learning mechanism have been implemented for bilingual lexical representations, and they often embrace an emergentist view that static linguistic representations (e.g., words, concepts, and grammatical structures) are emergent properties, dynamically acquired from the learning environment. Common learning algorithms can be classified roughly into *supervised* and *unsupervised* learning (see Zhao, 2017 for a brief introduction). These algorithms are often developed under the framework of connectionist or neural network models (aka Parallel Distributed Processing or PDP models; McClelland et al., 1986; for a bibliography and recent models based on connectionism, see Li and Zhao, 2020).

Developmental Models with Supervised Learning

French (1998) tested a Bilingual Simple Recurrent Network model (BSRN), which was based on the monolingual SRN model (Elman, 1990). The BSRN was trained on intermixed sentences from two artificially generated languages, with its input having a certain probability of switching between the two languages. The model's immediate task was to predict the next word in a sentence given the current word input. After training, a hierarchical cluster analysis was conducted on the hidden-node activations of the BSRN model, and results showed that words from the two languages became separated in the network's internal representations. The simulation results supported the hypothesis that the bilingual input environment itself is sufficient for the development of a distinct mental representation of each language, without invoking separate processing or storage mechanisms for the different languages.

Monner et al. (2013) developed a connectionist model in an effort to address a long-standing issue in bilingual language acquisition: To what extent the entrenchment of one's first language influences the learning of a second language? They tested the "less is more" hypothesis using a recurrent network model (Long Short-Term Memory or LSTM, Hochreiter and Schmidhuber, 1997) that learns the gender assignment and agreement in Spanish and French. In their network, increases of working memory were simulated using new cell assemblies in the model, whereas L1 entrenchment was simulated by training of the network with different length of L1 exposure before the onset of L2 (see more discussion below on this in the DevLex-II model). This approach allowed the researchers to dissociate and specify the effects due to age of L2 onset and those due to memory capacity. The authors concluded that their model supported the "less is more" hypothesis while at the same time showing L1 entrenchment effect as a function of L2 onset time.

During the last decade, there has been a fast-growing use of building semantic representations through the so-called "embedding" methods, which allows researchers to derive words' lexical-semantic representations for distributional properties in natural language usage. Many of the methods were based on supervised learning of large-scale monolingual database (such as *Word2Vec*, Mikolov et al., 2013). Following this direction, researchers have developed interests in cross-language word embeddings (see a brief review and the M2VEC model in Wang et al., 2019). A common strategy has been to build a mapping/transformation matrix between two, usually well pretrained, monolingual word embedding spaces (one as the source/input language and the other the target/output language). Such a strategy often needs a pre-built high-quality dictionary or parallel corpora to align the words/concepts in the two languages. Although this approach makes it a great addition to applied fields such as machine translation, it is less ideal for simulating L2 learners' bilingual lexical representations, which are dynamic and interactive in nature and more than just a mapping between two fully developed monolingual lexical spaces (keeping in mind Grosjean's earlier warning that bilinguals are not the sum of two monolinguals; Grosjean, 1989).

Developmental Models with Unsupervised Learning

In contrast to supervised learning, unsupervised learning algorithms do not use explicit error signal at the output level to adjust the network's weights. Among them, a popular type for bilingual models is the self-organizing map (SOM; Kohonen, 2001). On a SOM, a group of nodes with input connections are arranged on a two-dimensional lattice for the organization of external stimulus patterns. Each node has a vector associated with it to represent the weights of its input connections. At each training step of SOM, all nodes on the map are presented with an external input pattern (e.g., the phonological or semantic representation of a word). Some nodes will be activated, according to how similar their input weight vectors are to the input pattern; the node that has the highest activation is declared as the Best Matching Unit (BMU). A BMU's weight vectors are adjusted, along with the weights of its neighboring nodes to become more similar to the input pattern. As the result, they will respond to the same or similar inputs more strongly next time. Initially activation occurs in large areas of the map, but gradually learning becomes more focused and the neighborhood size reduces. This self-organizing process continues until all inputs have found some maximally responding nodes as their BMUs. Eventually, the map falls into a topography-preserving state, which means that inputs with similar features will activate nodes in nearby regions, yielding meaningful activity bubbles that can be visualized on the map. This property of the SOM enables researchers to explicitly examine and visualize the emergence of lexical-semantic structures in their models. It is a more desired feature that is absent in some supervised learning models as discussed above, where the internal representation is often "hidden" from the outside and needs to be analyzed through mathematical tools such as Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA).

Li and Farkas (2002) proposed a self-organizing model of bilingual processing (SOMBIP). The model was based on the SOM described above, with training data derived from linguistic corpora. The SOMBIP model included two SOM maps connected via Hebbian learning. One SOM was trained on the phonological representations of words and the other on the semantic representations of words. The SOMBIP learned bilingual input with mixed English and Chinese words simultaneously; and the frequency of the bilingual words exposed to the network was modulated according to the corpus data, rather than to an ad hoc probability of language switching as in BSRN. In the SOMBIP, the simultaneous learning of Chinese and English led to distinct lexical representations for the two languages, as well as structured semantic and phonological representations within each language. The SOMBIP also simulated a novice learner by having limited linguistic experience so that the network was exposed to fewer sentences in L2. It was shown that the novice network's representation of the L2 was more compressed and less clearly delineated, compared to that of the "proficient" network. The SOMBIP possessed more realistic linguistic and developmental properties than previous bilingual models and later evolved into the DevLex (Developmental Lexicon) models (Li et al., 2004; Li et al., 2007; Zhao and Li, 2010; Zhao and Li, 2013, see below for further discussions).

Recently, Peñaloza, et al. (2019) presented a SOM-based model to simulate the effect of AoA and L2 exposure on bilingual lexical access. Their model, BiLex, includes three connected SOMs, with one for common semantic/conceptual representation, and two for separate phonological representations of different languages (i.e. English and Spanish). Specifically, the model was applied to simulate the picture naming data on a case-by-case basis for 28 bilingual participants with different L2 proficiency levels and 5 monolinguals (as the base line). Their model incorporated important variables underlying the patterns of bilingual behavior, including language history regarding age of L2 acquisition, proficiency, exposure and language use. The best-fit set of values for parameters representing these variables were found based on an evolutionary algorithm

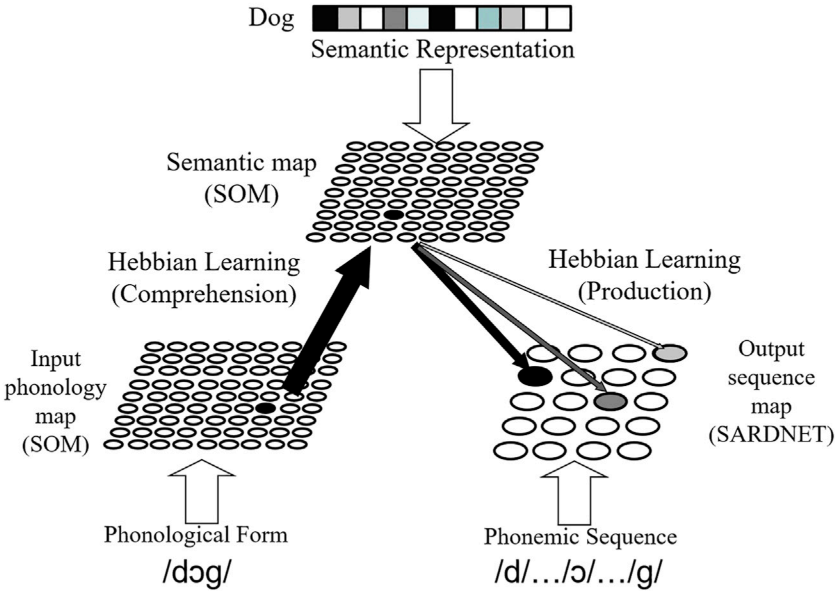


FIGURE 1 | The architecture of the DevLex-II model.

(Back, 1996) to model the data of each participant. The model's close match with real behavioral data from individual participants is a testimony that computational models, when properly constructed, can closely reflect realistic linguistic processes. Importantly, their simulations showed that early AoA and increased exposure can lead to well-organized representations on L2 phonological map and higher picture naming performance, while late AOA and limited exposure can lead to poor representations on L2 phonological map and lower picture naming performance. This pattern is in line with our simulation results from the DevLex-II, which we discuss next. The basic structure of Bilex resembles the well-developed theoretical framework of the RHM (Kroll and Stewart, 1994). However, the use of two predefined phonological maps for L1 and L2 assumes separated lexical representations of the two languages from the beginning of L2 learning, which could be problematic and makes it hard to simulate cross-language interferences. As discussed below, distinct or intermixed bilingual representations may emerge in the same underlying system such as the DevLex-II model through learning in the SOM or other computational algorithms.

THE DEVLEX-II MODEL: A SKETCH

The DevLex-II model is the main computation architecture of the current study. As an unsupervised multi-layer neural network model, the DevLex-II model has been successfully implemented in both monolingual and bilingual language learning (Li et al., 2007; Zhao and Li, 2010; Zhao and Li, 2013). As depicted in **Figure 1**, it consists of three basic levels for the representation and organization of linguistic contents, corresponding to phonological information, semantic information, and the output sequence of the lexicon, respectively. The core of the DevLex-II model is a SOM that handles lexical-semantic representation. This SOM is connected with two other SOMs, one for input (auditory) phonology, and the other for articulatory sequences of output phonology. During the training of DevLex-II, the semantic representation, input phonology, and output phonemic sequence of a word are simultaneously presented to and processed by their corresponding maps of the network, and the associative connections between maps are trained by the Hebbian learning rule. After the cross-map connections are stabilized, the activation of a word form can evoke the activation of a word meaning via form-to-meaning links, and we define this process as word comprehension in our model. Similarly, the activation of a word meaning can trigger the activation of an output sequence via meaning-to-sequence links, and we define this process as word production.

We have applied the DevLex-II model in various studies to simulate bilingual language learning (e.g., Zhao and Li, 2010) and cross-language priming (Zhao and Li, 2013). Specifically, to increase the connection to empirical data, we presented the model with an English-Chinese bilingual lexicon made up of 1,000 real words extracted from MacArthur-Bates Communicative Development Inventories (the CDI; English: Dale and Fenson, 1996; Chinese: Tardif et al., 1999). The input to the model was coded as vector representation of the phonemic, phonological, or semantic information of these words (see Zhao and Li, 2010 for technical details).

Simulations in Zhao and Li (2010), Zhao and Li (2013) included three learning conditions: simultaneous, early, and late, and the target lexicons were sent to the model in stages.

TABLE 1 | Number of each type of comprehension errors in L2 (English) across different late L2 learning stages. Results averaged over 5 simulations for each learning stage. Standard deviations are listed within the parentheses.

L2 Lexicon size	Late L2 learning									
	50	100	150	200	250	300	350	400	450	500
Total number of errors	0.4 (0.5)	28.4 (5.7)	41.4 (6.9)	52.4 (4.4)	65.8 (5.1)	63.8 (6.7)	73.2 (11.3)	100.6 (9.0)	97.6 (6.9)	104.8 (7.1)
Phonological confusions	0.4 (0.5)	6.2 (1.3)	7.2 (1.5)	12 (1.6)	17.8 (5.5)	20.6 (5.7)	25.8 (5.0)	33.4 (4.7)	35.2 (3.7)	45.2 (5.4)
Semantic confusions	0 (0)	8.8 (1.1)	15.6 (2.7)	18.2 (4.4)	20 (2.4)	24 (4.1)	26 (6.1)	35.8 (2.4)	33.2 (3.9)	32.2 (2.5)
Uncategorized Errors	0 (0)	6.6 (3.8)	12.6 (1.7)	15.6 (3.4)	23.2 (6.9)	17.8 (4.0)	20.2 (5.4)	30 (5.8)	28.2 (5.1)	25.4 (5.9)
Cross-language Errors	0 (0)	7.4 (2.1)	7.2 (6.1)	7 (4.5)	6.2 (5.2)	4.8 (6.3)	5.4 (3.2)	5.2 (6.1)	5.2 (4.0)	6.4 (8.3)

The simultaneous learning condition was designed to simulate a simultaneous bilingual learner who is exposed to both languages from early on. In the two sequential learning conditions, learning of L2 was delayed relatively to that of L1, either only slightly (early learning: after 100 L1 words were presented) or significantly (late learning: after 400 L1 words were presented). The exposure to L2 words in all three conditions was 10 stages, with 50 more new L2 words added at each stage.

The setup in Zhao and Li (2010), Zhao and Li (2013) allows a meaningful comparison of the three learning conditions on the effects that the consolidation of lexical organization in one language (usually L1) has on the lexical representation in the other language (usually L2). The modeling results indicate a strong age of acquisition (AoA) effect: When the learning of L2 is early relative to that of L1, functionally distinct lexical representations may be established for both languages; when the learning of L2 is significantly delayed relative to that of L1, fuzzy L2 representations occur due to the structural consolidation (or the "entrenchment") of the L1 lexicon and its impact on the L2 lexicon.

The present study extends the work of (Zhao and Li, 2010; Zhao and Li, 2013) to examine error patterns, using the same model structure and training parameters (see the Methods section in Zhao and Li, 2010 for technical details). Our analyses below focus on the sequential bilingual learning stages, specifically, on a variety of errors produced by the DevLex-II model under the early and late L2 learning conditions. These error patterns provide a window into the developmental changes underlying lexical-semantic representation, for not only the bilingual's L2 but also its interaction with the L1. They are also evaluated against empirical data that reflect patterns of L2 learners' pronunciation errors (e.g., Cutler et al., 2004; Zhang, 2009b; Wang and Chen, 2020).

ANALYSES OF ERROR PATTERNS IN BILINGUAL DEVLEX-II

L2 learners often have comprehension and production problems in their L2, producing errors that deviate from L1 speakers or listeners. In the current study, we have found interesting error patterns in DevLex-II's comprehension and production performance from our computational modeling. It is important to note a few general points here: 1) These errors show similar general patterns when either language is the L2 (Chinese or English), but the specific errors are language-dependent; 2) Our modeling parameters were held constant; being able to simulate both languages as the target L2 while holding modeling parameters constant shows an important and flexible feature of computational modeling; 3) We determined our modeling errors through examination of BMUs in the SOM maps: 1) If the activated unit on the semantic map is the BMU of the correct word meaning, it is taken that our network correctly comprehends this word; otherwise, the network makes a comprehension error; 2) If the activated units on the phonemic map match the BMUs of the phonemes making up the word in the correct order, it is taken that our network correctly produces this word; otherwise, the network makes an error in production.

Comprehension Errors

In **Tables 1** and **2**, we listed the average numbers of different types of comprehension errors that DevLex-II made at different stages of the late L2 learning condition (**Table 1**) and early L2 learning condition (**Table 2**). The results were averaged over 5 simulations trained with different random seeds, and the standard deviations were listed within the parentheses. Overall, four types of patterns were observed.

First, our model for late L2 learners showed a large proportion of errors related to phonological confusions/interferences, errors that were mainly with the comprehension of L2 (45.2 out of 104.8 total errors at the final stage in **Table 1**). For example, an activation of the English word *think* on the input phonology map led to the activation of *sink* on the semantic map. This type of error might be caused by the similarity in sound between two words in the L2 (and their representation similarities on the input phonology map): the/θ/versus/s/difference in L2 English does not exist in L1 Chinese, and therefore/s/took the place of/θ/. Other examples included *stove-stone*, *bump-jump*, *glass-grass*, *pull-pool*, *she-see*, *bug-big*, *light-like*, *blue-blow*, *chair-hair*, *wash-watch* (English as L2); *qing3* ("invite")-*qin1* ("kiss")[2], *zang1* ("dirty")—*zhang1* ("piece"), *bai2* ("white")—*bei4* ("carry") (Chinese as L2), and many more.

[2]The number in the Chinese phonetic transcription indicates the tone of the corresponding word

TABLE 2 | Number of each type of comprehension errors in L2 (English) across different early L2 learning stages. Results averaged over 5 simulations for each learning stage. Standard deviations are listed within the parentheses.

L2 Lexicon size	Early L2 learning									
	50	100	150	200	250	300	350	400	450	500
Total number of errors	42.8 (6.5)	18.2 (40.7)	26.6 (59.5)	0 (0)	3.6 (0.9)	6 (2.1)	7 (3.1)	16.8 (4.0)	21.2 (5.4)	28.2 (6.0)
Phonological confusions	1 (0.7)	3 (6.7)	5.6 (12.5)	0 (0)	1.6 (0.5)	4.2 (2.5)	4.2 (1.3)	6.6 (1.5)	12.6 (4.3)	13.4 (2.9)
Semantic confusions	4.6 (2.4)	1.8 (4.0)	2 (4.5)	0 (0)	1.4 (0.5)	1.4 (0.5)	2.6 (0.9)	7.6 (2.6)	7.8 (3.0)	12 (3.2)
Uncategorized Errors	35.2 (4.1)	10.4 (23.3)	15.2 (34.0)	0 (0)	0.6 (0.5)	0.8 (0.8)	0.6 (0.9)	2 (0.7)	1.8 (1.3)	4.4 (2.9)
Cross-language Errors	2.2 (3.9)	3.2 (7.2)	3.8 (8.5)	0 (0)	0 (0)	0 (0)	0 (0)	0.8 (0.4)	0 (0)	0 (0)

Second, semantic similarities also led to a large proportion of comprehension errors in our model (32.2 out of 104.8 total errors at the final stage). For example, an activation of the Chinese word *ge1ge* ("older brother") on the input phonology map led to the activation of *di1di* ("younger brother") on the semantic map. This is an example of incorrect comprehension (from phonology to semantics) due to within-language semantic interference. Most of these errors were within the L2 itself, such as *dog-cat*, *car-boat*, *pen-pencil*, *kick-drop*, *cut-tear*, *a-an*, *bench-couch*; and *hei1* ("black")-*lv4* ("green"), *mi4feng1* ("bee")-*ma3yi3* ("ant"), *ya1zi* ("duck")-*gong1ji1* ("rooster"). Such types of error might reflect the overlap/similarities between the representations of the words on the semantic map of our model.

Third, for the late L2 condition, certain comprehension errors could not be clearly categorized (25.4 out of 104.8 total errors at the final stage). For example, some activations on the input phonology map were not able to generate the activations of a BMU associated directly with a meaning (i.e., the model failed to comprehend the sound of the word). Also, some word forms evoked the activation of word meanings that were not related in any meaningful way. Examples included *will-jelly*, *nurse-little*, *sun-cheese*. Such errors were rare in our early L2 learning condition (see **Table 2**) and might reflect the unstable/inaccurate form-to-meaning links inconsistently built under our late L2 learning condition.

Finally, a very small proportion of comprehension errors were due to cross-language similarities (6.4 out of 104.8 total errors at the final stage as shown in **Table 1**). Most of them were due to phonetic similarities (i.e., cross-language homophones) and originated from L2: *a-e2* ("goose"), *tongue-tang2* ("sugar"), *hair-hei1* ("black"), *ear-ye2ye* ("grandpa"), *when-wan3* ("bowl") (see Li and Farkas 2002, for similar errors). Cross-language comprehension errors due to semantic similarities were found too, but their occurrence was extremely rare and only in the direction from L2 sound representation to L1 semantic representation. We only observed a few examples at the beginning of late L2 learning condition, such as *Mao1* ("cat")-*bear*; *shou3* ("hand")-*toe* (Chinese as L2), and *kiss-qin1* ("kiss"), *owl-ya1zi* ("duck"), *touch-reng1* ("throw") (English as L2).

Overall, in **Table 1**, we could observe the increment of total number of comprehension errors as more L2 words entered the training. In addition, roughly similar proportions of the three main types of comprehension errors occurred across the training stages of late learning condition (with a relative larger portion of phonological confusions towards the end of learning).

For the early L2 learning shown in **Table 2**, it is interesting to note that the total number of comprehension errors stayed at a low level across most stages, and most errors were regular confusions in either meaning or sound (13.4 and 12 out of 28.2 total errors respectively at the end), and the proportion of uncategorized errors was small (4.4 out of 28.2 at the end)[3]. The cross-language comprehension errors occurred rarely under the early L2 condition (0 at the end). Such a pattern was different from the late L2 learning where uncategorized errors were common, and it implied that the unstable form-to-meaning links (which often caused uncategorized errors) might not be a main driving force for early L2 learners' comprehension errors (which was low in number in the first place). In addition, from a developmental perspective, our model made fewer comprehension errors generally as L2 learning progressed.

Production Errors

In monolingual simulations (see Table 1 in Zhao and Li, 2013), DevLex-II showed lexical confusions, omissions, replacements, and incorrect sequencing of phonemes in production. The current bilingual simulations also showed many of the same patterns. However, there were certain production error patterns unique to our bilingual simulations, in particular for the late L2 learning condition. In **Table 3**, we present a summary of four types of production errors made by DevLex-II under both the late and early L2 (English) learning conditions: phoneme confusions, phoneme replacement, incorrect sequencing, and semantic confusions. The results were the averaged values across 5 simulations trained with different random seeds, and the standard deviations were in parentheses.

Under the late L2 learning condition, a large amount of production errors was due to confusions on the phonemes unique to L2 (37.4 out of 139.8 total errors). As shown on the row labeled as "phoneme confusions", sometimes our late L2 learning simulations could not distinguish between L2-unique

[3]Except at its earliest stage when the L1 was not fully developed, most such errors were comprehension failures without generating any meaningful BMU activations on the semantic map

TABLE 3 | Number of each type of production errors in L2 (English). Results averaged over 5 simulations for each learning condition. Standard deviations are listed within the parentheses.

	Late L2 learning	Early L2 learning	Error examples
Total errors	139.8 (35.2)	74.2 (26.1)	
Phoneme confusions	37.4 (4.3)	0.4 (0.9)	[ɔ:] confused with [ɒ] in English [ʐ] in Chinese confused with [z] in English
Phoneme replacement	67.4 (20.4)	29.6 (19.0)	*skate*[skeɪt] pronounced as [skeɪtʰ] *were* [wɜ:] pronounced as [wa:]
Incorrect Sequencing	78 (32.2)	60 (24.8)	*grandpa* ['ɡɹæmpɑ:] pronounced as ['ɡæɹmpɑ:] *telephone* ['tɛləfəʊn] produced as ['tɛəlfəʊn]
Semantic confusions	7 (2.9)	1 (0.7)	*bench* produced as *couch* [kaʊtʃ] *those* produced as *these* [ðiːz]

phonemes and similar but not identical phonemes in L1 on the output phonemic map. Examples of such confusions included an indistinguishable English phoneme [z] (as in *zebra*) with a Chinese phoneme [ʐ] (as *r* in *ri4*, "sun"), and English phonemes [ɔ:] (as in *born*) or [ɒ] (as in *pot*) with [o] (as *o* in *wo3*, "I") in Chinese. Other times our late L2 learning model could not distinguish between different L2-unique phonemes that are themselves similar on the output phonemic map. For example, English phonemes [ɔ:] as in *born* and [ɒ] as in *pot* are two phonemes not found in Chinese and therefore, they were often confused with each other on the output phonemic map when English was learned as L2. Other examples included [z] as in *zebra* with [ð] as in *then* (see Cutler et al., 2004 for phoneme confusion patterns of non-native listeners of English). Similarly, error patterns of this type of phonemic confusion were observed when Chinese was learned as L2 in our simulations. For instance, *c* ([ts']) and *ch* [tʂ'] are two phonemes not found in English and therefore they were often confused with each other on the phonemic map. Other examples included confusion of phonemes such as *j, q, x* ([tɕ], [tɕ'], [ɕ]), *z* and *zh* ([ts], [tʂ]), *s* and *sh* ([s], [ʂ]). Such phonemic confusions were very rare under the early L2 learning condition though (0.4 out of 74.2 total errors), given that the output phonemic map of early L2 learning was often much clearer and more organized (see the discussion section). These simulated patterns match up well with empirical findings (Yang and Yu, 2019; Wang and Chen, 2020; see further error analyses of an SLA corpus in next subsection). They are also consistent with speech learning theories indicating that early learners can create new phonetic categories more easily than late learners, and that such differences are due to the stabilization of the phonetic representation of L1 versus L2 over the lifespan of learning (see Flege, 1995; Flege, 2007). Flege's Speech Learning Theory also suggests that phonemes in the L2 that are similar to the L1 can actually cause more difficulty in learning, which is supported by our phonemic confusion data.

As shown on the row labeled as "phoneme replacement", some of the production errors were due to the phonemes in the pronouncing sequence replaced by other phonemes. For example, *were* [wɜ:] was wrongly pronounced as [wa:] in one simulation with English as L2. Similar examples included [ʃea] for *share* [ʃeə], [skeɪtʰ] for *skate* [skeɪt], [pʊi:] for *poor* [pʊə], and so on. Incorrect sequencing of phonemes in production was another salient error pattern that could be found in our simulations. For example, *grandpa* ['ɡɹæmpɑː] was wrongly pronounced as ['ɡæɹmpɑː] in one simulation with English as L2. Similarly, *telephone* ['tɛləfəʊn] was wrongly produced as ['tɛəlfəʊn] in another simulation. These types of patterns could also be found in the empirical error analyses of a Second Language Acquisition (SLA) corpus as shown in next section.

It is worth noting that incorrect sequencing could be accompanied with phoneme replacement and phoneme confusions, and it happened more often with words of greater length (such as multisyllabic words). For example, in one of the simulations, *closet* ['klɑːzət] was pronounced as ['klɑːəzd]. The effect of word length on lexical production was reported in our previous monolingual simulations and could be associated with the impact of individual working memory capacities on word articulation (see Li et al., 2007). This type of error reflects the challenges that language learners face when they need fine control of their articulatory organs to coordinate and execute sequential sound patterns, especially when the sequence belongs to an unfamiliar language (see the *sensorimotor integration hypothesis* from Hernandez and Li, 2007). Both late and early L2 learning showed a large amount of incorrect sequencing (78 and 60 respectively) along with phoneme replacement errors, but for the early L2 learning, such errors occurred less frequently and were more evenly distributed between L1 and L2.

Finally, our simulations also showed production errors caused by the confusions at the semantic level. Such errors were most salient in the late L2 learning condition and could be associated with the fuzzy L2 representations on the semantic map. For example, in one simulation, the word *bench* was wrongly named as *couch* [kaʊtʃ] and the word *those* was produced as *these* [ðiːz]. A close examination of the trained semantic map showed that the conceptual representation of bench and couch activated the same BMU on the map; therefore, they were confused in the first place during lexical production. Given that the late L2 learning condition produced much denser and fuzzier L2 semantic representation than the early L2 condition, it carried more semantic confusions as such (see discussion below on displaying and measuring density in representation). Such a type of semantic confusion could also be found in empirical studies. For example, contributors of Swan and Smith (1987), as quoted in Swan (1997), collected many vocabulary confusions at the semantic level (e.g., *think/hope, beat/hit/strike/knock*) in second language learners' L2 (English) production. An example observed from Chinese L2 learners of English, as quoted in Swan (1997), was that they often had confusions on the usage of

TABLE 4 | Total and average number of production errors shown in second language learners' L2 (Chinese). Data is averaged over 46 participants (with English as L1) from the COPA Corpus on TalkBank (Zhang, 2009a).

	Total errors	Average (SD)	Error examples
Total	2,177	47.33 (71.26)	
Phonological errors [*p]	1,549	33.67 (67.43)	**Tone confusions:** (*shi3* for 是 *shi4*); **Phoneme confusions/replacements:** (use *s[s]* for *sh[ʂ]* in 说 *shuo1,* use *c [ts'] for ch [tʂ'] in* 车 *che1, use j[tɕ] for q[tɕ'] in* 年轻 *nian2qing1, use an[an] for ang [ɑŋ] in* 光 *guang1*); **Incorrect sequencing** (*tuo1[t'uo] for* 偷*tou1 [t'əu], ling3dai4* for 带领*dai4ling3*, *xian1xin1* for 新鲜 *xin1xian1*)
Semantic errors [*s]	581	12.63 (7.83)	**Nouns:** 公司*company* -贸易 *business;* 工具 *tool* -玩具 *toy;*语文*literature* - 语言 *language;* 家 *home* - 房子*house;* 早上 *morning* - 晚上 *evening;* **Verbs**:做*do* - 踢*kick;* 去*go* - 回*back;* 穿 *wear (cloth)* -戴 *wear (hat);* **Pronouns:** 他*he* - 他们*they;* 你们*you guys* - 他们*they;* 那*that* - 这 *this*; **Classifiers:** 个*general* - 辆 *for vehicle*; 只 *for small animals* - 匹 *for horse*
Neologisms [*n]	42	0.92 (1.07)	经验纸*experience paper* - 简历 *resume;* 警察站 *police stand* - 警察局 *police station;* 冰盒子 *ice box* - 冰箱 *refrigerator* 服务人 *serving man* - 服务员 *server*

"small verbs" such as *come/go, do/make, bring/take*, which are semantically similar but with subtle differences. Similar examples such as 爱*love*/喜欢*like* and 预约*appointment*/约会*date* could also be found in American L2 learners of Chinese (Yuan, 2017).

Production Errors in SLA Corpus

The production errors shown in our model could also be compared with those from an SLA corpus (COPA Corpus on TalkBank, Zhang, 2009a), which included the elicited responses from L2 learners of Chinese to a fixed series of questions designed to measure the growth of their proficiency in Chinese (Zhang, 2009b). Specifically, we analyzed data from 46 learners of Chinese with English as their L1 in the corpus, aged between 19 and 56 years (M = 27.79, SD = 9.96). Although AoA information was not indicated in the corpus and despite many participants could speak Chinese relatively fluently, these participants were clearly not balanced bilingual or early L2 learners of Chinese. In the COPA Corpus, each participant's production errors at the word level were clearly annotated with error codes of the CHILDES (Child Language Data Exchange System; MacWhinney, 2000), and we conducted a frequency analysis of these errors using the CLAN program of the CHILDES.

As shown in **Table 4**, three main types of production errors were annotated in COPA Corpus by Zhang (2009). They were phonological errors (coded with [*p]), semantic errors (coded with [*s]), and neologisms (coded with [*n]). Generally consistent with our simulation results from late L2 learning condition (see the sections above), there were overall more phonological errors than the semantic errors [t (45) = 2.21, p = 0.032 in a paired-samples t test]. Within the phonology category, many errors were caused by the confusions of tones, a salient feature of Chinese as a tonal language but a difficult linguistic feature for learners of L2 Chinese (Hao, 2012; see also; Pelzl et al., 2021). We could also observe many phonemic confusions/replacements as in previous simulations. Examples included, but not limited to, *s*[s] for *sh*[ʂ], *c*[ts'] for *ch* [tʂ'], *j*[tɕ] for *q*[tɕ'], *an*[an] for *ang* [ɑŋ]. Incorrect sequencing could also be observed in the corpus. For example, a participant mistakenly pronounced the word *xin1xian1* (新鲜 *fresh*) as *xian1xin1*.[4]

Similar to our simulations in the late L2 learning, semantic errors in the COPA Corpus included confusion of a word with another word more or less semantically related to it, such as 那 *that*-这 *this,* 穿 *wear (cloth)*-戴 *wear (hat)*, 工具 *tool*-玩具 *toy,* and 只 *classifier for small animals*-匹 *classifier for horse.* Understandably, the confused words were often in the same grammatical category. As discussed below, we believe that this reflects the fuzzy and dense representations within the semantic categories of late L2 learners. There were also a small number of neologisms in the COPA Corpus, which were participant-generated pseudo words based on meaning of the target words. However, phonological information from the target word might also be mixed in the neologism, given that participants might keep the most part of a compound word in its original form but only changed one morpheme/character into a similar but inappropriate morpheme, like the case in using 警察站 *(police stand)* for 警察局 *(police station)*. Neologisms were not able to be simulated in our current study given that DevLex-II does not have a separate morpheme layer, although this could be one of the future research directions.

Displaying and Measuring Density in Representation

As discussed previously, the topography-preserving property of SOM allows the researchers to visually examine the emergence of lexical-semantic structures in their models. Here we show how lexical items from the two languages are represented under the late L2 learning (**Figure 2**) and early L2 learning (**Figure 3**) conditions. **Figure 2A** shows a semantic map[5] at the final stage of a late L2 learning condition. The circled gray regions indicate the nodes that represent the L2 (English) words and the other regions represent the L1 (Chinese) words learned by the model. Inspecting the bilingual representations, we can find that the words are not evenly represented for L1 and L2 by the model. Some areas are very dense while others very sparse. Lexical items in L2 are represented more densely on the maps in a more disorganized fashion, and they have higher chances to be confused with each other (being projected to the same BMU).

[4]Please note that this example may also be categorized as semantic errors since the two characters (*xin1*新and *xian1* 鲜) of the word are also morphemes by themselves with the meaning of *fresh*

[5]Phonological maps are not displayed here, but their representations are similar

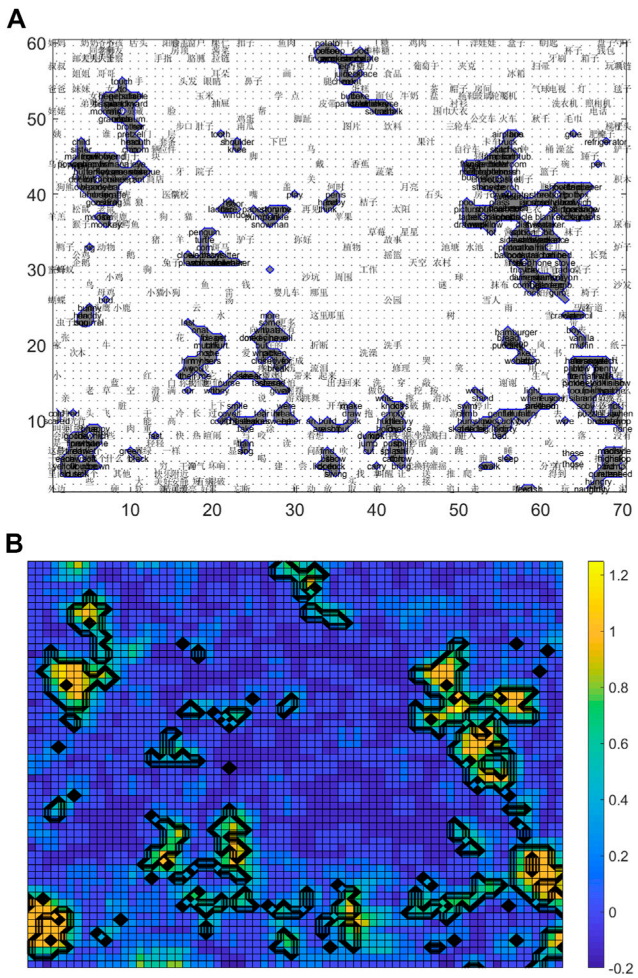


FIGURE 2 | Bilingual lexical-semantic representations of late L2 learning in the DevLex-II model. Circled areas correspond to L2 (English) words. Top: **(A)** Semantic map; Bottom **(B)** Heat map with density measure of each node, brighter color indicates higher density level.

Compared with L1 words, L2 words occupy only small and fragmented regions (neighborhoods on the map), dispersing throughout the map. In addition, the boundary between L1 items and L2 items on the maps is fuzzy.

To explore the differences in the density of the regions that L1 and L2 occupied on the map, we developed a method to calculate the density of units in their semantic and phonological neighborhoods. Specifically, we defined a unit's *density* as the number of words represented as BMUs in its direct neighborhood divided by the total number of units of its neighborhood, which is usually nine, but can be six or four, depending on whether the tested unit is on the border or in the corner of the map. The value of this density measure ranges from 0 when a unit has no words represented as BMUs in its neighborhood, to 1.0 when its entire neighborhood including itself are occupied by words. The results of this calculation showed that under the late L2 learning conditions the density of the L2 regions reached a very high level (0.64 and 0.75 on average for the semantic maps and phonological maps, respectively). **Figure 2B** shows a heat map of **Figure 2A** with each unit's density level represented by color. The high density of the small and isolated regions occupied by L2 can be clearly observed with the bright colors, reflecting the compact and fuzzy representation of L2 items.

Figure 3 presents a semantic map and its corresponding heatmap under the early L2 learning condition. Comparing it with **Figure 2**, we can find that the relative onset time of L2 vs. L1 plays an important role in modulating the overall representational structure of L2. For the early L2 learning condition, our network shows clear distinct lexical representations of the two languages at both the semantic (**Figure 3A**) and phonological level (not shown here). The results imply that the early learning of two languages allows the system to easily separate the lexicons during learning. In addition, as shown by the heatmap in **Figure 3B**, L2 representations are less crowded with lower density measures (bluer in color) on the L2-occupied regions on both phonological and semantic maps (0.35 and 0.29 respectively).

To further explore the density distributions of L2 and L1 on the maps under different learning conditions, a simulation-based 2 by 2 mixed-design ANOVA was conducted with learning condition (early vs. late) as the between-subject factor and language (L2 vs. L1) as the within-subject factor. The data were based on 10 simulations (5 for early and 5 for late learning), and only results for semantic maps are reported here. Significant main effects were found for both factors. The main effect of learning condition [$F\ (1,8) = 369.51$, $p < 0.001$, partial $\eta^2 = 0.979$] suggested that overall late L2 learning generated more crowded representation than early L2 learning. The main effect of language was significant too [$F\ (1,8) = 900.20$, $p < 0.001$, partial $\eta^2 = 0.991$]; and it showed that L2 had more crowded representation than L1.

A significant interaction between learning condition and languages can also be observed in our simulations [$F\ (1,8) = 179.70$, $p < 0.001$, partial $\eta^2 = 0.957$]. The interaction can be clearly found on the line graph of the average density levels of the four groups (**Figure 4**). Specifically, a post-hoc test on the simple effects of learning condition showed a significantly ($p < 0.001$) more crowded L2 representation under the late L2 learning condition ($M = 0.64$, $SD = 0.025$) than under the early L2 learning condition ($M = 0.29$, $SD = 0.043$); however, there is no significant difference ($p = 0.999$) between L1 density levels under the two learning conditions (L1 representations are always clear).

GENERAL DISCUSSIONS

The comprehension and production errors as shown in our bilingual learning models reflect the dynamic interactions among bilinguals' two lexicons in the DevLex-II model. We suggest that these errors mainly come from two sources, namely the fuzzy linguistic representations of L2 on each map and the inaccurate/unstable connections between the maps. Such considerations based on computational models are informative to

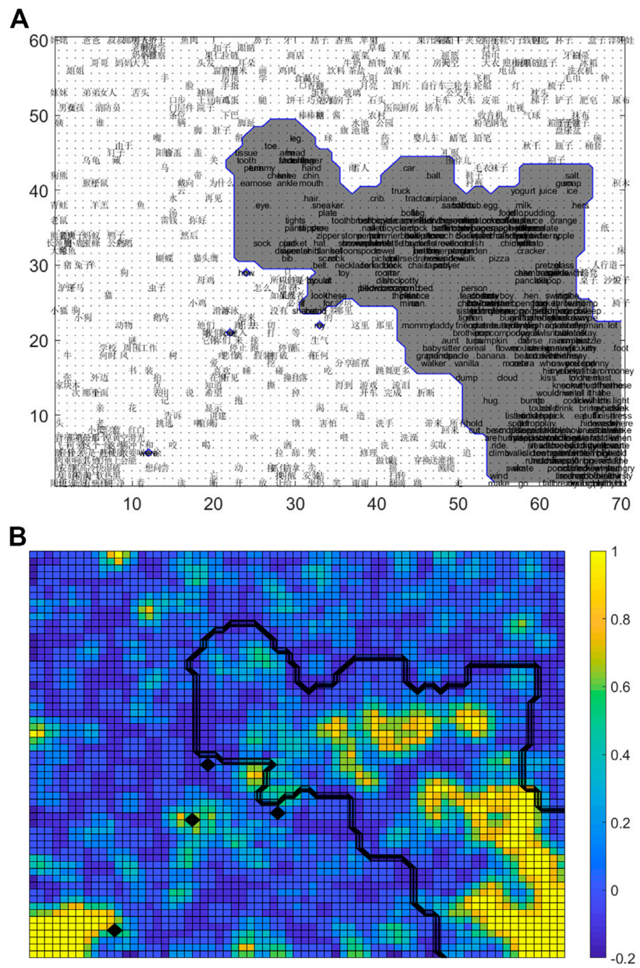


FIGURE 3 | Bilingual lexical-semantic representations of early L2 learning in the DevLex-II model. Circled areas correspond to L2 (English) words. Top: **(A)** Semantic map; Bottom **(B)** Heat map with density measure of each node, brighter color indicates higher density level.

our understanding of the nature of fuzzy representations in the bilingual lexicon.

Fuzzy L2 Representations

One possible source of these error-prone difficulties in production and comprehension could be the fuzzy L2 representations of the L2 phonology and semantics, and the output sequence maps. As can be found in **Figure 2** and the associated analyses of word density, there is a very compact and fuzzy representation of L2 items under the late L2 learning condition. There are two ways in which we characterize the fuzziness in the dense L2 areas: on the one hand, many L2 words are densely (and sometimes imprecisely) represented on the adjacent units and even by the same BMUs; on the other hand, the organization of L2 lexicon is fuzzy in the sense that the boundaries among word categories are blurred and overlapped, especially for those words that are similar phonologically or semantically. As a consequence, the retrieval of the sound or the semantic content of a word could be difficult because the competition between words is strong and could thus result in a higher confusion rate, contributing to the higher number of comprehension errors of L2. Word density is relatively low for the L1 words in general, and the organization of L1 representations is clearer and more precise. They are more robust than words in high density areas and thus more resistant to competition. Consequently, a clearer, more precise, and less crowded lexical-semantic representation of L2 makes early bilinguals less prone to lexical errors in their L2, compared with late L2 learners.

In late L2 learning, L2 representations are fuzzy because they are often parasitic on or auxiliary to those of L1 words, in the sense that the locations of many isolated L2 words depend on how similar they are to the L1 words in meaning (for the semantic map) and in sound (for the phonological map). For example, on the semantic map shown in **Figure 2A**, the English word *go* and *walk* is located next to the Chinese verbs like *pao3* ("run") and *pa2* ("crawl") since they are similar in meaning. Similar examples could also be found on one phonological map (not shown here): English word *cat* is close to the Chinese word *kai1* ("open") since they sound similar; other examples include *ear* close to *ye2ye* ("grandpa"), *hair* close to *hei1* ("black"). Interestingly, these are the places where few cross-language comprehension errors occurred at the beginning stages of late L2 learning. As shown in **Table 1**, the ratio of such cross-language comprehension errors over the total number of errors gradually decreases as learning progresses, indicating a gradually reduced parasitism of L2 on L1, even for the late L2 learning condition. Another interesting finding is that the cross-language interference is unidirectional, that is, the comprehensions of L2 words are affected by L1 knowledge only. There is little evidence of direct interference from L2 to L1 in our simulations. This is perhaps because in late L2 learning, the system has not reached a level of proficiency to produce backward L2-to-L1 influence. Such findings are consistent with the theoretical perspectives of emergentism, according to which parasitism arises not simply as a function of AoA, but as a function of the interaction between L1 and L2 (Li and Zhao, 2013; Hernandez et al., 2005). Specifically, the bilingual's L1 may be "entrenched", such that the lexical structure established early on becomes resistant to radical changes when L2 learning occurs late in life, causing only L2 to be parasitic on L1 rather than the other way around.

The compact and fuzzy L2 representations on the semantic map could also contribute to more production errors of L2. As shown in **Table 3**, semantic confusions (e.g. *bench-couch*) may generate production errors between synonyms. These two words are represented by the same BMU on the semantic map in one simulation with English as the L2, indicating that our model confuses these two L2 words in the first place. As a result, the model produces the sound of a wrong but semantically related word. A real-world example of this type of error would be that a participant mistakenly called a bench as couch in a picture naming task in her L2 (see similar examples in McMillen et al., 2020). Additionally, for those L2 words that are not overlapped with each other but are close enough at the semantic level, the high competition among them could result

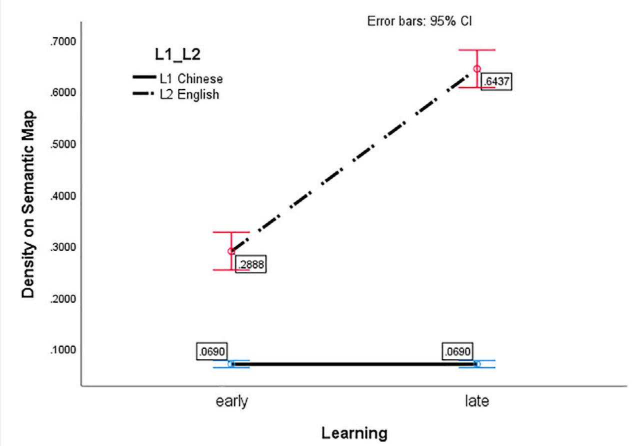


FIGURE 4 | A line graph of the average density levels of the L1 and L2 regions on the semantic map under the two learning conditions.

in shorter reaction times and less accuracy when being produced in picture naming tasks (Gollan et al., 2008), or result in higher rate of tip-of-the-tongue state in L2 (Kreiner and Degani, 2015). As discussed earlier, such production errors can be found in real SLA corpus as shown in **Table 4**.

On the output phoneme map, confusions between similar phonemes were also found, and these confusions contribute to many production errors as described above. Specifically, in late L2 learning, the subtle differences between some L2-unique phonemes are not highly distinguishable in a system that has already committed itself to the L1 phonemic inventory, thus these similar phonemes are projected to the same BMU on the map. Also, some of the new phonemes in L2 are conveniently "attracted" to similar but not identical L1 phonemes, acting like the "magnets" in the phonemic space for late L2 learners. Phonological errors shown in **Table 4** also provide many such examples with Chinese as L2. This pattern is highly consistent with the well-documented findings that adult L2 learners often have greater difficulties in accurately perceiving or producing the phonemic contrasts that do not exist in their L1 (e.g.,/r/and/l/ for Japanese learners of English, Flege, 1995; Flege, 2007; see also Zhang and Yin, 2009; Han, 2013 for detailed discussions on the commonly observed pronunciation problems of Chinese learners of English as L2).

Unstable Connections

Another major source of L2 lexical errors could be the weak or inaccurate form-to-meaning or meaning-to-sequence links between the maps, simulating the unstable connections between these different linguistic aspects in real language learning situations. The fuzzy L2 presentations again may contribute to some failures of building reliable between-map links. In the DevLex-II model, associative connections between maps are trained *via* the Hebbian learning rule, a biologically inspired mechanism, whose success requires a consistent co-occurrence of different linguistic aspects belonging to the same word. Under the late L2 learning condition, due to the fuzzy boundaries between L2 and L1 and within the compacted L2 region, BMUs corresponding to the same word may be subject to quick change of their coordinates on the maps. As a consequence, associative connections might be weak or inaccurate and cannot overcome the randomness in the model (which is generated by the connections' initial random weights). Indeed, Cook et al. (2016) showed that a fuzzy nonnative phono-lexical representation may lead to inaccurate form-to-meaning mappings in a Pseudo-Semantic Priming (PSP) task of L2 for American adult L2 learners of Russian. Many uncategorized comprehension errors occurred in late L2 learning (see **Table 1**) may be caused by these unreliable associative connections. It is worth noting that, under the early L2 learning condition, the boundaries between and within the two languages are much clearer on both phonology and semantic maps (**Figure 3**), thus effectively reducing the number of uncategorized (or arbitrary) comprehension errors.

Adult L2 learners often face a big challenge in adjusting themselves to better coordinate their articulatory apparatus to execute sequential sound patterns in an unfamiliar language. Different from comprehension, the articulation of sounds must be a sensorimotor process, and the accuracy of L2 pronunciation depends on the speaker's motor control and effortful coordination of the articulatory apparatus (tongue, lips, jaw, larynx, etc.). The difficulties in building correct meaning-to-phonemes links for L2 words in our late L2 learning condition reflect this challenge, especially for words with considerable length. The longer the target word is, the more ordered one-to-more links the model needs to learn, and the higher the chance of incorrect sequencing of phonemes and the replacement of certain ambiguous L2 phonemes (see also similar patterns in child L1 based on the DevLex model; Li et al., 2007).

The Dynamic Interplay Between L1 and L2

The current study clearly demonstrates that to understand effects of AoA on L2 acquisition, the computational modeling approach is important, especially with regard to understanding the dynamic interplay between L1 and L2, the process of L1 entrenchment, L2 parasitism, and the semantic, orthographic, and phonological organizations of lexical structures. Our simulations with the DexLex-II model show that such interactions play an important role in the development of the lexical representation systems across learning stages. When L2 is introduced late relative to L1 learning, L2 learners develop L2 representations parasitic on their L1 representations due to previously consolidated L1. Late L2 representations are fuzzy and under-differentiated in both phonological and semantic systems. Such fuzzy representations contribute to errors in L2 word comprehension and production as shown in our simulations. When L2 is introduced early relative to L1 learning, clear and distinct lexical representations of the two languages emerge in learning, and fewer errors are observed.

Why does late L2 learning lead to lexical representations so different from those in the early L2 learning condition? We believe that this "age" effect in L2 learning may reflect the changing learning dynamics and neural plasticity of the learning system.

In the late learning condition, L2 is introduced at a time when the learning system has already dedicated its resources and representational structure to L1, and L1 representations has been consolidated. So, L2 can only use existing structures and associative connections that are already established by the L1 lexicon. This is the sense in which we say that the L2 lexicon is parasitic to the L1 lexicon (Hernandez et al., 2005). In terms of the network's plasticity, the decrement of the neighborhood sizes on each map at a later stage of learning also significantly constrains its plasticity for radical re-organization. Therefore, as reflected at multiple levels of our model, L2 representations are constrained by well-developed L1 to fragmented areas. In contrast, for early L2 learning, the network still has significant plasticity and can continually reorganize the lexical space for L2. Rather than becoming parasitic to the L1 lexicon, early learning allows the L2 lexicon to present significant competition against the L1 lexicon. Our computational modeling findings suggest that the nature of bilingual representation is the result of a highly dynamic and competitive process in which early learning significantly constrains later development, shaping the time course and structure of later language systems.

Our simulation results are consistent with many previous theoretical frameworks that emphasize the dynamic interactive nature of bilingualism (Hernandez et al., 2005; Hernandez and Li, 2007; Li, 2015). Moreover, newer theoretical formulations highlight this dynamic interaction in terms of emergentism and the ecosystem (see Claussenius-Kalman et al., 2021), influenced by strong competitions between bilinguals' two languages across a developmental timeline. For late L2 learners, their L1 knowledge and skills are already well established, and highly resistant to change (i.e., "entrenched") in the face of new input from a new language. Once the structural consolidation in L1 has reached a point of entrenchment, the organization of L2 will have to tap into existing representational resources and structure of L1. According to the "sensorimotor integration hypothesis" (Hernandez and Li, 2007), entrenchment is accompanied by changes in neural plasticity, particularly in sensorimotor integration, such that highly flexible neural systems for developing fine-grained articulatory motor actions and for sequence processing are in deficit or are no longer available. Indeed, recent neurocognitive findings provide evidence that points to differences in neuroplasticity: L2 speakers fail to establish a neural network that connects L2 lexico-semantic representation with sensorimotor integration, in contrast to the L1 network that establishes strong connections between language processing areas and sensorimotor brain systems; see Figure 5 in Zhang et al. (2020). Such findings have significant theoretical and practical implications for L2 learning and representation, as discussed in Li and Jeong (2020) from a neurocognitive perspective.

Future Directions

DevLex-II has been proven to be a powerful tool for studying both monolingual and bilingual lexical development. In the future, we plan to further extend its scope to help us better understand L2 learning and representation.

First, DevLex-II is essentially a developmental model, but it could be extended to integrate both lexical learning and processing, and simulate a wide variety of empirical findings quantitatively. By incorporating a spreading activation mechanism on the semantic map, Zhao and Li (2013) has successfully simulated the effects of age of acquisition on cross-language semantic priming. Similar spreading activation mechanism, if added onto the phonological map and cross-map connections, will have the potential to simulate the phonology-based priming effects and word recognitions as shown in empirical studies such as (Cook et al., 2016; Gor, 2018).

Second, DevLex-II can be used to systemically examine how individual L2 learners' different cognitive abilities can affect their L2 learning outcomes, given other factors (such as AoA, L2 exposure) being equal. Cognitive scientists have been interested in whether executive control abilities such as working memory and processing speed might predict individual learners' success in L2 (see Miyake and Friedman, 1998; Kormos, 2015; Wen et al., 2017). In the original monolingual version of DevLex-II, we have successfully simulated individual differences in word production by adjusting our model's serial-recall ability in phonological short-term memory with a "memory gating parameter" (Li et al., 2007, p.593). Such a strategy could also be used for modeling bilingual processing and is consistent with the current trend of testing computational models on individual differences data (as shown in Peñaloza et al., 2019).

Finally, it will also be interesting to examine if fuzzy L2 representations may eventually be overcome in our model by integrating new modules. Zhang et al. (2020) showed that, compared with L2 speakers, L1 speakers engaged a more integrated brain network connecting key areas for language and sensorimotor integration during lexico-semantic processing. Naturally, a related question would be if late L2 learners can eventually acquire L1-like representations in L2 by utilizing extra cognitive resources or using new learning strategies. In a recent article, Li and Jeong, (2020) proposed the approach of *Social L2 Learning* (SL2) that focuses on grounding L2 learning in a social interaction framework, which focuses on "learning through real-life or simulated real-life environments where learners can interact with objects and people, perform actions, receive, use, and integrate perceptual, visuospatial, and other sensorimotor information, which enables learning and communication to become embodied." Most recently, Li and Lan (2021) also pointed out that *digital language learning* (DLL) may enable "L1-like representations in the L2, through the use of interactive and socially relevant contexts and multimodal/multisensory information", and such DLL approach may lead to brain changes in both function and structure. Social learning has been well accepted in L1 studies as an important contributor to successful language acquisition in children, and computational modeling research has also compared models with and without social cues. For example, Yu and Ballard (2007) incorporated social-interactive cues that are based on mother-child interactions, suggesting that such a model performed significantly better than models without such cues.

Along this new line of research that highlights social learning, the DevLex-II model could consider methods to build embodied semantic representations into the L2 lexicon by incorporating sensorimotor cues, social cues, and even affective-emotional cues from the learning environment. In addition, one could consider a growing SOM mechanism (e.g., Farkas and Li, 2002) to enable more resources for late L2 for the processing of embodied perceptual-spatial-sensorimotor features. Such studies could incorporate important information based on neurocognitive evidence that involves processing in both the neocortical and subcortical brain regions (see Green and Abutalebi, 2013; Stocco et al., 2014; see Grant et al., 2019, for a review). This new direction using the computational modeling approach, in conjunction with behavioral and neurocognitive studies, will lead to significant insights into the mechanisms and principles underlying individual difference in L2 learning and representation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

XZ and PL both contributed to the conception, design, and data analysis of the study. They contributed equally to the preparation of the manuscript.

ACKNOWLEDGMENTS

Preparation of this article was made possible in part by a grant from the Hong Kong Research Grants Council (GRF-15601520) and a Faculty Startup Fund from the Hong Kong Polytechnic University to PL, and a Faculty Development Grant from Emmanuel College to XZ. The authors wish to express gratitude to Jing Wang for help on an earlier draft of the article.

REFERENCES

Back, T. (1996). *Evolutionary Algorithms in Theory and Practice*. New York: Oxford University Press.

Bordag, D., Gor, K., and Opitz, A. (2021). Ontogenesis Model of the L2 Lexical Representation. *Bilingualism* 1, 1–17. doi:10.1017/S1366728921000250

Claussenius-Kalman, H., Hernandez, A., and Li, P. (2021). Expertise, Ecosystem, and Emergentism: Dynamic Developmental Bilingualism. *Brain Lang.* 222, 105013. doi:10.1016/j.bandl.2021.105013

Cook, S. V., and Gor, K. (2015). Lexical Access in L2: Representational Dificit or Processing Constraint *Ml* 10 (2), 247–270. doi:10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy Nonnative Phonolexical Representations Lead to Fuzzy Form-To-Meaning Mappings. *Front. Psychol.* 7, 1345. doi:10.3389/fpsyg.2016.01345

Cutler, A., Weber, A., Smits, R., and Cooper, N. (2004). Patterns of English Phoneme Confusions by Native and Non-native Listeners. *The J. Acoust. Soc. Am.* 116, 3668–3678. doi:10.1121/1.1810292

Dale, P. S., and Fenson, L. (1996). Lexical Development Norms for Young Children. *Behav. Res. Methods Instr. Comput.* 28, 125–127. doi:10.3758/bf03203646

Diependaele, K., Lemhöfer, K., and Brysbaert, M. (2013). The Word Frequency Effect in First- and Second-Language Word Recognition: a Lexical Entrenchment Account. *Q. J. Exp. Psychol.* 66 (5), 843–863. doi:10.1080/17470218.2012.720994

Dijkstra, T., and van Heuven, W. J. B. (1998). "The BIA Model and Bilingual Word Recognition," in *Localist Connectionist Approaches to Human Cognition*. Editors J. Grainger and A. M. Jacobs (Mahwah, NJ: Erlbaum).

Dijkstra, T., and van Heuven, W. J. B. (2002). The Architecture of the Bilingual Word Recognition System: From Identification to Decision. *Bilingualism* 5 (3), 175–197. doi:10.1017/s1366728902003012

Dijkstra, T., Haga, F., Bijsterveld, A., and Sprinkhuizen-kuyper, I. (2012). "Lexical Competition in Localist and Distributed Connectionist Models of L2 Acquisition," in *Memory, Language and Bilingualism: Theoretical and Applied Approaches*. Editors J. Altarriba and L. Isurin (Cambridge: Cambridge University Press), 48–73.

Dijkstra, T., Wahl, A., Buytenhuijs, F., Van Halem, N., Al-Jibouri, Z., De Korte, M., et al. (2019). Multilink: a Computational Model for Bilingual Word Recognition and Word Translation. *Bilingualism* 22 (4), 657–679. doi:10.1017/S1366728918000287

Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford, UK: Oxford University.

Elman, J. L. (1990). Finding Structure in Time. *Cogn. Sci.* 14, 179–211. doi:10.1207/s15516709cog1402_1

Farkas, I., and Li, P. (2002). "Modeling the Development of Lexicon with a Growing Self-Organizing Map," in Proceedings of the Sixth Joint Conference on Information Science, March 08–13, 2002. Editor H.J. Caulfield (North Carolina, USA: Research Triangle Park), 553–556.

Flege, J. E. (1995). "Second Language Speech Learning: Theory, Findings and Problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Editor W. Strange (Baltimore: York Press), 233–277.

Flege, J. E. (2007). Language Contact in Bilingualism: Phonetic System Interactions. *Lab. Phonol. 9*. Editor J. Cole and J. I. Hualde (Berlin, Germany: Walter de Gruyter), 353–382.

French, R. M. (1998). "A Simple Recurrent Network Model of Bilingual Memory," in Proceedings of the 20th Annual Conference of the Cognitive Science Society. Editors M. A. Gernsbacher and S. J. Derry (Mahwah, NJ: Erlbaum), 368–373.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., and Morris, S. K. (2005). Bilingualism Affects Picture Naming but Not Picture Classification. *Mem. Cogn.* 33 (7), 1220–1234. doi:10.3758/bf03193224

Gollan, T. H., Montoya, R. I., Cera, C., and Sandoval, T. C. (2008). More Use Almost Always Means a Smaller Frequency Effect: Aging, Bilingualism, and the Weaker Links Hypothesis☆. *J. Mem. Lang.* 58, 787–814. doi:10.1016/j.jml.2007.07.001

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., and Opitz, A. (2021). Fuzzy Lexical Representations in Adult Second Language Speakers. *Front. Psychol.* 12, 732030. doi:10.3389/fpsyg.2021.732030

Gor, K. (2015). "Phonology and Morphology in Lexical Processing," in *The Cambridge Handbook of Bilingual Processing*. Editor J. Schwieter (Cambridge: Cambridge University Press), 173–199. Cambridge Handbooks in Language and Linguistics.

Gor, K. (2018). Phonological Priming and the Role of Phonology in Nonnative Word Recognition. *Bilingualism* 21 (3), 437–442. doi:10.1017/s1366728918000056

Grant, A., Legault, J., and Li, P. (2019). "What Do Bilingual Models Tell Us about the Neurocognition of Multiple Languages," in *The Handbook of the Neuroscience of Multilingualism*. Editor J. Schwieter (Wiley-Blackwell), 48–74. doi:10.1002/9781119387725.ch3

Green, D. W., and Abutalebi, J. (2013). Language Control in Bilinguals: The Adaptive Control Hypothesis. *J. Cogn. Psychol.* 25 (5), 515–530. doi:10.1080/20445911.2013.796377

Grosjean, F., and Li, P. (2013). *The Psycholinguistics of Bilingualism*. New York, NY: John Wiley & Sons.

Grosjean, F. (1989). Neurolinguists, Beware! the Bilingual Is Not Two Monolinguals in One Person. *Brain Lang.* 36 (1), 3–15. doi:10.1016/0093-934x(89)90048-5

Grosjean, F. (2013). "Bilingualism: A Short Introduction," in *The Psycholinguistics of Bilingualism*. Editors F. Grosjean and P. Li (Hoboken, NJ: John Wiley & Sons), 5–25.

Han, F. (2013). Pronunciation Problems of Chinese Learners of English. *ORTESOL J.* 30, 26–30.

Hao, Y.-C. (2012). Second Language Acquisition of Mandarin Chinese Tones by Tonal and Non-tonal Language Speakers. *J. Phonetics* 40, 269–279. doi:10.1016/j.wocn.2011.11.001

Hernandez, A. E., and Li, P. (2007). Age of Acquisition: Its Neural and Computational Mechanisms. *Psychol. Bull.* 133, 638–650. doi:10.1037/0033-2909.133.4.638

Hernandez, A., Li, P., and MacWhinney, B. (2005). The Emergence of Competing Modules in Bilingualism. *Trends Cogn. Sci.* 9 (5), 220–225. doi:10.1016/j.tics.2005.03.003

Hochreiter, S., and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Comput.* 9 (8), 1735–1780. doi:10.1162/neco.1997.9.8.1735

Jiang, N. (2015). "Six Decades of Research on Lexical Representation and Processing in Bilinguals," in *The Cambridge Handbook Of Bilingual Processing*. Editor J. Schwieter (Cambridge: Cambridge University Press), 29–84. Cambridge Handbooks in Language and Linguistics. doi:10.1017/CBO9781107447257.002

Kohonen, T. (2001). *Self-organizing Maps*. 3rd ed.. Berlin: Springer.

Kormos, J. (2015). "Individual Differences in Second Language Speech Production," in *The Cambridge Handbook Of Bilingual Processing*. Editor J. Schwieter (Cambridge: Cambridge University Press), 369–388. Cambridge Handbooks in Language and Linguistics. doi:10.1017/CBO9781107447257.017

Kreiner, H., and Degani, T. (2015). Tip-of-the-tongue in a Second Language: The Effects of Brief First-Language Exposure and Long-Term Use. *Cognition* 137, 106–114. doi:10.1016/j.cognition.2014.12.011

Kroll, J. F., and Stewart, E. (1994). Category Interference in Translation and Picture Naming: Evidence for Asymmetric Connections between Bilingual Memory Representations. *J. Mem. Lang.* 33 (2), 149–174. doi:10.1006/jmla.1994.1008

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., and Zwitserlood, P. (2008). Native Language Influences on Word Recognition in a Second Language: A Megastudy. *J. Exp. Psychol. Learn. Mem. Cogn.* 34, 12–31. doi:10.1037/0278-7393.34.1.12

Li, P., and Farkas, I. (2002). "A Self-Organizing Connectionist Model of Bilingual Processing," in *Bilingual Sentence Processing*. Editors R. Heredia and J. Altarriba (North-Holland: Elsevier Science Publisher), 59–85. doi:10.1016/s0166-4115(02)80006-1

Li, P., and Grant, A. (2019). Scaling up: How Computational Models Can Propel Bilingualism Research Forward. *Bilingualism* 22, 682–684. doi:10.1017/s1366728918000755

Li, P., and Jeong, H. (2020). The Social Brain of Language: Grounding Second Language Learning in Social Interaction. *Npj Sci. Learn.* 5, 1–9. doi:10.1038/s41539-020-0068-7

Li, P., and Lan, Y.-J. (2021). Digital Language Learning (DLL): Insights from Behavior, Cognition, and the Brain. *Bilingualism*, 1–18. doi:10.1017/S1366728921000353

Li, P., and Zhao, X. (2017). "Computational Modeling," in *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*. Editors A. M. B. de Groot and P. Hagoort (Hoboken, NJ: John Wiley & Sons), 208–229.

Li, P., and Zhao, X. (2020). "Connectionism,". originally published 2012, last updated 2020 in *Oxford Bibliographies Online (Linguistics)*. Editor M. Aronoff (Oxford University Press). Available at: www.oxfordbibliographies.com.

Li, P., and Zhao, X. (2021). Computational Mechanisms of Development? Connectionism and Bilingual Lexical Representation. *Bilingualism*, 1–2. doi:10.1017/s1366728921000432

Li, P., Farkas, I., and MacWhinney, B. (2004). Early Lexical Development in a Self-Organizing Neural Network. *Neural Networks* 17, 1345–1362. doi:10.1016/j.neunet.2004.07.004

Li, P., Zhao, X., and Mac Whinney, B. (2007). Dynamic Self-Organization and Early Lexical Development in Children. *Cogn. Sci. A Multidisciplinary J.* 31, 581–612. doi:10.1080/15326900701399905

Li, P., and Zhao, X. (2013). "Connectionist Bilingual Representation," in *Foundations of Bilingual Memory*. Editors R. R. Heredia and J. Altaribba (New York, NY: Springer), 63–84.

Li, P. (2015). "Bilingualism as a Dynamic Process," in *Handbook of Language Emergence*. Editors B. MacWhinney and W. O'Grady (Malden, MA: John Wiley & Sons), 511–536. doi:10.1002/9781118346136.ch23

Li, R., Faroqi-Shah, Y., and Wang, M. (2019). A Comparison of Verb and Noun Retrieval in Mandarin-English Bilinguals with English-speaking Monolinguals. *Bilingualism* 22, 1005–1028. doi:10.1017/s1366728918000913

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Lawrence Erlbaum.

MacWhinney, B. (2013). "The Logic of the Unified Model," in *The Routledge Handbook of Second Language Acquisition*. Editors S. Gass and A. Mackey (New York: Routledge), 211–227.

McClelland, J. L., and Rumelhart, D. E. (1981). An Interactive Activation Model of Context Effects in Letter Perception: I. An Account of Basic Findings. *Psychol. Rev.* 88, 375–407. doi:10.1037/0033-295x.88.5.375

McClelland, J., Rumelhart, D., and the PDP Research Group (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2. Cambridge, MA: MIT Press.

McMillen, S., Griffin, Z. M., Peña, E. D., Bedore, L. M., and Oppenheim, G. M. (2020). "Did I Say Cherry ?" Error Patterns on a Blocked Cyclic Naming Task for Bilingual Children with and without Developmental Language Disorder. *J. Speech Lang. Hear. Res.* 63 (4), 1148–1164. doi:10.1044/2019_JSLHR-19-00041

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J. (2013). "Distributed Representations of Words and Phrases and Their Compositionality," in *Advances in Neural Information Processing Systems*, 3111–3119.

Miyake, A., and Friedman, N. P. (1998). "Individual Differences in Second Language Proficiency: Working Memory as Language Aptitude," in *Foreign Language Learning: Psycholinguistic Studies on Training and Retention*, 339–364.

Monner, D., Vatz, K., Morini, G., Hwang, S.-O., and DeKeyser, R. (2013). A Neural Network Model of the Effects of Entrenchment and Memory Development on Grammatical Gender Learning. *Bilingualism* 16 (2), 246–265. doi:10.1017/S1366728912000454

Pelzl, E., Lau, E. F., Guo, T., and DeKeyser, R. M. (2021). Advanced Second Language Learners of Mandarin Show Persistent Deficits for Lexical Tone Encoding in Picture-to-word Form Matching. *Front. Commun.* 6, 689423. doi:10.3389/fcomm.2021.689423

Peñaloza, C., Grasemann, U., Dekhtyar, M., Miikkulainen, R., and Kiran, S. (2019). BiLex: A Computational Approach to the Effects of Age of Acquisition and Language Exposure on Bilingual Lexical Access. *Brain Lang.* 195, 104643. doi:10.1016/j.bandl.2019.104643

Rumelhart, D., and McClelland, J. L. (1986). "On Learning the Past Tenses of English Verbs," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Editors J. L. McClelland and D. E. Rumelhart, and the PDP Research Group (MIT Press), Vol. 2, 216–271.

Shirai, Y. (2019). *Connectionism and Second Language Acquisition*. New York, NY: Routledge.

Stocco, A., Yamasaki, B., Natalenko, R., and Prat, C. S. (2014). Bilingual Brain Training: A Neurobiological Framework of How Bilingual Experience Improves Executive Function. *Int. J. Bilingual.* 18 (1), 67–92. doi:10.1177/1367006912456617

M. Swan and B. Smith (Editors) (1987). *Learner English: A Teacher's Guide to Interference and Other Problems, Cambridge Handbooks for Language Teachers* (Cambridge, UK: Cambridge University Press).

Swan, M. (1997). "The Influence of the Mother Tongue on Second Language Vocabulary Acquisition and Use," in *Vocabulary: Description, Acquisition and Pedagogy*. Editors N. Schmitt and M. McCarthy (Cambridge University Press), 156–180.

Tardif, T., Gelman, S. A., and Xu, F. (1999). Putting the "Noun Bias" in Context: a Comparison of English and Mandarin. *Child. Dev.* 70, 620–635. doi:10.1111/1467-8624.00045

Tsoukala, C., Frank, S. L., Van Den Bosch, A., Valdés Kroff, J., and Broersma, M. (2021). Modeling the Auxiliary Phrase Asymmetry in Code-Switched Spanish-English. *Bilingualism* 24, 271–280. doi:10.1017/S1366728920000449

Wang, X., and Chen, J. (2020). The Acquisition of Mandarin Consonants by English Learners: The Relationship between Perception and Production. *Languages* 5 (2), 20. doi:10.3390/languages5020020

Wang, H., Henderson, J., and Merlo, P. (2019). "Weakly-supervised Concept-Based Adversarial Learning for Cross-Lingual Word Embeddings," in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, Hong Kong, November 3–7, 2019 (EMNLP-IJCNLP), 4419–4430. doi:10.18653/v1/d19-1450

Wen, Z., Biedroń, A., and Skehan, P. (2017). Foreign Language Aptitude Theory: Yesterday, Today and Tomorrow. *Lang. Teach.* 50 (1), 1–31. doi:10.1017/S0261444816000276

Xu, Q., Markowska, M., Chodorow, M., and Li, P. (2021). Modeling Bilingual Lexical Processing through Code-Switching Speech: A Network Science Approach. *Front. Psychol.* 12, 662409. doi:10.3389/fpsyg.2021.662409

Yang, C., and Yu, A. (2019). The Acquisition of Mandarin Affricates by American L2 Learners. *Taiwan J. Linguist.* 17, 91–122. doi:10.6519/TJL.201907_17(2).0004

Yu, C., and Ballard, D. H. (2007). A Unified Model of Early Word Learning: Integrating Statistical and Social Cues. *Neurocomputing* 70 (13-15), 2149–2165. doi:10.1016/j.neucom.2006.01.034

Yuan, A. (2017). A Brief Analysis on the Common Mistakes Made by American Students in Learning Chinese as A Second Language. *Chin. Lang. Teach. Methodol. Technol.* 1 (2), 9–16. Article 3.

Zhang, F., and Yin, P. (2009). A Study of Pronunciation Problems of English Learners in China. *Asian Soc. Sci.* 5, 141–146. doi:10.5539/ass.v5n6p141

Zhang, X., Yang, J., Wang, R., and Li, P. (2020). A Neuroimaging Study of Semantic Representation in First and Second Languages. *Lang. Cogn. Neurosci.* 35, 1223–1238. doi:10.1080/23273798.2020.1738509

Zhang, Y. (2009a). *SLABank COPA Corpus*. [Data set]. TalkBank. Available at: https://sla.talkbank.org/TBB/slabank/Mandarin/COPA/engL1 (Retrieved November 9, 2021). doi:10.21415/T5ZK6W

Zhang, Y. (2009b). *A Tutor For Learning Chinese Sounds through Pinyin*. Pittsburgh, PA: Carnegie Mellon University. Unpublished doctoral dissertation.

Zhao, X., and Li, P. (2010). Bilingual Lexical Interactions in an Unsupervised Neural Network Model. *Int. J. Bilingual. Educ. Bilingual.* 13, 505–524. doi:10.1080/13670050.2010.488284

Zhao, X., and Li, P. (2013). Simulating Cross-Language Priming with a Dynamic Computational Model of the Lexicon. *Bilingualism* 16, 288–303. doi:10.1017/s1366728912000624

Zhao, X. (2017). "Connectionism in Linguistic Theory," in *The Oxford Research Encyclopedia of Linguistics*. Editor M. Aronoff (New York, NY: Oxford University Press). Available at: http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-69. doi:10.1093/acrefore/9780199384655.013.69

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 31 January 2022
doi: 10.3389/fpsyg.2021.768362

Check for updates

Employing General Linguistic Knowledge in Incidental Acquisition of Grammatical Properties of New L1 and L2 Lexical Representations: Toward Reducing Fuzziness in the Initial Ontogenetic Stage

Denisa Bordag[1,2] *and Andreas Opitz*[1]*

[1] *Herder Institute, Leipzig University, Leipzig, Germany,* [2] *University of Haifa, Haifa, Israel*

OPEN ACCESS

Edited by:
Jeanine Treffers-Daller,
University of Reading,
United Kingdom

Reviewed by:
Marianna Marcella Bolognesi,
University of Bologna, Italy
Irina Elgort,
Victoria University of Wellington,
New Zealand

****Correspondence:***
Andreas Opitz
andreas.opitz@uni-leipzig.de

Specialty section:
This article was submitted to
Language Sciences,
a section of the journal
Frontiers in Psychology

Received: *31 August 2021*
Accepted: *22 December 2021*
Published: *31 January 2022*

Citation:
Bordag D and Opitz A (2022)
Employing General Linguistic
Knowledge in Incidental Acquisition
of Grammatical Properties of New L1
and L2 Lexical Representations:
Toward Reducing Fuzziness
in the Initial Ontogenetic Stage.
Front. Psychol. 12:768362.
doi: 10.3389/fpsyg.2021.768362

The study explores the degree to which readers can use their previous linguistic knowledge, which goes beyond the immediate evidence in the input, to create mental representations of new words and how the employment of this knowledge may reduce the fuzziness of the new representations. Using self-paced reading, initial representations of novel identical forms with different grammatical functions were compared in native German speakers and advanced L2 German learners with L1 Czech. The results reveal that although both groups can employ general knowledge about German grammar when establishing new representations, the L1 native speakers outperform the L2 learners: Their new representations have more precise structure and are better differentiated from related representations with respect to their grammatical information. Modeling consequences of these findings are discussed in the context of the Ontogenesis Model of the L2 Lexical Representation and the Fuzzy Lexical Representation Hypothesis.

Keywords: mental lexicon, conversion, second language acquisition, fuzzy representation, incidental acquisition, word categories (parts-of-speech)

INTRODUCTION

In recent years, the properties of newly acquired lexical representations have gained more attention compared to those that are well established and frequently used. The focus has been primarily on the acquisition of meaning and its integration in the semantic network (for L1 e.g., Perfetti et al., 2005; Breitenstein et al., 2007; Clay et al., 2007; Mestres-Missé et al., 2007; Borovsky et al., 2010; Tamminen and Gaskell, 2013; for L2 e.g., Elgort, 2011; Bordag et al., 2015a, 2017a, 2018) and on the equivalent questions regarding word form (for L1 e.g., Shtyrov et al., 2010; for L2, e.g., Bordag et al., 2017b). However, there is basically no research exploring the initial representations of grammatical features in natural languages, in particular in the incidental acquisition scenarios when they need to be inferred from the linguistic context. Studies related to such topics usually have a different focus or background. For example, several studies investigate grammar acquisition in incidental learning, but their primary concern is not to explore the mental representations of the newly established grammatical features, but usually rather to assess learning gains under different reading conditions. Such studies address either effects of reading on overall grammar competence (i.e., not focusing on a particular grammar feature, e.g., Elley and Mangubhai, 1983) or a single grammatical feature, but are not concerned with its mental representation or how it interacts with previously acquired grammatical knowledge as is the case with our study. As an example, Aka (2020) explores the efficiency with which Japanese learners of English acquire to-infinitives used as nouns during reading while varying the amount of the target features in the input. Other authors, such as Shintani and Ellis (2010) or Song and Sardegna (2014) address incidental grammar acquisition of individual grammatical features (prepositions and -s plurals, respectively) during reading within a similar framework. Outside the area of research on reading, grammar acquisition is explored in studies with novice learners [e.g., noun-adjective agreement in Russian in Denhovska et al. (2016); plural -s and copula be in Shintani (2015)] and in artificial grammar learning (e.g., Grey et al., 2014; Monaghan et al., 2021; Rebuschat et al., 2021). In both cases, effects of previously acquired grammatical knowledge on the acquisition of features of the target grammar are not in focus and neither is the mental representation of the acquired features.

Similarly, the first versions of the most recent frameworks and approaches, such as the Ontogenesis Model of the L2 Lexical Representation (OM; Bordag et al., 2021a) and the related Fuzzy Lexical Representation (FLR) Hypothesis (Cook and Gor, 2015; Gor and Cook, 2020; Gor et al., 2021) that address the development of individual lexical representations and their quality only marginally touch upon the grammatical aspects. The OM addresses the development of lexical representations along three dimensions: the dimension of linguistic domains, the dimension of mappings between domains, and the dimension of networks of lexical representations. The dimension of linguistic domains that constitute a lexical entry has several sub-domains. The model focuses on the phonological, orthographic, and semantic domains as they comprise information which is stored at the lexical entry according to a general consensus. With respect to grammar, the situation is more complex, which is one of the reasons why it has not been addressed in the model blueprint. While some aspects of grammar such as agreement or word order are assumed to be handled on the processing level, other aspects, in particular morphosyntactic features, are assumed to be stored at the lexical entry. For the processing level of grammar, there is, to our knowledge, no model or approach that would operate with the concept of fuzziness. However, there are approaches that address topics that could be related to fuzziness in a broader sense such as the Shallow Structure Hypothesis (e.g., Clahsen and Felser, 2006, 2017). According to this hypothesis, L2 learners dispose of the same processing architecture and mental-processing mechanisms as L1 speakers, but they have "problems building or manipulating abstract syntactic representations in real time" (Clahsen and Felser, 2017, p. 2) and underuse syntactic information in online processing. Consequently, their processing could be seen as "fuzzier." However, it remains to be explored to which degree the concept of fuzziness would need to be adapted to suite also processing models in which fuzziness is seen more as a binary property.[1] With respect to the representational level that we address in our study, grammatical features at the lexical entry subsume both the so-called internal grammatical features with fixed values that need to be acquired (e.g., word class, grammatical gender, or declension class of nouns, number of singularia and pluralia tantum, subcategorisation frame, conjugational class or aspect of verbs, and declension class of adjectives), and the so-called external features with variable values that need to be set anew during processing each time (e.g., number, case, tense, grammatical voice, and gender of adjectives) (Bordag and Pechmann, 2009). For models supposing the existence of a so-called lemma as a component of a lexical entry (in addition to, e.g., a phonological form, earlier 'lexeme'), such as the Interactive Activation Model (Dell, 1986) or the Levelt Model (Levelt, 1989; Levelt et al., 1999), the lemma is supposed to be where such morphosyntactic features are stored [but cf., e.g., the Independent Network Model of Caramazza (1997) that dispenses with the notion of a lemma]. The part of grammar with a representational character and that is represented at the lexical entry would be a candidate for the grammar/morphosyntactic domain in the OM model.

The concept of fuzziness plays an important role in the OM model and is further developed in the FLR hypothesis (in particular Gor et al., 2021) that shares its focus on the quality of lexical encoding with the lexical quality hypothesis developed for L1 reading (Perfetti and Hart, 2002; Perfetti, 2007). Lexical representations undergo a developmental change during which the degree of fuzziness decreases untill a target stage is reached. This target stage of a lexical representation's ontogenesis, for which fuzziness is reduced to zero, is called 'optimum.' Fuzzy lexical representations are described in the FLR hypothesis as having imprecise, low-resolution or fuzzy encoding of their form and/or meaning, and potentially also the mapping between them. Their less distinct boundaries result in their reduced differentiation from neighboring representations.

The OM assumes that the development of a lexical representation can follow various scenarios depending on multiple factors such as the learning conditions or the current state of the learner's mental lexicon. As an example, the authors describe several possible developmental curves for the acquisition of the semantics of the word *dandelion* (for more details please cf. Bordag et al., 2021a, pp. 9–10 and Figure 5). In the simplest scenario, the word form is directly linked to an already existing semantic representation (possibly via the L1 form for novice learners, De Groot et al., 1994; see Bordag et al., 2017a for a detailed description), as it is typically the case when L1–L2 vocabulary pairs are learned, and the translation

[1] Clahsen and Felser (2017) explicitly classify their Shallow Structure Hypothesis as a multiple-pathways model that assumes at least two processing routes that operate in parallel. One of the routes involves creating detailed syntactic structures; the other one is syntactically shallower (i.e., characterized by deficiencies in hierarchical syntactic organization). Learners also have access to both paths in principle in their L2, but they have to rely much more on the shallower one.

equivalency is given. In this case, there is a sudden rise of the semantic ontogenetic curve toward the optimum. In more complex scenarios, the equivalency may need to be discovered in a cumulative way and initially only highly fuzzy semantic representations emerge that consist of, for instance, only very general features (e.g., 'a kind of blossoming flower'), or that comprise a specific but incomplete set of features. Over time, such representations may get more precise and semantically richer. This is typically the case when the meaning needs to be inferred in incidental vocabulary acquisition and depends on the input quality with respect to the available cues (Ellis and Collins, 2009). In such a case, fuzziness is reduced more gradually and the rise toward the optimum is less steep and/or may proceed in jumps.

Fuzziness is primarily viewed as a property of less familiar words (i.e., whose representations are lower on the ontogenetic curve that culminates in the optimum) in both L1 and L2. However, since less familiar words are more numerous in the L2 mental lexicon and because L2 learners experience more difficulties with encoding the phonological form and meaning of L2 words, establishing strong mappings between them, and integrating new L2 lexical entries in the lexical network, fuzziness is more pervasive in the L2 compared to L1[2]. Though the OM explicitly and the FLR hypothesis implicitly assume that fuzziness also affects grammatical encoding, the topic is not developed in either approach. As we will show in the current study, the basic concepts of these approaches such as fuzziness or optimum can also help to understand the characteristics and the initial development of new lexical representations at the grammatical level.

In our study, we ask to what degree readers or listeners can use their previous linguistic knowledge, which goes beyond the immediate evidence in the input, to create the mental representation of a new word and how the employment of this knowledge may reduce the fuzziness of the new representations. We were particularly interested in whether new mental representations are idiosyncratic in that they contain only grammatical information that could be derived directly from the linguistic context in which the new word appeared, or whether their establishment is assisted also by the information anchored in the reader's general knowledge about the grammar of the language, and if yes, how the engagement of this knowledge interacts with the fuzziness of the representation.

We found an empirical domain suitable for addressing this question in the area of German morphology. In German, every verb can be turned into a noun via a morphological process called conversion or zero-derivation. The crucial point of that process is that it operates without overt affixation (hence the name 'zero-derivation'): The product, the conversion noun, is formally identical to (some) morphological forms of the base verb. Thus, a German infinitive form like SPIELEN ('to play') can be converted into a form-identical conversion noun 'das SPIELEN' ('the playing'). This process is highly productive, and any German verb can be turned into an uncountable neuter (with respect to gender) noun this way. Though the mental representation of conversion nouns is still controversial, the most recent research supports the hypothesis that German deverbal conversion nouns are nested as word-category-specific subentries under a basic lexical entry that comprises also a subentry for a verbal representation. In their priming experiments with grammaticality judgments, Bordag and Opitz (2021) and Opitz and Bordag (2021) compare priming between formally identical primes and targets while manipulating the function of the primes, one of which being that of conversion [e.g., prime: *das – SPIELEN* ('the playing'), target: *wir – SPIELEN* ('we play')[3]]. The comparison of the priming effects in the different prime conditions (identical, inflected, infinitive, conversion noun, and inflected countable noun) allowed the authors to assess whether different, partially different, or the same representations were accessed. The priming patterns suggest the existence of complex lexical entries where the upper level is word-class neutral, and the lower levels (subentries) are specified for word classes and word-class-specific information (verbal sub-entry for verbal forms and nominal subentry for conversion nouns).

One of the questions we ask in our study is whether readers, who encounter a particular word form in a text such as an inflected verb form in 3rd person plural, establish only a simple lexical entry comprising only the given, i.e., verbal, information, or whether they can establish a more complex lexical entry also containing the conversion noun subentry based on linguistic generalization (for which they do not find any cue in the immediate linguistic context, though); and vice versa: Does the presentation of a new word as a conversion noun lead to establishing a simple nominal representation or does the new word's lexical entry also contain the verbal component?

Previous research on how existing general linguistic knowledge can affect representation of new linguistic information is scarce and emerges rather as a by-product in studies addressing other aspects of grammar acquisition. Bordag et al. (2015b) explored incidental acquisition of grammatical features of verbs during reading. They focused on subcategorisation and the (ir)regularity status of verbs. These two verb properties differ in that while a dominant, more frequent category can be determined for (ir)regularity status (namely the regular conjugation), no such generalization is possible for subcategorisation – a verb can be transitive or intransitive with basically the same chance. In their experiments, native and non-native speakers of German read short texts followed by several sentences that participants had to read in a self-paced manner. The introductory texts contained a conjugated novel verb repeated three times, whose meaning participants could derive from the context. The verb was then repeated in one of the self-paced sentences. In the congruent condition, the properties of the verb complied with its properties in the introductory text [e.g., the same subcategorisation frame or the same conjugation type (regular vs. irregular)]. In the incongruent condition, one of

[2]These difficulties are primarily accounted for by two factors: (1) Later learning onset in L2 that is associated with lower learning outcomes (e.g., DeKeyser, 2012; Hartshorne et al., 2018; Bylund et al., 2021) and (2) the L1 mental lexicon and phonological system already being in place when L2 lexical learning starts. L2 learners thus need to overcome the influence of L1 in developing the L2 system (e.g., Jarvis, 2000; Barrios and Hayes-Harb, 2020, 2021; Llompart, 2021).

[3]The two words were presented in two steps; participants made a grammaticality judgment over the whole phrase.

the two properties was violated, e.g., the verb was presented in a different subcategorisation frame than in the introductory text or it was presented as regular while conjugated as irregular in the introductory text or vice versa.

Bordag et al. (2015b) found that both native and advanced non-native readers could derive and store the information about the subcategorisation frame of a novel verb after just three occurrences in a text. However, contrary to the L2 learners, the L1 participants seemed indifferent to the (ir)regularity status of the novel verbs as it was presented in the introductory texts: No matter whether the verb was conjugated regularly or irregularly, the irregular conjugation was always perceived as a violation in the self-paced reading test phase. The authors interpret the finding through a "learning by unlearning effect": in their long experience with their native language, the L1 readers learned that regular conjugation is productive and that the set of irregular verbs is a rather small, closed group of verbs and they are certain to know all its members. Having learned this, they cease to acquire information about conjugation type from input and instead assume – based on their general knowledge about the language – that all new verbs are regular. If an unknown irregular form appears, they consider it implausible, irrespective of the evidence in the input, so the actual evidence in the particular context is overridden by the general knowledge. These findings indicate that – where applicable – L1 participants not only draw generalizations about linguistic properties and categories, but that these generalizations can also drive the acquisition and affect the setting of properties in newly established representations. On the other hand, L2 learners seem to be more driven by the actual input when acquiring linguistic properties of new words and less able to employ general knowledge about the language (cf. also the stronger focus on verbatim information in L2, e.g., Sampaio and Konopka, 2012; Bordag et al., 2021b; and the L2 form prominence, e.g., Jiang and Zhang, 2019). How the involvement of general knowledge may interact with the degree of fuzziness of the newly established representation, which is one of the aims of this study, has been neither directly explored nor actually addressed thus far.

THE PRESENT STUDY

In the present study we applied a method similar to Bordag et al. (2015b) to test whether participants can use general knowledge about the acquired language to establish complex lexical entries that contain information which goes beyond the immediate evidence in the text. More specifically, we asked whether an encounter with each of the forms (verb and conversion noun) triggers the establishment of a new lexical entry that also contains the representation of the other one, or whether for instance, only the encounter with the more basic form (from which the other form is derived, i.e., presumably the verb) enables the establishment of a complex representation containing also the specifications for the derived functions (i.e., the de-verbal, converted noun form).

In addition, we wanted to explore whether the employment of generalized grammatical information differs for native speakers and advanced learners of German and how this may be related to the higher degree of fuzziness observed for L2 representations compared to L1 so far, primarily in the domains of phonology and semantics (Bordag et al., 2021a). As previous research indicates, L2 learners might have a limited ability to engage this knowledge (as it is also typically at a lower level of acquisition compared to L1) and may thus be more dependent on the verbatim, word-form-related information in the input in general and when establishing new lexical entries in particular. To our knowledge, no previous research targeting a direct comparison between adult L1 and L2 acquisition in this area has been reported.

In the experiments in this study, participants read short German texts that contain a novel, previously unknown lexical item repeated twice. Each text is followed by a sentence read in a self-paced reading manner that includes the critical item either as an inflected verb form, an infinitive verb form, a conversion form, or a countable noun form. All forms shared the same stem and the ending –*en* that had a different function for each form. Thus, the target forms in the SPR sentences were all formally identical and differentiated only through the slightly different syntactic context of the sentence part which preceded them. This way we could compare reading times of form-identical words that differed only in their grammatical specifications (verbal forms, conversion nouns, and countable nouns) that were either present in the previous input (i.e., in the short texts), or not. In addition, using this version of the self-paced reading task enables to test the acquired knowledge for every single new word directly after it had occurred in input for the first time (compared to, for example, priming experiments).

The countable noun condition was included to serve as a kind of control condition. The countable noun (e.g., die MIETEN 'the rents') is homonymous with the other forms [e.g., MIETEN can also mean *they rent, we rent, to rent,* (*the*) *renting*], but its derivation is not a productive process in German: Not all German verbs have such derivations (their number is rather very limited) and neither their base forms (die Miete – 'the rent' in nominative singular), nor their meaning or grammatical gender are predictable from the verb stem from which they are derived historically. Previous research showed that homonymous countable nouns are represented as separate lexical entries (Bordag and Opitz, 2021; Opitz and Bordag, 2021). Therefore, if participants establish lexical entries that are precise and thus distinctly differentiable from other representations, they should not process the countable noun in the SPR sentence as the recently established verb/conversion noun representation, but rather respond to it as a new representation encountered for the first time (alternatively: respond to it as a violation). In this case, we should expect longer reading times for this control condition than for the other conditions. Contrary to the existence of the verb and the corresponding conversion noun that mutually condition themselves, the existence of the countable noun entry cannot be extrapolated from the more general, productive rules of the German language. Crucially, this control condition shared with the other experimental conditions the fact that the critical word was formally identical. Thus, any differences in reading times could not be caused by differences in form overlap.

Participants were tested in two experimental versions, A and B. The two versions differed in the function of the novel word which appeared in the short preceding texts: In version A, the new word was presented as a conjugated verb form; in version B it was a conversion noun form.

We hypothesized that if readers employ more general linguistic knowledge about the German language system when establishing a new representation, the resulting representations would be different to those if readers establish the representation relying solely on the information available in the immediate input. Since all German verbs can be converted into conversion nouns, readers could establish a complex lexical entry also containing the conversion noun grammatical information when encountering an inflected verb form or containing also the verbal grammatical information when encountering a conversion noun form based on their previous grammatical knowledge. In this case, we would expect the same reading times in both SPR verbal conditions (inflected and infinitive) and in the conversion condition, because in all cases participants would be accessing an already established entry containing full grammatical information (Hypothesis 1). However, if participants could not access more general linguistic knowledge ("to every verb there is a conversion noun" or "to every conversion noun, there is a verb"), they would only be able to establish a simpler entry containing only the grammatical information (verbal or conversion noun) that appeared in the text. In this case, we would expect longer reading times in the SPR condition that contains the form that did not appear in the initial text (Hypothesis 2). The longer reading times would either arise because that (part of the) representation could not be established yet based on the previous input and may become established only during reading of the SPR sentence in which it appears for the first time, or because this first-time-occurring form (in the SPR sentence) would be perceived as a violation because its word class is incongruent with the information readers induced and represented based on the previous text input (along the same argumentation as presented for the countable noun condition above).

By employing versions A (verbal form in text) and B (conversion noun in text), we want to explore whether readers' ability to use more general linguistic knowledge for acquisition and thus to establish complex lexical entries is dependent on or modulated by the grammatical type of input. We stipulated that participants will be either able to establish a more complex lexical entry comprising both the verbal and the conversion noun information irrespective of whether a verbal form or a conversion noun appears in the input (Hypothesis 1A), or that their ability to employ general grammatical knowledge will be limited or otherwise modulated when one of the forms (verbal or conversion noun) appears in the text input (Hypothesis 1B). For example, the fact that for every verb there is a conversion noun might be easier to generalize and employ in acquisition than that for every conversion noun there is a verb. These differences or asymmetries could be related to the fact that e.g., the higher frequency of the verbal forms compared to the conversion noun forms, conversion nouns are derived from verbs and thus more specific, or – in the case of the L2 learners – in language instruction the typical information shared in the classroom is that one can make a noun from every verb by using the neutral article *das* (formulation of a one-directional rule).

With respect to the differences between the two populations, we expect L1 speakers to be better at using their general linguistic knowledge for acquisition than the L2 learners (cf. also Bordag et al., 2015b) and that the L2 representations may manifest greater fuzziness than the L1 representations. However, since our L2 learners are very advanced, they might already possess the same abilities in this respect as the L1 speakers despite the explored linguistic phenomena not having equivalents in their native language. No similar homonymy of forms with corresponding functions exists in Czech, however, the concept of conversion is familiar to Czech native speakers as it exists, for example, between adjectives and nouns. The explored type of conversion in German is structurally very easy, completely regular, and very productive. It thus enables L2 learners to enlarge their competence significantly at very low costs. As such, conversion is typically learnt and mastered already rather early in L2 German, at the latest at the B1 level (at least for the Czech learners). Its formation in German is significantly easier than the formation of the Czech derived noun that corresponds in its function to the German conversion noun (in German: *sprechen – das Sprechen*, *mieten – das Mieten*, in Czech: *mluvit – mluvení*, *pronajmout – pronajmutí*). It can be thus safely expected that Czech learners at B2/C1 level are well familiar with the phenomenon.

Based on previous research, we also expect that critical effects may appear at the spill-over region in addition to the novel word itself. This is in line with Reichle et al. (2009) model of eye-movement control called "E-Z Reader 10," according to which processing difficulty can occur either at the lexical or post-lexical processing stage. The lexical processing stage comprises a word-familiarity check and lexical access, while the higher-order post-lexical processing involves the integration of the currently fixated word *n* "into the higher-level representations that readers construct online" (Reichle et al., 2009, p. 5). Given that the word form is the same in all our conditions, we can stipulate that it can pass the word-familiarity check without differences related to the different functions of the critical word. However, internal properties of the new representation are relevant for both lexical access and integration of the critical word into higher-level representations. Therefore, we also analyze the spill-over region, in which a word-class mismatch or grammatical properties mismatch (countable noun vs. non-countable conversion noun) between the novel item in the introduction text and the SPR sentence might play a stronger role due to difficulties in integrating a word with an unexpected word class or grammatical properties into the sentence context.

We first present the results of both experiments for the L1 and then for the L2 group. We decided on this order of presentation because our primary question is whether generalized linguistic knowledge is employed during establishment of new lexical entries. We assume that it is more likely to find evidence for it with adult native language speakers, which is why we address this group first. In the second step, we address the same question for advanced L2 learners to explore whether the L2 acquisition procedures work the same as in adult L1. In addition, we examine the patterns of results of both groups to explore whether there are indications of fuzziness in the initially established representations, which we expect especially in L2. Finally, we present an overall analysis of all four experiments that directly compares the L1 and the L2 data and confirms the patterns observed in the language separate analyses.

NATIVE PARTICIPANTS: EXPERIMENTS L1A AND L1B

In both experiments, participants read short texts in which a novel word (pseudoword) was introduced. After each text, participants read sentences in a self-paced reading manner. In some of the sentences the novel word appeared again, but partially in a different grammatical form.

Methods and procedures for both experiments were mostly identical, except for the grammatical form of the novel word introduced in the text. In Experiment L1A, the novel word was introduced as an inflected verb; in Experiment L1B, it was introduced as a conversion noun.

In the following we report all methods for Experiment L1A and L1B together, highlighting the aspects in which the experiments differed.

Methods

Participants

In Experiment L1A, 72 native speakers (56 female and 16 male) were tested with a mean age of 26.9 years ($sd = 7.90$, range = 18–56). Most participants were university students.

In Experiment L1B, a total of 70 native speakers (48 female and 22 male) were tested with a mean age of 28.9 years ($sd = 6.9$, range = 18–56). None participated in Experiment L1A. Most participants were university students.

Materials

Items

Twenty-four concrete German verbs with a very low frequency were selected that were mostly unknown to L2 learners at B2 to C1 level as assessed in a pre-test. These verbs were later replaced by pseudoverbs to guarantee that the critical words in the study were completely unknown to all participants (e.g., *gaffen* ‘to gawp’ was replaced by pseudoverb *brössen*). The pseudoverbs were constructed using the computer program Wuggy (Keuleers and Brysbaert, 2010) and followed German orthography and phonotactics (see Hulstijn, 1992). Care was taken that they did not resemble existing words in other languages, in particular in Czech and English. **Table 1** lists all novel verbs used in the experiment with their corresponding low-frequency counterparts.

TABLE 1 | List of items.

Low-frequency word	English translation	Novel word (pseudoword)
Schnitzen	‘To carve’	Fienen
Trödeln	‘To dawdle’	Belfen
Roden	‘To uproot’	Paufen
Gaffen	‘To gawk’	Brössen
Flanieren	‘To stroll’	Jollen
Flattern	‘To flutter’	Tinfen
Plaudern	‘To twaddle’	Zöcheln
Gröhlen	‘To bawl’	Jühnen
Hausieren	‘To peddle’	Rahnen
Kippeln	‘To tipple’	Döcheln
Lispel	‘To lisp’	Plimmen
Nisten	‘To nest’	Wucken
Gurgeln	‘To gurgle’	Zwaulen
Flunkern	‘To fib’	Meifen
Keimen	‘To germinate’	Hunken
Haaren	‘To shed (hair)’	Kleupen
Dösen	‘To doze’	Nieben
Schnurren	‘To purr’	Elmen
Modern	‘To molder’	Lörren
Schielen	‘To squint’	Gäpfen
Brodeln	‘To seethe’	Sülfen
Schlüpfen	‘To hatch’	Fähsen
Rascheln	‘To rustle’	Alzen
Schweißen	‘To weld’	Schünen

Texts

Twenty-four short texts were constructed in such a way that the meaning of the 24 verbs could be inferred from them. They comprised 3–5 sentences. The low-frequency verb itself was replaced by a pseudoword. Each pseudoword appeared in its corresponding text twice.

In Experiment L1A, the pseudoword appeared both times as an inflected verb form: Once in 3rd person singular in present tense (e.g., *er brösst*, meaning ‘he gawks’), and once inflected 3rd person plural (e.g., *viele Leute brössen*, meaning ‘many people gawk’).

In Experiment L1B, the pseudoword appeared as a nominalized form (a conversion noun) that was presented twice in the text, once with the article ‘das’ (e.g., *für das Brössen*, meaning ‘for the gawking’) and once in genitive with the article ‘des’ and genitive inflection on the noun (e.g., *wegen des Brössens*, meaning ‘because of the gawking’).

The final selection of the 24 texts was a result of a sequence of two pre-tests, in which all novel words were replaced by a dummy word *xarren*/*Xarren*. Participants were instructed to guess the meaning of the dummy word for each text and rate on a six-point scale how confident they were regarding their guess and how easy it was to deduce the meaning. Additionally, they rated the readability of each text and could leave additional comments regarding each text. In the first pre-test, 36 candidate texts were rated by native speakers ($N = 48$). The texts were presented in two versions, once with the dummy word in the function of an inflected verb (*xarren*), once in a function of a conversion noun (*Xarren*). For each participant, half of the texts appeared with the dummy word in one function and the second half in the other function. Before the second pre-test, the texts were optimized and submitted to another group of native speakers for rating. The 24 texts that scored best in the second pre-test were chosen as final text items for the experiment. The summary in

TABLE 2 | Properties of texts introducing the novel words.

	Text condition			
	Inflected verb form		Conversion noun form	
	Mean	(*SD*)	Mean	(*SD*)
Text length (in words)	62.8	(14.8)	64.3	(14.9)
Average sentence length (in words)	15.6	(4.39)	16.2	(4.43)
Readability	5.46	(0.76)	5.19	(0.98)
Ease of deducing the meaning	5.10	(1.26)	4.88	(1.22)
Confidence in deducing the meaning	5.15	(1.22)	4.85	(1.30)

Readability, ease of deducing the meaning and participants' confidence were measured on 6-point Likert scales (1–6).

Table 2 shows that the texts with a dummy word as an inflected verb and as a converted noun did not differ statistically with respect to their general readability, the ease of deducing the novel word's meaning, participants confidence in deducing the meaning, and text length.

In addition to the 24 texts, 6 filler texts were created that were similar to the critical texts but contained existing words only.

SPR Sentences

For each text, four critical sentences were created, each of them containing the novel word ending in *–en* (e.g., *BRÖSSEN*). However, in each of the sentences the novel word was used in a different function forming the four conditions of the experiment. In order to avoid orthographic cues (nouns are written with initial capitals in German), all SPR sentences were presented in capital letters (see examples below).

(1) Infinitive condition
The novel word is used as an infinitive verb form (e.g., *sie wollen brössen*, meaning 'they want to gawk').
Example:
VIELE LEUTE WOLLEN NUR BRÖSSEN, ANSTATT SELBST ETWAS ZU TUN.
"A lot of people just want to [gawk] instead of doing something themselves."

(2) Inflected condition
The novel verb was used as an inflected verb form ending in *–en* (i.e., in 3rd person plural, e.g., *sie brössen*, meaning 'they gawk').
Example:
VIELE LEUTE KOMMEN NUR UND BRÖSSEN, ANSTATT SELBST ETWAS ZU TUN.
"A lot of people just come and [gawk] instead of doing something themselves."

(3) Conversion noun condition
The novel verb was used as a conversion noun in nominative or accusative case, i.e., preceded by the definite article *das* and ending in *–en* [e.g., *durch das Brössen*, meaning 'due to (the) gawking'].
Example:
VIELE LEUTE KOMMEN NUR FÜR DAS BRÖSSEN, ANSTATT SELBST ETWAS ZU TUN.
"A lot of people come just for the [gawking] instead of doing anything themselves."

(4) Countable noun condition
The sentence contained a concrete, countable noun in plural that was formally homonymous with the novel verb as it appeared in the text, but there was no clear meaning relationship between them (e.g., *für die vielen Brössen*, meaning 'for the many/for all the ...'). The plurality of the context was unambiguously indicated by a preceding definite or indefinite numeral requiring a plural. Note that in contrast to this countable noun condition, all conversion nouns (as in condition 3 and as introduced in the texts in Experiment L1B/L2B) are singularia tantum (i.e., they do not have any plural form) by definition. Thus, the countable noun in plural here cannot be interpreted as a conversion noun.
Example:
DIE LEUTE KOMMEN NUR FÜR DIE VIELEN BRÖSSEN, ANSTATT SELBST ETWAS ZU TUN.
"A lot of people come just for all the/for the many ... instead of doing anything themselves."

As evident from the above examples, the parts of the sentences that followed the novel word were always identical in all four conditions and they were at least four words long. The part preceding the novel word that determined its word class and other grammatical properties could not be the same across all the conditions, but care was taken that there was as much overlap between the four conditions as possible.

In order to guarantee that the assumed differences in reading times are not due to reading differences that would be inherent to the four SPR sentences themselves, a pre-test was run that measured the reading times on the novel words and the words immediately following them within the sentences while no introductory texts were presented. Forty participants of the pre-test read all SPR sentences in all experimental conditions (i.e., with the novel word either as an inflected verb, an infinitive, a conversion form, or a countable noun) without any introductory texts. The participants were distributed over 4 lists such that each participant saw only one item in one of the four conditions, but each saw all four conditions equally often. None of the pre-test participants took part in the actual experiments. No differences in reading times were observed at the position of the critical word n (the novel word): $F(3{,}895.7) = 0.29$, $p = 0.834$; or the spill-over region, i.e., the following word, position $n + 1$: $F(3{,}899.2) = 1.47$, $p = 0.220$.[4]

The SPR sentences were related in topic to the previous text, but there was no vocabulary overlap between them and the texts except for the novel word. For each text, either none, one, or two filler SPR sentences were constructed that were also related by topic but consisted only of vocabulary typically known by the targeted learner group. The number of SPR sentences varied in order to avoid participants' strategies and/or expectations when

[4]Statistical analyses were conducted in parallel to those reported in detail for the experiment data (i.e., linear mixed effects models with log-transformed reaction times etc., for details see below).

the new word would appear in the SPR part of the experiment and also to deflect participants' attention away from the novel words.

For the comprehension task, a related sentence was created for each text that formed a statement that was either consistent with the meaning of the text or not. The statements referred to propositions of either the texts or the filler SPR sentences. However, they did not mention or refer to the novel word. The purpose of the task was to keep participants attentive to the texts.

Procedure

Prior to the experiment, participants were given written instructions, informing them that they were to read texts for comprehension and that comprehension statements would follow each text. The instructions also mentioned that the texts might contain unknown vocabulary from regional dialects or special registers, but that they were to try to grasp the text's meaning, nonetheless. The stimuli were presented using the E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA, United States).

Each trial consisted of three parts: reading of a text, reading of one to three SPR sentences, and assessing a comprehension statement. A trial started with the presentation of an introductory text that included the novel word or only known vocabulary (filler texts). Participants read the text silently and pressed the space bar when they were finished. When pressing the space bar, an SPR sentence written in capital letters appeared, initially with all words masked with Xs. When pressing the space bar again, the next word was revealed and the previous one was masked with Xs again (self-paced reading with a moving window, cf. Just et al., 1982). Reading times were measured. The number of SPR sentences following each text varied from one to three. One of the sentences was always the critical sentence in one of the experimental conditions.

After the presentation of the SPR sentence(s), a comprehension statement referring to the introductory text or one of the filler SPR sentences appeared on screen, and participants had to decide whether the statement was true or false by pressing one of the corresponding buttons. After the participant's response was registered, the next trial started with an inter-stimulus interval of 1,000 ms.

Items were distributed over four experimental lists and each subject was administered to one of those lists. Each list contained all 24 texts (and 6 filler texts), but for each text only one critical SPR sentence in one of the four conditions was presented. The number of conditions was counterbalanced across lists such that each participant saw six items in each condition and that four complementary experimental lists formed a complete set. Each participant thus read each text only once followed by one of the four possible SPR conditions. Within each list, the order of trials was pseudo-randomized for each participant with fixed positions of the filler trials and the restrictions that no more than three trials with the same answer to the comprehension statement and no more than two trials with the same experimental condition followed in succession. The first trial of the experiment was always a filler trial. One session of the experiment took about 35–40 min.

Data Preparation and Analyses

Statistical analyses for all experiments reported in the present paper were performed using linear mixed-effect models employing the software R (R Core Team, 2020). Models were fitted using the *mixed* function of package *afex* (Singmann et al., 2021). All models included random intercepts for participants and items. For all analyses, the maximal model structure was attempted (Barr et al., 2013). However, when the maximal model did not converge, the error term structure was systematically reduced using the procedure suggested by Singmann (2021). The structure of the final model is noted in the results for each analysis. For *post hoc* comparisons of contrasts of significant main effects and interactions, contrasts of estimated (marginal) means were performed using the package *emmeans* (Lenth et al., 2021) and the Bonferroni adjustment for multiple comparisons was applied for those contrasts. For the treatment of outliers, reaction time data were first winsorized with a 5% criterion, i.e., with the 2.5 and 97.5 percentile as boundaries, meaning that for each participant, all data points that fell below the 2.5th percentile or above the 97.5th percentile were set to these boundary values[5]. Additionally, and in order to compensate for non-normality of the distribution, all reaction times (in ms) were log-transformed (natural log) prior to statistical analyses. The same procedures were carried out for each of the reported experiments.

Results

Reaction times were analyzed on the positions n (the novel word) and $n + 1$ (the word following the novel word; spill-over region)[6] for the four conditions (inflected verb form, infinitive verb form, conversion noun and countable noun).

Table 3 and **Figure 1** summarize the results of mean response latencies of Experiment L1A (inflected form introduced in the texts) and Experiment L1B (conversion noun introduced in texts) L1B.

The two L1 experiments (L1A and L1B) were analyzed together. The analysis of latencies therefore contained fixed effects for the factors Condition, Position (n vs. $n + 1$), and Textform (inflected form vs. conversion form introduced in the texts, i.e., experiment L1A vs. L1B). The results of the final model [log(RT) $\sim$ Condition * Position * Textform + (Position + Textform | Item) + (Position | Participant)] are summarized in **Table 4**. They reveal a main effect of Condition [$F(3{,}6453.4) = 44.10$, $p < 0.001$] and a significant interaction of Condition and Textform [$F(3{,}6453.4) = 4.05$, $p = 0.007$]. Importantly, there was also a significant higher-level 3-way interaction of Condition:Position:Textform [$F(3{,}6452.0) = 2.83$, $p = 0.037$] indicating that the effect of Condition was moderated by an interaction of both Position and Textform. Following this significant 3-way interaction, *post hoc* comparisons of estimated (marginal) means were computed with p-adjustment for the accumulated alpha error according

[5]This follows recommendations of data treatment of L2 reaction times (Nicklin and Plonsky, 2020).

[6]Exploratively, we also looked at later spill-over regions, i.e., positions $n + 2$ and $n + 3$. However, at these positions the effects observed on n and $n + 1$ were already receding and mostly not significant. We thus analyzed and report only the n and $n + 1$ positions.

TABLE 3 | Results of Experiments L1A and L1B (mean RTs in ms and SDs).

	L1A (inflected form in texts)				L1B (conversion noun in text)			
	Position *n*		Position *n* + 1		Position *n*		Position *n* + 1	
Condition	**RT**	**(*SD*)**	**RT**	**(*SD*)**	**RT**	**(*SD*)**	**RT**	**(*SD*)**
Inflected	516.8	(267.9)	487.4	(181.1)	498.0	(222.1)	495.1	(159.5)
Infinitive	508.1	(269.9)	477.7	(168.2)	494.3	(224.4)	495.8	(178.1)
Conversion	510.9	(283.8)	481.0	(165.5)	495.6	(233.3)	455.2	(138.0)
Noun	635.8	(476.1)	559.6	(271.4)	546.8	(308.3)	538.4	(232.7)

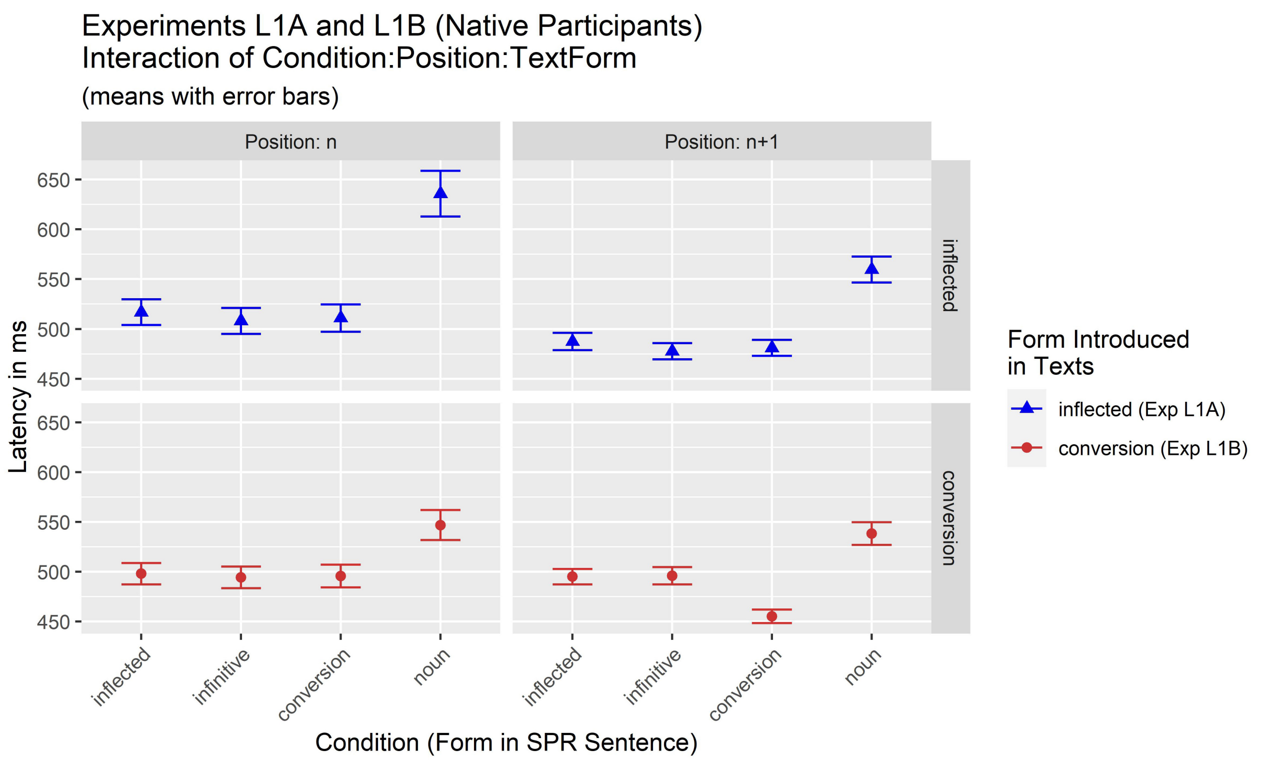


FIGURE 1 | Results of Experiments L1A and L1B: native participants. (Mean latencies of critical regions in SPR sentences).

to the Bonferroni procedure to investigate potential differences between conditions in different combinations of Position and Textform. Results (see **Table 5**) indicate that when the novel word was introduced as an inflected form (Experiment L1A), the pattern of results was essentially the same for positions n and $n + 1$: Responses to three of the four conditions were equally fast (i.e., the inflected, infinitive, and conversion condition; all $p > 0.999$), while responses for the countable noun condition were significantly slower (all $p < 0.001$). In contrast, when the novel word was introduced as a conversion form (Experiment L1B), the pattern of significant differences differed for positions n and $n + 1$. At position n, the results resembled the pattern also seen in Experiment L1A: there were slower responses to countable nouns compared to all other conditions. However, the effect was not so pronounced, as can be seen from the p-values that reveal the significance of the difference between the countable noun condition and the infinitive ($p = 0.010$) and the conversion ($p = 0.039$) condition, while there was only a marginal difference to the inflected condition ($p = 0.081$). This reduced difference is also visible in the numerical differences at position n (see also **Figure 1**): While in Experiment L1A there was a numerical difference of ca. 123 ms between the three faster and the slowest noun conditions, this was reduced to ca. 51 ms in Experiment L1B.

However, at the spill-over region ($n + 1$) a pattern emerged that is different to the so-far generally attested pattern of slower responses to countable nouns compared to (equally) faster responses to the three other conditions. While the noun condition still yielded the slowest responses, the conversion noun here elicited the fastest responses, also differing significantly from both the inflected ($p < 0.001$) and the infinitive ($p = 0.004$) condition.

TABLE 4 | Mixed model ANOVA table for Experiments L1A and L1B (native participants).

Effect	df	*F*	*p*-value	Signif.
Condition	3, 6453.38	44.10	<0.001	***
Position	1, 60.73	0.20	0.658	
TextForm	1, 140.25	0.19	0.665	
Condition:Position	3, 6457.96	1.19	0.314	
Condition:TextForm	3, 6453.38	4.05	0.007	**
Position:TextForm	1, 140.00	1.27	0.261	
Condition:Position:TextForm	3, 6451.96	2.83	0.037	*

*Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10.*

To sum up, the difference between the faster conditions and the slowest (i.e., countable noun) condition was more pronounced in Experiment L1A when the novel word was introduced as an inflected word compared to Experiment L1B when the novel word was introduced as a conversion noun. In addition, an effect of faster responses to conversion nouns was observed when the novel word was introduced as a conversion noun in the text (Experiment L1B), but only at position $n + 1$.

Discussion

With respect to the research question regarding the employment of general linguistic knowledge when establishing new lexical entries, we conclude that native speakers employ knowledge about grammar that goes beyond the information encoded in the immediate input when establishing mental representations of new words. This is indicated by the observation that participants showed no delays when presented with a form that was not in the preceding input, but whose existence could be inferred from the general knowledge about the German grammar: at position n inflected verb forms, infinitives, and conversion nouns were read equally fast regardless of the form presented in the preceding texts. At the same time, participants also showed sensitivity to the grammatical information in the input. It manifested itself with longer reading times in the countable noun condition. This form was preceded by a plural numeral in the SPR sentence, so that a noun in the plural could be predicted. However, when a word form appeared in the SPR sentence that was homonymous to the new word which participants had just acquired (via the preceding texts), but with the grammatical properties of a countable noun, participants had problems with lexical access and/or integrating this form in the sentence which resulted in the longer reading times. This also indicates that the recently established representation (based on the text input) was grammatically precise enough to be distinctly differentiated from another new representation with the same word form that participants encountered later (in the SPR sentence).

It is notable that at position n, the implausibility or surprizal effect was greater in Experiment L1A where inflected verb forms

TABLE 5 | Pairwise contrasts of estimated marginal means for the predictor ‘Condition’ (by Position and TextForm) for Experiments L1A and L1B (native participants).

Contrast of condition	Textform	Position	Estimate	*SE*	df	*t*-ratio	*p*-value	Signif.
Inflected – infinitive	Inflected	*n*	0.021	0.022	6451.0	0.997	1.000	
Inflected – conversion	Inflected	*n*	0.017	0.022	6451.0	0.788	1.000	
Inflected – noun	Inflected	*n*	−0.127	0.022	6451.0	−5.892	<0.001	***
Infinitive – conversion	Inflected	*n*	−0.005	0.022	6451.0	−0.209	1.000	
Infinitive – noun	Inflected	*n*	−0.149	0.022	6451.0	−6.889	<0.001	***
Conversion – noun	Inflected	*n*	−0.144	0.022	6451.0	−6.680	<0.001	***
Inflected – infinitive	Inflected	$n + 1$	0.017	0.022	6451.0	0.812	1.000	
Inflected – conversion	Inflected	$n + 1$	0.005	0.022	6451.0	0.212	1.000	
Inflected – noun	Inflected	$n + 1$	−0.102	0.022	6451.0	−4.735	<0.001	***
Infinitive – conversion	Inflected	$n + 1$	−0.013	0.022	6451.0	−0.600	1.000	
Infinitive – noun	Inflected	$n + 1$	−0.120	0.022	6451.0	−5.547	<0.001	***
Conversion – noun	Inflected	$n + 1$	−0.107	0.022	6451.0	−4.947	<0.001	***
Inflected – infinitive	Conversion	*n*	0.015	0.022	6454.24	0.678	1.000	
Inflected – conversion	Conversion	*n*	0.005	0.022	6454.24	0.251	1.000	
Inflected – noun	Conversion	*n*	−0.054	0.022	6454.24	−2.469	0.081	+
Infinitive – conversion	Conversion	*n*	−0.009	0.022	6454.24	−0.427	1.000	
Infinitive – noun	Conversion	*n*	−0.069	0.022	6454.24	−3.148	0.010	*
Conversion – noun	Conversion	*n*	−0.059	0.022	6454.24	−2.720	0.039	*
Inflected – infinitive	Conversion	$n + 1$	0.009	0.022	6454.25	0.391	1.000	
Inflected – conversion	Conversion	$n + 1$	0.083	0.022	6454.25	3.797	0.001	**
Inflected – noun	Conversion	$n + 1$	−0.057	0.022	6454.25	−2.615	0.054	+
Infinitive – conversion	Conversion	$n + 1$	0.074	0.022	6454.25	3.406	0.004	**
Infinitive – noun	Conversion	$n + 1$	−0.066	0.022	6454.25	−3.006	0.016	*
Conversion – noun	Conversion	$n + 1$	−0.140	0.022	6454.25	−6.412	<0.001	***

*p-value adjustment: Bonferroni method; degrees-of-freedom method: Satterthwaite. Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10.*

were presented in the text than in Experiment L1B where a conversion noun was presented in the text. This indicates that the word class status that was present in the input did influence the mental representation of the new word. This is further supported by the shorter reading times in the conversion condition in Experiment L1B at position $n + 1$ where the conversion noun was also presented in the input.

In the following two experiments we explored whether advanced L2 learners employ the generalized knowledge about German in the same way as the native speakers and whether their initial representations have lower resolution on the grammatical level, i.e., are more fuzzy.

NON-NATIVE PARTICIPANTS: EXPERIMENTS L2A AND L2B

The two experiments with non-native participants were structured and analyzed in exactly the same way as their L1 counterparts and thus only the information about the participants and the results is presented. As mentioned in the Introduction, in Czech, which was the participants' L1, there is no analogous process to the zero-derivation found for verbs and conversion nouns in German.

Participants

All non-native participants were native speakers of Czech who learned German as a foreign language. Their language proficiency in German was assessed prior to the actual experiments. Three different measures were obtained for each participant: a version of the Goethe Test, an online version of DiaLang (subtest on lexical knowledge), and a self-evaluation questionnaire. The classification by the three tests was not always consistent, with participants scoring at the B2 level in some test(s) and on C1 level at the other(s). Only those participants who scored at the B2 and/or C1 levels according to the Common European Framework of Reference for Languages (CEFR) in any of the three tests were selected for participation in the following experiments. As mentioned in the introduction, it can be safely assumed that Czech learners at B2/C1 level are well familiar with the investigated grammatical phenomena.

In Experiment L2A, the final group of non-native participants comprised 72 learners (62 females and 10 males) with a mean age of 23.8 years ($sd = 7.3$, range = 18–65).

In Experiment L2B, the final group of non-native participants comprised 68 learners (55 female and 13 male) with a mean age of 24.9 years ($sd = 4.4$, range = 19–41). None of the L2B participants took part in Experiment L2A.

Results

Results of Experiments L2A and L2B are summarized in **Table 6** and **Figure 2**.

The results of the final model [log(RT) ∼ Condition * Position * Textform + (Condition + Position || Item) + (Condition + Position || Participant)] are summarized in **Table 7**. They reveal significant main effects of Condition [$F(3,32.04) = 15.27$, $p < 0.001$] and Position [$F(1,108.18) = 45.87$, $p < 0.001$] and a significant interaction of Condition:Position [$F(3,6044.78) = 8.94$, $p < 0.001$]. Importantly, there was also a significant 3-way interaction of Condition:Position:Textform [$F(3,6044.78) = 5–64$, $p < 0.001$] indicating that the effect of Condition was moderated by an interaction of both Position and Textform. In order to investigate potential differences between the conditions of this interaction, *post hoc* comparisons of estimated (marginal) means were computed with *p*-adjustment for the accumulated alpha error according to the Bonferroni procedure. The results of these comparisons are summarized in **Table 8**. For experiment L2A, in which the novel word was introduced as an inflected verb form, results for position n yielded a pattern similar to that for L1 participants: While latencies for the inflected, the infinitive, and the conversion condition were equally fast (all $p > 0.999$), they were faster than the noun condition (infinitive $p = 0.003$; conversion condition $p = 0.020$; and inflected condition $p = 0.088$). At position $n + 1$, the noun condition was slower only than the conversion condition ($p = 0.030$), while none of the other comparisons yielded significant differences (all $p \geq 0.230$).

A different pattern was seen when the novel word was introduced as a conversion form (Experiment L2B). While at position n the noun condition again elicited the slowest responses (all $p \leq 0.007$), the situation for the three faster conditions was more diverse. The conversion (648.1 ms) condition was significantly faster than the inflected condition (726.6 ms) ($p < 0.001$) and the infinitive condition (688.6 ms) did not differ significantly from either of them ($p = 0.090$ and $p = 0.203$). At the same time, at position $n + 1$, no significant differences between conditions were observed.

Discussion

The analyses of Experiments L2A and L2B reveal a different pattern of results depending on the form presented in the introductory text. Results suggest that in Experiment L2A, when the form in the introductory text was an inflected verb form, L2 learners could establish a mental representation that comprised also the conversion noun information as indicated by the verbal forms and the conversion noun having been read equally fast when presented in the SPR sentence. The L2 participants also reacted with longer reading times in the countable noun condition indicating that it was not a part of the mental representation they established for the new word. In this respect their results mirror those of the L1 participants in Experiment L1A. However, the effect was distinctly weaker at the spill-over region $n + 1$ where the countable noun differed significantly only from the conversion condition.

In Experiment L2B, however, the evidence that the L2 participants could establish a mental representation for both the verbal and the conversion noun forms when presented with the conversion form in the introductory texts is less convincing. First, results show that participants were fastest when reading the conversion form in the SPR sentence (at position n) indicating the superiority of this component of the new mental representation compared to the verbal component. This idea is especially supported by the observation that

TABLE 6 | Results of Experiments L2A and L2B (mean RTs in ms and SDs).

	L2A (inflected form in texts)				L2B (conversion noun in text)			
	Position *n*		Position *n* + 1		Position *n*		Position *n* + 1	
Condition	**RT**	***(SD)***	**RT**	***(SD)***	**RT**	***(SD)***	**RT**	***(SD)***
Inflected	635.0	(345.0)	543.0	(194.6)	726.6	(481.2)	557.9	(253.2)
Infinitive	617.4	(335.4)	526.6	(183.5)	688.6	(465.8)	550.6	(268.9)
Conversion	629.6	(344.1)	526.6	(244.0)	648.1	(433.1)	520.1	(229.2)
Noun	715.0	(479.0)	572.5	(266.3)	800.8	(520.2)	537.3	(237.8)

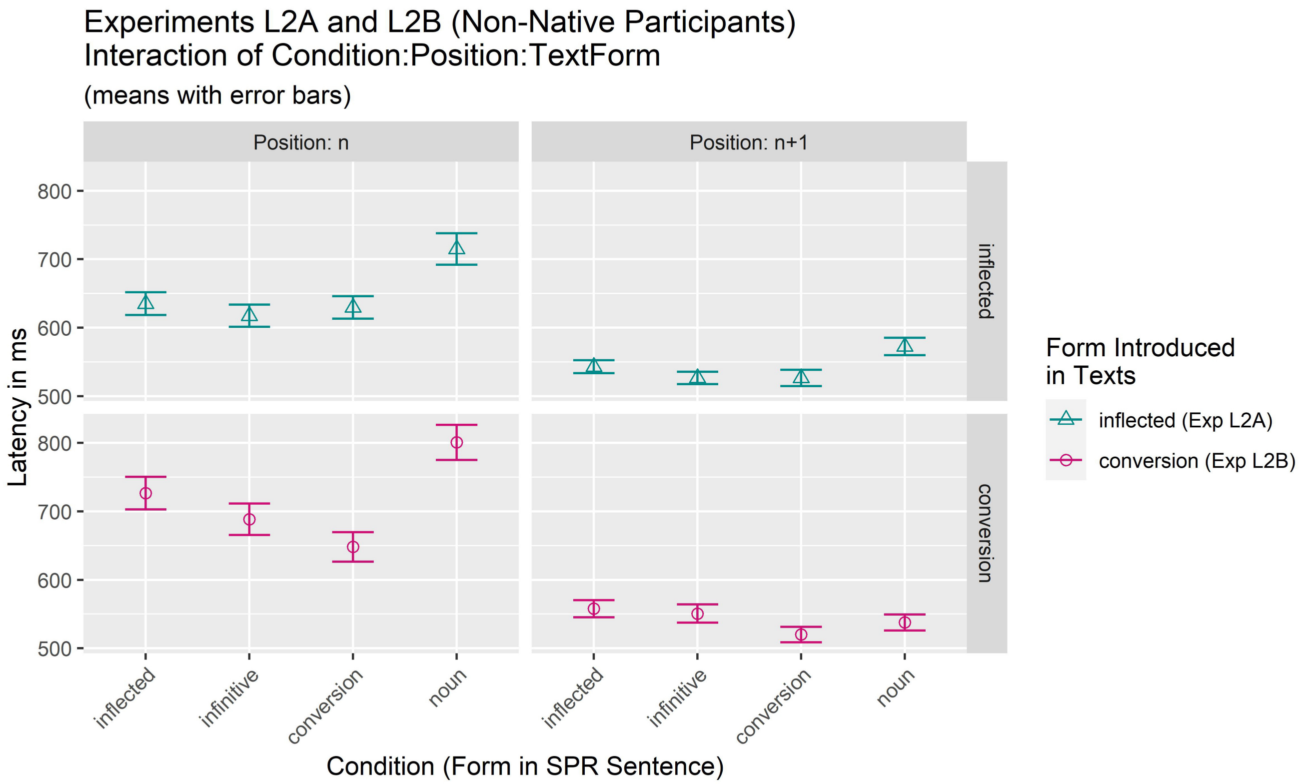


FIGURE 2 | Results of Experiments L2A and L2B: non-native participants. (Mean latencies of critical regions in SPR sentences).

the inflected verb form is read significantly slower than the conversion form at position n, suggesting that the representation of the verbal component after a conversion noun in input was only weak and possibly fuzzier. Moreover, the analyses further revealed that the effect indicating processing difficulties when reading the countable noun was not present at position $n + 1$, which suggests that the established mental representation of the new form presented in the introductory text did not enable a clear differentiation between the conversion noun and another noun (countable) that should be a separate entity. This contrasts sharply with the results of the L1 experiments at both positions. Overall, the new L2 representation established after the presentation of the conversion form in the introductory text is fuzzier than the new representation established after the presentation of the verbal form in the introductory text in L2 and also fuzzier than the new representation established under the same conditions (conversion form in the introductory text) in L1. The fuzziness seems to result from the fact that when the more specific, derived conversion form is presented in the input, the L2 learners are unable to employ their general, possibly limited knowledge about German grammar so effectively as in the case when the more basic verbal form is encountered.

Finally, we investigated whether the different pattern of results obtained for native participants (Experiments L1A and L1B) and non-native speakers (Experiments L2A and L2B) could be substantiated also statistically. We therefore carried out an additional overall analysis of all four experiments containing the additional factor Language (i.e., L1 vs. L2). We were especially interested in whether the

TABLE 7 | Mixed model ANOVA table for Experiments L2A and L2B (non-native participants).

Effect	df	*F*	*p*-value	Signif.
Condition	3, 32.04	15.27	<0.001	***
Position	1, 108.18	45.87	<0.001	***
TextForm	1, 137.97	0.35	0.555	
Condition:Position	3, 6044.78	8.94	<0.001	***
Condition:TextForm	3, 218.16	2.50	0.060	+
Position:TextForm	1, 138.00	3.50	0.063	+
Condition:Position:TextForm	3, 6044.78	5.64	<0.001	***

*Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10.*

interaction of Condition:Position:Textform seen for both L1 (Experiments L1A and L1B) and L2 (Experiments L2A and L2B) participants separately was moderated by the factor Language in the overall analysis. This was indeed the case. The final model [log(RT) ∼ Condition * Position * Textform * Language + (Condition + Position + Language || Item) + (Condition + Position || Participant)] yielded a significant 4-way interaction of Condition:Position:Textform:Language [$F(3,12273.57) = 7.67$, $p < 0.001$; for full model results see **Table 9**].

GENERAL DISCUSSION

In the present study, we addressed the question whether readers use previously acquired, generalized, grammatical knowledge to establish new lexical entries that would contain information not deducible from the immediate input and whether this ability depends on the properties of the word that appears in the input (i.e., verbal form vs. conversion noun form in our experiments). We further explored how these two aspects, i.e., (a) the engagement of previously acquired grammatical knowledge and (b) the specific properties of the newly encountered word interact with fuzziness as a characteristic property of not yet firmly acquired representations, typical especially for L2 learners (Bordag et al., 2021a,b; Gor et al., 2021). We took advantage of the existence of homonymous forms in the German language that can have various functions in the text and focused on the relationship between verbs and conversion nouns derived from them by a productive process. Recent evidence indicates that conversion noun information and the corresponding verbal information are represented within a joint, structured lexical entry as two distinct components (Bordag and Opitz, 2021; Opitz and Bordag, 2021).

The experiments yielded different patterns of results for native speakers and advanced L2 learners of German that partly depended on the properties of the word form presented in the introductory texts. For the L1 speakers, this factor played a minor role: they could access both the verbal and the conversion noun representational component in the SPR sentence equally fast, irrespective of which of the two forms was presented in the input (according to Hypothesis 2A). This indicates that they have good, reliable linguistic knowledge about the generalizable grammatical relations between a verb and a conversion noun and they can employ this knowledge when establishing new representations [in line with the findings of Bordag et al. (2015b)]. At the same time, their new representations of such forms have high enough resolution to be recognized and processed as different from other words that share the same surface form but have incompatible grammatical properties (homonymous countable noun forms).

Advanced L2 learners of German also possess some generalized grammatical knowledge and can employ it to a similar extent to native speakers when establishing new representations, but with specific limitations (according to Hypothesis 2B). When the form in the text input is the base (i.e., verbal) form, they can induce that the to-be-established entry needs to comprise both the verbal and the conversion noun component and their results mirror those of the L1 natives as can be seen in Experiments L2A and L1A. However, when the form in the text input is the more specific, less frequent and a derived conversion noun, the representation they establish is more incomplete and fuzzy: the verbal component is present to some degree (since the verbal conditions are still faster than the grammatically unrelated countable noun condition at least at position *n*), but it is obviously less well established than the conversion noun component as evidenced by the processing delay compared to the conversion condition. This 'internal' fuzziness within the lexical verb/conversion noun representation is accompanied by 'external' effects of fuzziness that reduces the differentiation of this representation from other, similar representations – such as the countable noun representation. It indicates that the L2 learners were able to establish a noun representation within the verb/conversion noun entry, but that this representation was not specific, clearly defined enough in its grammatical properties. In particular, the feature 'uncountable' or 'singulare tantum' which is characteristic for conversion nouns was only weakly represented in the new L2 representation. Therefore, the countable noun presented in plural contexts in the SPR sentences did not lead to pronounced, strong and lasting incongruence effects seen for L1 (cf. position $n + 1$ in Experiment L2B). This parallels the findings which the FLR hypothesis reports as evidence for fuzzy semantic and in particular phonological representations, in which an imprecise or a missing representation/encoding of a feature can lead to semantic or phonological confusions (e.g., Ota et al., 2009; Darcy et al., 2013; Cook and Gor, 2015; Cook et al., 2016; Llompart and Reinisch, 2019). Obviously, also grammatical fuzziness manifests itself through less distinct boundaries (in this case of the grammatical components of the representation), which leads to deficiencies in differentiation from neighboring representations.

All these findings are in accordance with the FLR hypothesis and the OM. While both frameworks are based primarily on the evidence from phonology, orthography and semantics, the presented study delivers evidence supporting these approaches also in the area of grammar. As suggested in the OM, morphosyntax/grammar may be another domain within the dimension of linguistic domains that comprises phonological, orthographic, and semantic domains in the model. Though the topic of fuzziness and its reduction during the ontogenetic development in the grammatical domain has not been directly addressed in previous studies, reconsideration of some of the previous findings indicates that grammar could be recognized as another domain in the model at which fuzziness operates in a similar way like in the other domains.

The OM is a model of individual lexical representations, and this is also the primary scope of the FLR hypothesis. As mentioned in the Introduction, a substantial part of grammar is handled by the mechanisms and procedures that operate on representations in the mental lexicon but may not be part of them – thus they are addressed neither by the OM nor the FLR hypothesis[7]. However, the aspects of grammar that are assumed to be stored in individual lexical entries, such as the word-class information or number information of pluralia tantum, could form the contents of the grammar or – maybe more precisely – morphosyntactic feature domain and could be grasped by the OM using its central concepts of the optimum, fuzziness, and ontogenesis.

Considering the whole grammatical domain of a single lexical entry, its optimum would be reached when all grammatical features of a given word class in the given language are acquired, including a stable representation of correctly set fixed values of the internal features. Missing, unstable or incorrectly set features would be factors that would determine the degree of fuzziness in this domain, analogically to how fuzziness is captured in the FLR hypothesis for the other domains. From a

[7]Though the FLR hypothesis discusses only representational fuzziness (which is also in the focus of our present study), it acknowledges the existence of fuzziness at the processing level, too (Gor et al., 2021).

TABLE 9 | Mixed model ANOVA table for all 4 experiments.

Effect	df	*F*	*p*-value	Signif.
Condition	3, 39.36	18.81	<0.001	***
Position	1, 78.57	26.86	<0.001	***
TextForm	1, 278.00	0.04	0.842	
Language	1, 257.00	31.41	<0.001	***
Condition:Position	3, 12272.73	5.69	<0.001	***
Condition:TextForm	3, 438.80	4.13	0.007	**
Position:TextForm	1, 277.99	0.88	0.350	
Condition:Language	3, 439.01	1.97	0.118	
Position:Language	1, 277.99	35.82	<0.001	***
Text.Form:Language	1, 278.02	0.55	0.459	
Condition:Position:TextForm	3, 12272.73	1.06	0.364	
Condition:Position:Language	3, 12273.57	5.05	0.002	**
Condition:TextForm:Language	3, 439.01	1.46	0.225	
Position:TextForm:Language	1, 277.99	4.79	0.029	*
Condition:Position:TextForm: Language	3, 12273.57	7.68	<0.001	***

*Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10.*

TABLE 8 | Pairwise contrasts of estimated marginal means for the predictor 'Condition' (by Position and TextForm) for Experiments L2A and L2B (non-native participants).

Contrast of condition	Textform	Position	Estimate	*SE*	df	*t*-ratio	*p*-value	Signif.
Inflected – infinitive	Inflected	*n*	0.029	0.025	715.6	1.175	1.000	
Inflected – conversion	Inflected	*n*	0.008	0.024	669.8	0.316	1.000	
Inflected – noun	Inflected	*n*	−0.069	0.028	118.4	−2.478	0.088	+
Infinitive – conversion	Inflected	*n*	−0.022	0.024	868.8	−0.894	1.000	
Infinitive – noun	Inflected	*n*	−0.098	0.027	128.7	−3.605	0.003	**
Conversion – noun	Inflected	*n*	−0.077	0.026	145.8	−2.986	0.020	*
Inflected – infinitive	Inflected	*n* + 1	0.027	0.025	715.6	1.077	1.000	
Inflected – conversion	Inflected	*n* + 1	0.043	0.024	669.8	1.779	0.454	
Inflected – noun	Inflected	*n* + 1	−0.030	0.028	118.4	−1.083	1.000	
Infinitive – conversion	Inflected	*n* + 1	0.016	0.024	868.8	0.683	1.000	
Infinitive – noun	Inflected	*n* + 1	−0.057	0.027	128.7	−2.089	0.232	
Conversion – noun	Inflected	*n* + 1	−0.073	0.026	145.8	−2.856	0.030	*
Inflected – infinitive	Conversion	*n*	0.062	0.026	769.8	2.437	0.090	+
Inflected – conversion	Conversion	*n*	0.115	0.025	722.4	4.596	<0.001	***
Inflected – noun	Conversion	*n*	−0.095	0.029	127.1	−3.316	0.007	**
Infinitive – conversion	Conversion	*n*	0.053	0.025	933.6	2.124	0.203	
Infinitive – noun	Conversion	*n*	−0.157	0.028	138.0	−5.617	<0.001	***
Conversion – noun	Conversion	*n*	−0.209	0.026	156.7	−7.941	<0.001	***
Inflected – infinitive	Conversion	*n* + 1	0.006	0.026	769.8	0.251	1.000	
Inflected – conversion	Conversion	*n* + 1	0.055	0.025	722.4	2.188	0.174	
Inflected – noun	Conversion	*n* + 1	0.031	0.029	127.1	1.076	1.000	
Infinitive – conversion	Conversion	*n* + 1	0.048	0.025	933.6	1.950	0.309	
Infinitive – noun	Conversion	*n* + 1	0.024	0.028	138.0	0.870	1.000	
Conversion – noun	Conversion	*n* + 1	−0.024	0.026	156.7	−0.909	1.000	

*p-value adjustment: Bonferroni method; degrees-of-freedom method: Satterthwaite. Significance codes: ***p < 0.001; **p < 0.01; *p < 0.05; +p < 0.10.*

more differentiated perspective, ontogenetic development of the individual features toward their optima can be considered, too. As an example, research on grammatical gender (e.g., Bordag and Pechmann, 2007) indicates that during its ontogenesis, a fuzzy phase of gender value computation based on various sources (phonological form of the word, its L1 gender value, unstably set L2 gender value) precedes the final, optimum, stage when the gender value is firmly fixed and automatically retrieved (not computed each time anew). Similarly, the results of the present study suggest that the examined newly established L2 representations follow a developmental trajectory from a weak representation of a word class information that leads to processing difficulties when accessing the verbal component of the verb/conversion noun representation and from low-resolution representation of the fixed number that compromises the differentiation from homonymous, but separate countable noun representations (positions $n + 1$ of Experiments 2A and 2B) toward a more precise grammatical representation that manifests itself in functional equivalence comparable to the L1.

Though the present study delivers promising results in areas such as incidental vocabulary acquisition, grammar acquisition, and research on FLR and ontogenetic development of individual representations at the grammar domain, more research is clearly needed to substantiate the presented claims. As an example, the current study was limited in that we explored advanced L2 learners and hypothesized about the ontogenetic development of their newly established representations based on the comparison with L1 and on acquisition in two differently difficult learning contexts (experimental versions A and B). In order to gain a clearer picture of, for instance, such developmental aspects, future research should address comparisons between participants at different proficiency levels and in longitudinal studies. Moreover, examining different L1–L2 pairings, for example, could help determine the role of cross-linguistic transfer in resolving fuzziness. With respect to practical aspects of language instructions, it would also be interesting to explore whether fuzziness in the explored area can be reduced by particular teaching methods or training.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article and the R script used to generate all reported results are publicly available in the Mendeley database: doi: 10.17632/548vjgyk6t.1.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by German Research Council (DFG). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (grant number BO 3615/6-1 to DB).

ACKNOWLEDGMENTS

The authors acknowledge support from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing.

REFERENCES

Aka, N. (2020). Incidental learning of a grammatical feature from reading by Japanese learners of English as a foreign language. *System* 91:102250. doi: 10.1016/j.system.2020.102250

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

Barrios, S., and Hayes-Harb, R. (2020). The relationship between native English-speaking learners' perception and lexical representation of Hindi affricates. *J. Acoust. Soc. Am.* 148, 2812–2813. doi: 10.1121/1.5147843

Barrios, S., and 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. *Front. Comm.* 6:689470. doi: 10.3389/fcomm.2021.68947

Bordag, D., Gor, K., and Opitz, A. (2021a). Ontogenesis Model of the L2 Lexical Representation. *Bilingualism* 2021, 1–17. doi: 10.1017/S1366728921000250

Bordag, D., Kirschenbaum, A., Opitz, A., and Tschirner, E. (2015b). Incidental acquisition of grammatical features during reading in L1 and L2. *Stud. Sec. Lang. Acq.* 38, 445–483. doi: 10.1017/S0272263115000315

Bordag, D., Kirschenbaum, A., Rogahn, M., Opitz, A., and Tschirner, E. (2017a). Semantic representation of newly learned L2 words and their integration in the L2 Lexicon. *Stud. Sec. Lang. Acq.* 39, 197–212. doi: 10.1017/S0272263116000048

Bordag, D., Kirschenbaum, A., Rogahn, M., and Tschirner, E. (2017b). The role of orthotactic probability in incidental and intentional vocabulary acquisition L1 and L2. *Sec. Lang. Res.* 33, 147–178. doi: 10.1177/0267658316665879

Bordag, D., Kirschenbaum, A., Tschirner, E., and Opitz, A. (2015a). Incidental acquisition of new words during reading in L2: inference of meaning and its integration in the L2 Mental Lexicon. *Bilingualism* 18, 372–390. doi: 10.1017/S1366728914000078

Bordag, D., and Opitz, A. (2021). Representation and processing of overtly identical complex forms in L1 and L2: The case of conversion and its 'siblings'. *Stud. Sec. Lang. Acq.* 2021:S0272263121000206. doi: 10.1017/S0272263121000206

Bordag, D., Opitz, A., Polter, M., and Meng, M. (2021b). Non-native readers are more sensitive to changes in surface linguistic information than native readers. *Bilingualism* 24, 1–13. doi: 10.1017/S1366728920000772

Bordag, D., Opitz, A., Rogahn, M., and Tschirner, E. (2018). The Strength of meaning Representations of new words with different properties in the L2 semantic network. *Ment. Lex.* 13, 74–104. doi: 10.1075/ml.17005.bor

Bordag, D., and Pechmann, T. (2007). Factors influencing L2 gender processing. *Bilingualism* 10, 299–314. doi: 10.1017/S1366728907003082

Bordag, D., and Pechmann, T. (2009). Externality, internality, and (in)dispensability of grammatical features in speech production: evidence from Czech declension and conjugation. *J. Exp. Psychol.* 35, 446–465. doi: 10.1037/a0014874

Borovsky, A., Kutas, M., and Elman, J. (2010). Learning to use words: Event-related potentials index single-shot contextual word learning. *Cognition* 116, 289–296. doi: 10.1016/j.cognition.2010.05.004

Breitenstein, C., Zwitserlood, P., de Vries, M. H., Feldhues, C., Knecht, S., and Dobel, C. (2007). Five days versus a lifetime: Intense associative vocabulary training generates lexically integrated words. *Rest. Neurol. Neurosci.* 25, 493–500.

Bylund, E., Hyltenstam, K., and Abrahamsson, N. (2021). Age of acquisition–not bilingualism–is the primary determinant of less than nativelike L2 ultimate attainment. *Bilingualism* 24, 18–30. doi: 10.1017/S1366728920000188

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cogn. Neuropsychol.* 14, 177–208. doi: 10.1080/026432997381664

Clahsen, H., and Felser, C. (2006). Grammatical processing in language learners. *Appl. Psycholing.* 27, 3–42. doi: 10.1017/S0142716406060024

Clahsen, H., and Felser, C. (2017). Some notes on the Shallow Structure Hypothesis. *Stud. Sec. Lang. Acq.* 40, 1–14. doi: 10.1017/s0272263117000250

Clay, F., Bowers, J. S., Davis, C. J., and Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *J. Exp. Psychol.* 33, 970–976. doi: 10.1037/0278-7393.33.5.970

Cook, S. V., and Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *Ment. Lex.* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Darcy, I., Daidone, D., and Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *Men. Lex.* 8, 372–420. doi: 10.1075/ml.8.3.06dar

De Groot, A. M., Dannenburg, L., and van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *J. Mem. Lang.* 33, 600–629. doi: 10.1006/jmla.1994.1029

DeKeyser, R. (2012). Interactions between individual differences, treatments, and structures in SLA. *Lang. Learn.* 62, 189–200. doi: 10.1111/j.1467-9922.2012.00712.x

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psycholog. Rev.* 93:283. doi: 10.1037/0033-295X.93.3.283

Denhovska, N., Serratrice, L., and Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: the role of frequency and working memory. *Lang. Learn.* 66, 159–190. doi: 10.1111/lang.12142

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Lang. Learn.* 61, 367–413. doi: 10.1111/j.1467-9922.2010.00613.x

Elley, W. B., and Mangubhai, F. (1983). The impact of reading on second language learning. *Read. Res. Q.* 19, 53–67. doi: 10.2307/747337

Ellis, N., and Collins, L. (2009). Input and second language acquisition: the roles of frequency, form, and function introduction to the special issue. *Mod. Lang. J.* 93, 329–335. doi: 10.1111/j.1540-4781.2009.00893.x

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., and Opitz, A. (2021). Fuzzy lexical representations in adult second language speakers. *Front. Psychol.* 12:732030. doi: 10.3389/fpsyg.2021.732030

Gor, K., and Cook, S. V. (2020). A mare in a pub? Nonnative Facilitation in Phonological Priming. *Sec. Lang. Res.* 36, 123–140. doi: 10.1177/0267658318769962

Grey, S., Williams, J. N., and Rebuschat, P. (2014). Incidental exposure and L3 learning of morphosyntax. *Stud. Sec. Lang. Acq.* 36, 611–645. doi: 10.1017/S0272263113000727

Hartshorne, J. K., Tenenbaum, J. B., and Pinker, S. (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition* 177, 263–277. doi: 10.1016/j.cognition.2018.04.007

Hulstijn, J. H. (1992). "Retention of inferred and given word meanings: Experiments in incidental vocabulary learning," in *Vocabulary and Applied Linguistics*, eds P. J. Arnaudand and H. Béjoint (London: Macmillan), 113–125. doi: 10.1007/978-1-349-12396-4_11

Jarvis, S. (2000). Methodological rigor in the study of transfer: identifying L1 influence in them interlanguage lexicon. *Lang. Learn.* 50, 245–309. doi: 10.1111/0023-8333.00118

Jiang, N., and Zhang, J. (2019). Form prominence in the L2 lexicon: further evidence from word association. *Sec. Lang. Res.* 37, 69–90. doi: 10.1177/0267658319827053

Just, M. A., Carpenter, P. A., and Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *J. Exp. Psychol.* 111, 228–238. doi: 10.1037/0096-3445.111.2.228

Keuleers, E., and Brysbaert, M. (2010). Wuggy: a multilingual pseudoword generator. *Behav. Res. Methods* 42, 627–633. doi: 10.3758/BRM.42.3.627

Lenth, R. V., Buerkner, P., Herve, M. J., Love, J., Riebl, H., and Singmann, H. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means. Version 1.5.4*. doi: 10.1080/00031305.1980.10483031 Availabe online at: https://cran.r-project.org/package=emmeans/ (accessed November 29, 2021).

Levelt, W. J., Roelofs, A., and Meyer, A. S. (1999). A theory of lexical access in speech production. *Behav. Brain Sci.* 22, 1–38. doi: 10.1017/S0140525X99001776

Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge MA: The MIT Press.

Llompart, M. (2021). Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism* 24, 481–496. doi: 10.1017/S1366728920000656

Llompart, M., and Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Biling. Lang. Congn.* 22, 1085–1100. doi: 10.1017/S1366728918000925

Mestres-Missé, A., Rodriguez-Fornells, A., and Münte, T. F. (2007). Watching the brain during meaning acquisition. *Cereb. Cort.* 17, 1858–1866. doi: 10.1093/cercor/bhl094

Monaghan, P., Ruiz, S., and Rebuschat, P. (2021). The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Mixed effects models reveal local and global effects on acquisition. *Sec. Lang. Res.* 37, 261–289. doi: 10.1177/0267658320927741

Nicklin, C., and Plonsky, L. (2020). Outliers in L2 research in applied linguistics: a synthesis and data analysis. *Annu. Rev. Appl. Ling.* 40, 26–55. doi: 10.1017/S0267190520000057

Opitz, A., and Bordag, D. (2021). The impact of orthography on lexical access: the case of capitalisation and word category information in L1 and L2 German. *Stud. Sec. Lang. Acq.* 2021, 1–16. doi: 10.1017/S0272263121000711

Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition* 111, 263–269. doi: 10.1016/j.cognition.2008.12.007

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Sci. Stud. Read.* 11, 357–383. doi: 10.1080/10888430701530730

Perfetti, C. A., and Hart, L. (2002). The lexical quality hypothesis. *Precurs. Funct. Lit.* 11, 67–86.

Perfetti, C. A., Wlotko, E. W., and Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *J. Exp. Psychol.* 31, 1281–1292. doi: 10.1037/0278-7393.31.6.1281

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Rebuschat, P., Monaghan, P., and Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition* 206:104475. doi: 10.1016/j.cognition.2020.104475

Reichle, E. D., Warren, T., and McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychon. Bull. Rev.* 16, 1–21. doi: 10.3758/PBR.16.1.1

Sampaio, C., and Konopka, A. E. (2012). Memory for non-native language: the role of lexical processing in the retention of surface form. *Memory* 21, 537–544. doi: 10.1080/09658211.2012.746371

Shintani, N. (2015). The incidental grammar acquisition in focus on form and focus on forms instruction for young beginner learners. *TESOL Q.* 49, 115–140. doi: 10.1002/tesq.166

Shintani, N., and Ellis, R. (2010). The incidental acquisition of English plural–s by Japanese children in comprehension-based and production-based lessons: A process-product study. *Stud. Sec. Lang. Acq.* 32, 607–637. doi: 10.1017/S0272263110000288

Shtyrov, Y., Nikulin, V. V., and Pulvermueller, F. (2010). Rapid Cortical Plasticity Underlying Novel Word Learning. *J. Neurosci.* 30, 16864–16867. doi: 10.1523/JNEUROSCI.1376-10.2010

Singmann, H. (2021). *Mixed Model Reanalysis of RT data*.Available online at: https://cran.r-project.org/web/packages/afex/vignettes/afex_mixed_example.html (accessed November 26, 2021).

Singmann, H., Bolker, B., Westfall, J., Aust, F., and Ben-Shachar, M. S. (2021). *afex: Analysis of Factorial Experiments. R package version 1.0-1*. Available online at: https://CRAN.R-project.org/package=afex (accessed July 22, 2021).

Song, J., and Sardegna, V. G. (2014). EFL learners' incidental acquisition of English prepositions through enhanced extensive reading instruction. *RELC* J. 45, 67–84. doi: 10.1177/0033688214522623

Tamminen, J., and Gaskell, M. G. (2013). Novel word integration in the mental lexicon: evidence from unmasked and masked semantic priming. *Q. J. Exp. Psychol.* 66, 1001–1025. doi: 10.1080/17470218.2012.724694

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychology

ORIGINAL RESEARCH
published: 28 July 2022
doi: 10.3389/fpsyg.2022.742965

Morphological Priming Effects in L2 English Verbs for Japanese-English Bilinguals

*Jessie Wanner-Kawahara[1], Masahiro Yoshihara[1,2], Stephen J. Lupker[3], Rinus G. Verdonschot[4] and Mariko Nakayama[1]**

[1]International Graduate Program in Language Sciences, Graduate School of International Cultural Studies, Tohoku University, Sendai, Japan, [2]Research Institute for Letters, Arts, and Sciences, Waseda University, Tokyo, Japan, [3]Department of Psychology, University of Western Ontario, London, ON, Canada, [4]Max Plank Institute for Psycholinguistics, Nijmegen, Netherlands

OPEN ACCESS

Edited by:
Denisa Bordag,
Leipzig University, Germany

Reviewed by:
Francesco Vespignani,
Università degli Studi di Padova, Italy
Jonathan Grainger,
Centre National de la Recherche Scientifique (CNRS), France
Say Young Kim,
Hanyang University, South Korea
Rosane Silveira,
Federal University of Santa Catarina, Brazil

****Correspondence:***
Mariko Nakayama
mariko.nakayama.d5@tohoku.ac.jp

Specialty section:
This article was submitted to Language Sciences, a section of the journal Frontiers in Psychology

Received: *17 July 2021*
Accepted: *21 June 2022*
Published: *28 July 2022*

Citation:
Wanner-Kawahara J, Yoshihara M, Lupker SJ, Verdonschot RG and Nakayama M (2022) Morphological Priming Effects in L2 English Verbs for Japanese-English Bilinguals. Front. Psychol. 13:742965. doi: 10.3389/fpsyg.2022.742965

For native (L1) English readers, masked presentations of past-tense verb primes (e.g., fell and looked) produce faster lexical decision latencies to their present-tense targets (e.g., FALL and LOOK) than orthographically related (e.g., fill and loose) or unrelated (e.g., master and bank) primes. This facilitation observed with morphologically related prime-target pairs (morphological priming) is generally taken as evidence for strong connections based on morphological relationships in the L1 lexicon. It is unclear, however, if similar, morphologically based, connections develop in non-native (L2) lexicons. Several earlier studies with L2 English readers have reported mixed results. The present experiments examine whether past-tense verb primes (both regular and irregular verbs) significantly facilitate target lexical decisions for Japanese-English bilinguals beyond any facilitation provided by prime-target orthographic similarity. Overall, past-tense verb primes facilitated lexical decisions to their present-tense targets relative to both orthographically related and unrelated primes. Replicating previous masked priming experiments with L2 readers, orthographically related primes also facilitated target recognition relative to unrelated primes, confirming that orthographic similarity facilitates L2 target recognition. The additional facilitation from past-tense verb primes beyond that provided by orthographic primes suggests that, in the L2 English lexicon, connections based on morphological relationships develop in a way that is similar to how they develop in the L1 English lexicon even though the connections and processing of lower level, lexical/orthographic information may differ. Further analyses involving L2 proficiency revealed that as L2 proficiency increased, orthographic facilitation was reduced, indicating that there is a decrease in the fuzziness in orthographic representations in the L2 lexicon with increased proficiency.

Keywords: morphological priming, fuzzy lexicon, bilinguals, L2 English, proficiency

INTRODUCTION

Word recognition studies involving bilinguals have focused mainly on understanding the relationship between first (L1) and second language (L2) representations. Some of this focus stems from the debate on language selectiveness vs. non-selectiveness of lexical access (see Jiang, 2015 for a review), which now seems to favor the language non-selective access hypothesis (see Dijkstra, 2005 for a review). Understanding the structure and inner workings of the bilingual's L2 lexicon itself has now become another important focus of bilingual visual word recognition studies (e.g., Bordag et al., 2021). Relevant studies commonly address the question of whether they are the same as or different from those of the L1 lexicon. Some studies have suggested that L2 representations are fuzzy, meaning that lexical items in a second language may be encoded in a less precise manner than in the first language of a bilingual (Cook et al., 2016). This idea has been investigated in some detail for form-meaning mappings (in the L2). In contrast, there is, at this point, little information concerning whether the same is true for morphological level representations.

With respect to visual word recognition, some studies have reported that proficient L2 readers produce the same pattern of results as L1 readers (e.g., Witzel et al., 2011, transposed letter/character priming effects; Nakayama et al., 2013, frequency attenuation of repetition priming effects), suggesting that certain aspects of how L2 readers process and represent L2 words seem similar to those of L1 readers. Other studies with proficient L2 readers, however, have shown different patterns of results from those of L1 readers. For example, the word frequency effect has been found to be greater in L2 than in L1 (e.g., Duyck et al., 2008). It has also been shown that near-homophones (ROCK vs. LOCK) can produce an interference effect in a semantic relatedness judgment task (i.e., are ROCK and KEY related?) in L2 but not in L1 readers (Ota et al., 2009). This latter result seems to indicate that certain phono-lexical representations (e.g., those having a non-native /l/ - /r/ contrast) might indeed be fuzzy (i.e., stored imprecisely) and are therefore hard to separate for Japanese-English bilinguals. Furthermore, lexical competition, the process during which orthographically similar words compete with each other during the word recognition process (e.g., Segui and Grainger, 1990; Davis and Lupker, 2006), appears to be absent, or at least greatly diminished, for L2 readers (e.g., Qiao and Forster, 2017; Nakayama and Lupker, 2018; Jiang, 2021). Weak lateral inhibition in L2 learners has also been reported in the auditory domain (Gor and Cook, 2020). Although different behavioral results do not inevitably indicate that L1 and L2 lexicons are organized qualitatively differently (e.g., see Brysbaert et al., 2017), it is certainly the case that understanding both the similarities and differences between how words are processed and represented in the L1 and L2 lexicons is critical to gaining a clear picture of the bilingual language system.

In the present research, we explored a potential difference between L2 and L1 English lexicons by examining the representations of morphological relationships of L2 words for Japanese-English bilinguals. As previous studies have shown that representations for L2 word forms appear to differ from those for L1 words (e.g., Qiao and Forster, 2017; Nakayama and Lupker, 2018),[1] the question of whether differences between L1 and L2 representations also exist for representations of morphological relationships is clearly of interest. The type of morphological relationship examined here was that between past and present tense verb forms. Because of the extensiveness of the literature on this issue, we limit our discussion to masked priming visual word recognition experiments that investigate L2 English processing.

We focused on this particular morphological relationship because previous studies examining L1 (native) English readers have reliably observed significant masked priming effects between past-tense verb primes and their present-tense targets (fell-FALL, boiled-BOIL). Such priming effects indicate that there are special connections between the two types of words, due to their morphological relationship in the L1 English reader's lexicon (e.g., Forster et al., 1987; Crepaldi et al., 2010). In contrast to L1 studies, previous studies investigating L2 English readers have reported mixed results: some studies reporting that L2 readers produce priming patterns similar to those of L1 readers (Feldman et al., 2010; Voga et al., 2014) and others reporting that they do not (e.g., Silva and Clahsen, 2008; Clahsen et al., 2013). The latter pattern suggests that non-native readers of a language are not as sensitive to the morphological structure, or to morphological exceptions, as native readers are. Therefore, the L2 morphological makeup of particular words may perhaps constitute another source of fuzziness in the mental lexicon (similar to meaning-form mappings; Cook and Gor, 2015). However, we should point out that the question of whether L2 readers represent the relationship of past-tense verbs and their present-tense forms similarly to L1 readers does not have a clear answer at present.

In these masked priming lexical decision experiments, researchers typically compare the speed at which targets are responded to when the targets are preceded by a brief (40–60 ms) presentation of a prime that is in some way related to its target (e.g., orthographically, phonologically, semantically, morphologically, etc.) versus a prime that is unrelated to its target. When responses to targets are differentially affected by related versus unrelated primes (i.e., faster or slower), the latency difference is called a priming effect. This priming effect is thought to occur due to primes that are in some way related to their targets pre-activating their target's representations based on that relatedness. Significant priming effects, therefore, indicate that representations of primes and targets share processing structures as a result of their related property. Priming based on morphological relationships, that is, a morphological priming effect, is, therefore, thought to reflect some kind of connectedness between the representations of prime and target words based on morphology.

Forster et al. (1987) were among the first to demonstrate masked priming effects of the sort being examined here, that is, between irregular inflectional past-tense verbs and their present-tense forms (e.g., kept-KEEP) for L1 English readers. In their monolingual experiment using a 60 ms prime duration, targets primed by their past-tense forms were responded to significantly faster than the same targets primed by unrelated primes (e.g., kept-KEEP vs. navy-KEEP). Past-tense primes, in fact, facilitated target recognition as much as identity primes did (e.g., keep-KEEP = kept-KEEP < navy-KEEP; 36 ms vs. 37 ms effects), indicating that a past-tense verb has the ability to access or pre-activate present-tense verb representations as efficiently as the verb itself for native speakers. Using regular

[1]Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., and Opitz, A. (under review; this issue). Fuzzy lexical representations in adult second language speakers. Front. Psychol. 12.

past-tense and present-tense form pairs (e.g., boiled-BOIL), Silva and Clahsen (2008) also found significant priming effects relative to unrelated primes (jump-BOIL—a 55 ms effect). Similar to Forster et al. (1987), the size of the morphological priming effect was not statistically different from the size of the parallel identity priming effect (67 ms). Crepaldi et al. (2010), using a slightly shorter prime duration (42 ms), replicated the significant priming effect for irregular past-tense and present-tense verb pairs (e.g., fell-FALL vs. hope-FALL, 25 ms) with stringently controlled stimuli. Importantly, they added a crucial control condition, orthographic primes (e.g., fill-FALL). Specifically, the orthographic similarity of their orthographic primes and targets was matched to the orthographic similarity of their morphological primes and targets. The inclusion of orthographic primes thereby guarded against the possibility that any priming observed for the fell-FALL pairs might have been orthographically based. The morphological primes produced a significant (21 ms) priming effect using the orthographic condition as a baseline.

Finally, Pastizzo and Feldman (2002) tested both regular and irregular verbs in a single experiment and found significant morphological priming effects, again using orthographic control primes as baselines, for both irregular (e.g., fell-FALL vs. fill-FALL; a 33 ms effect) and regular verbs (billed-BILL vs. billion-BILL; a 44 ms effect), although only a non-significant (15 ms) effect was observed for a different group of irregular verb pairs, that is, pairs that had low form overlap and varied in word length (taught-TEACH vs. taunts-TEACH). As such, for L1 English readers, priming effects for past-tense and present-tense verb pairs have been reliably observed, and such is the case for both regular (Pastizzo and Feldman, 2002; Silva and Clahsen, 2008; Feldman et al., 2010) and irregular verb pairs (Forster et al., 1987; Crepaldi et al., 2010; Feldman et al., 2010). Reliable priming effects observed for past-tense and present-tense word pairs by L1 English readers have been taken to imply that representations of the two words are shared and/or intimately connected in L1 lexicons due to their morphological relationship.

What should be noted at this point is that there is some disagreement among morphological processing models as to how the representations of past- and present-tense forms are shared and/or connected in the L1 lexicon. One point of contention is whether past-tense forms are represented and, hence, processed differently based on their inflectional regularity or not. Some models have proposed that two different cognitive mechanisms are employed for processing past-tense forms (see Pinker and Ullman, 2002). In these models, regular past-tense forms are decomposed *via* the application of morpho-syntactic rules (i.e., verb stem + regular past-tense suffix). Only the verb stem's representation is stored in the lexicon, which is identical to the representation of the present-tense form. For this reason, and consistent with Silva and Clahsen's (2008) results, regular past-tense forms would be expected to prime their present-tense forms just as efficiently as the present-tense forms themselves in masked priming experiments. In contrast, irregular past-tense forms cannot be decomposed by morpho-syntactic rules. Therefore, they are stored in their full form in the lexicon. Morphological priming for irregular past- and present-tense forms is thus a result of the two forms being connected by their formal and morphological/semantic relationship in the lexicon. In contrast to this view of morphological processing, other models require only one mechanism. Some of these models posit explicit representations relating to the morphological structure of words in the lexicon, while others do not (for a review, see Feldman and Weber, 2012; Milin et al., 2018). Despite differences among the models, past-tense forms are assumed to be processed similarly regardless of their inflectional regularity, and morphological priming is explained by shared and/or connected lexical representations for past- and present-tense forms.

In contrast to the consistent results for L1 readers, previous studies produced inconsistent results for L2 readers (in line with the idea that L2 readers might have inexact, or fuzzy, morphological representations). Some experiments showed no morphological priming effect for past-tense verb primes and their present-tense targets in situations, in which an effect has been observed for L1 English readers (e.g., Silva and Clahsen, 2008; Clahsen et al., 2013). Silva and Clahsen (2008), for instance, had both Chinese-English and German-English bilinguals make lexical decisions to regular verb targets (e.g., BOIL) that were preceded by past-tense primes (e.g., boiled), identity primes (e.g., boil), or unrelated primes (e.g., jump). Although their control group of L1 readers showed a significant priming effect for morphologically related pairs (55 ms), which was statistically as strong as the identity priming effect (67 ms), for the two groups of L2 readers, facilitation from past-tense verb primes was absent. Despite the null morphological priming effect, both L2 groups nevertheless showed significant identity priming effects (84 ms and 59 ms for Chinese- and German-English bilinguals, respectively), indicating that those individuals were capable of processing masked L2 primes. The lack of a morphological priming effect for L2 English readers was replicated by Clahsen et al. (2013), in which the same stimulus set used by Silva and Clahsen (2008) was tested with a group of Arabic-English bilinguals. Past-tense primes again failed to facilitate target recognition for bilinguals relative to unrelated primes (e.g., boiled-BOIL = jump-BOIL), although a significant repetition priming effect was again observed for the L2 readers (e.g., boil-BOIL < jump-BOIL). Lack of priming effects from past-tense verb primes in these experiments suggests that at least for regular verbs, there is no underlying connection or shared representation between past-tense verbs and their present-tense forms in the lexicons of L2 readers. In line with the view that past-tense forms are processed differently depending on their regularity, Silva and Clahsen (2008) and Clahsen et al. (2013) have taken their results to suggest that morpho-syntactic processing is less effective for L2 readers than for L1 readers.

Other studies using past- and present-tense pairs with L2 English readers, on the other hand, did find a significant pattern of priming effects that was similar to the one typically observed for L1 English readers (Feldman et al., 2010; Voga et al., 2014), suggesting that the representations and processing of past-tense forms for L2 and L1 readers could be similar. Voga et al. (2014), for example, using the same set of stimuli used by Silva and Clahsen (2008), but with a slightly shorter prime duration (50 ms), found significant priming effects from regular past-tense primes for their Greek-English bilinguals (e.g., boiled-BOIL < jump-BOIL). Furthermore, the past-tense primes facilitated their targets to a degree that was statistically equivalent to that of the identity primes (66 ms and 54 ms effects, respectively). This priming pattern is exactly the priming pattern observed with L1 English readers by Silva and Clahsen (2008).

Some additional support for similar underlying connections among morphologically related word pairs for L1 and L2 English readers (i.e., past-tense inflectional morphological pairs) comes from a study by Feldman et al. (2010), a study that is directly relevant to the present experiments. Feldman et al. tested a group of Serbian-English bilinguals with the same set of regular verb pairs and the two types of irregular verb pairs (e.g., billed-BILL, fell-FALL, taught-TEACH) used with L1 readers by Pastizzo and Feldman (2002). The data from the bilinguals showed that, relative to *orthographic control primes* (billion-BILL), a significant 23 ms priming effect was observed for regular verbs (billed-BILL) although not for the irregular verbs irrespective of the degree of form overlap (fell-FALL = fill-FALL, a 3 ms difference, taught-TEACH = taunts-TEACH, an 11 ms difference). A re-analysis of Pastizzo and Feldman's (2002) L1 data (i.e., 9 items that produced high error rates for Serbian-English bilinguals were removed for a better and more direct comparison of the L1 vs. L2 data), however, did show reliable morphological priming effects for irregular verb types (20 and 19 ms effects), although the effect for regular verbs (42 ms) was somewhat larger.

What is also important to note is that when Feldman et al. (2010) examined priming effects measured against *unrelated primes*, a somewhat different pattern emerged. Specifically, for the L2 readers, past-tense primes produced significant priming effects that were not statistically different across the three verb types (23, 33, and 32 ms effects for fell-FALL, taught-TEACH, and billed-BILL type pairs, respectively). The same priming pattern was also observed for L1 readers; past-tense primes produced significantly faster responses to targets, and there was no interaction with verb type (20, 22, and 30 ms effects). Therefore, in general, Feldman et al.'s results indicate that L2 readers produce a similar result pattern to that of L1 readers. Further, their results suggest that any behavioral difference between L1 and L2 readers in these types of experiments may be due to the differential impact of orthographic primes on the word recognition process.

More specifically, for L2 readers, orthographic similarity almost always facilitates target processing (Nakayama and Lupker, 2018; Jiang, 2021; Kida et al., 2022; also see Qiao and Forster, 2017). For example, Nakayama and Lupker (2018), using a 67 ms prime duration, reported that orthographically related pairs such as time-TILE produced facilitation for Japanese-English bilinguals, even though the same prime-target pairs produced an inhibitory effect for L1 English readers. For L1 readers, the inhibitory effect from orthographically similar primes is assumed to occur through the process of lexical competition among the representations activated by the prime. That is, due to the precision of prime encoding, the prime's lexical representation successfully competes with and inhibits the target's representation (as well as all other activated representations). This competition/inhibition process delays the target's lexical representation from reaching the recognition threshold when it is presented for a lexical decision. In contrast, for L2 readers, facilitation from orthographic relationships can be explained as a consequence of fuzziness, that is, that L2 word forms may be encoded in a less precise manner than L1 forms (e.g., Cook and Gor, 2015). Essentially, for L2 readers, representations of words with similar forms are not easily distinguishable. As a result, many orthographically similar candidates are activated by the masked prime and, equally importantly, remain active at the point that the target is presented because the prime does not prevail in the competition process (see Footnote 1). The result is that orthographically similar prime-target pairs virtually always facilitate L2 word recognition. Consequently, when examining morphological relationships in the L2 lexicon, it is critical that the impact of facilitation due to orthographic similarity be controlled because morphologically related word pairs are typically orthographically similar. That is, for L2 readers, if morphological priming effects are measured by using unrelated primes as the baseline, those effects would be contaminated by the effects of orthographic similarity.

Matching lexical and participant characteristics that could also differentially affect response times between the relevant conditions is also crucial. In studies that examine morphological priming, controlled lexical characteristics typically include the frequency of words, neighborhood density, and the degree of prime-target orthographic overlap (e.g., Crepaldi et al., 2010; Feldman et al., 2010). Further, a general assumption in designing the stimuli for this type of experiment is that these characteristics are similar across all participants. This assumption seems reasonable for L1 readers tested in most studies, who are typically college/university students, as their background in acquiring their L1 is likely to be relatively homogeneous. However, one would expect there to be more variability in such characteristics among L2 readers. Factors such as a participant's age of acquisition of their L2 (e.g., Veríssimo et al., 2018) and L1 background (e.g., Nakayama and Lupker, 2018) could affect how the participant processes words, or at least which words the participant is familiar with and to what degree (e.g., Brysbaert et al., 2017). Research with L1 readers has also shown that certain language skills may be related to how precisely words are represented in the lexicon (e.g., Andrews and Lo, 2012), and it is possible that this conclusion may hold true for L2 readers as well. At present, however, there appears to be no agreed-upon method for gauging language skills in L2 readers. Researchers typically use any one of a number of different L2 proficiency tests to assess or control the L2 language skills of participants, and these tests tend to evaluate different types of language skills (e.g., vocabulary, grammar, and comprehension) for different types of settings (e.g., daily communication, business, and academia). It is, of course, far from clear as to whether the scores from these different tests are reflecting L2 proficiencies in a similar way, making it somewhat difficult to compare the results from morphological priming studies with L2 readers.

Essentially, previous studies focused on the morphological relationship of past-tense verbs and their present-tense forms with L2 readers have not yielded fully consistent results. In the present research, we conducted two masked priming lexical decision experiments with Japanese-English bilinguals in order to provide additional empirical evidence concerning the potential development of a special underlying connection in those readers' lexicons as a result of the morphological relationship between the past- and present-tense forms. Experiment 1 was designed based on Feldman et al.'s (2010) experiment with Serbian-English bilinguals. We followed Feldman et al.'s design because priming effects were assessed relative to both unrelated primes and orthographic control primes in that experiment. Because both bilinguals and L1 readers were tested using the same procedure and stimuli in Feldman et al.'s experiments, we wished to determine how our bilingual results, obtained in a similar experimental setting, would look in reference to their results.[2]

In summary, the purpose of the present research was to investigate whether the representations of morphological relationships found in L1 English readers' lexicons are similar to those found in L2 English readers' lexicons. If a pattern paralleling that shown by L1 English readers is found, then morphological representations are likely encoded in a similar fashion in the mental lexicon of L2 English readers. Specifically, we examined the question of whether representations of inflectional verbal morphology are present in the L2 English lexicon of Japanese-English bilinguals by conducting two masked priming experiments. The setup of the stimuli and the experimental design followed largely those used by Feldman et al. (2010), although a new set of stimuli was selected in order to better suit the breadth of English vocabulary knowledge of our particular bilingual groups. As the results of Experiment 1 were not entirely conclusive with respect to our research question, Experiment 2 was conducted to test the replicability of the main results found in Experiment 1.

EXPERIMENT 1

Method

Participants

A total of 93 Japanese-English bilinguals participated in Experiment 1. Forty-five were recruited from Tohoku University and 48 were recruited from Waseda University. Data collection was conducted in each respective institution. The participants' L1 language was Japanese, and they were reasonably proficient in English (i.e., they all obtained scores equal to or higher than 610 on TOEIC or 530 on TOEFL ITP or Grade 2 on EIKEN; Eiken Foundation of Japan, n.d.).[3] Fifty-one participants were male and 42 were female. The mean age of participants at the time of the experiment was 20.85 (*SD* = 3.22). The age they started learning English was, on average, 9.81 (*SD* = 3.40). The time they had spent in an English-speaking region was, on average, 6.32 months (*SD* = 19.59). Each participant received a 1,000-yen gift card (roughly equivalent to US$9.00) for their participation.

Stimuli

A total of 81 verbs were selected as targets. Following Feldman et al. (2010), there were three types of verb conditions, each involving 27 targets: (1) Regular verbs (REG) were verbs that take the "-ed" ending to form the past tense (e.g., look-looked; dream-dreamed), (2) Irregular Length Preserved verbs (IRLP) were verbs which do not take the "-ed" ending and, therefore, their past tense is formed irregularly; however, their present- and past-tense forms have the same letter length (e.g., fall-fell; sell-sold), and (3) Irregular Length Varied verbs (IRLV) were verbs which do not take the "-ed" ending, their past tense is formed irregularly and their present- and past-tense forms have different letter lengths (e.g., meet-met; pay-paid). Each target verb was paired with three types of primes: a morphological prime that was the past-tense form of its target (e.g., looked-LOOK, fell-FALL, met-MEET), an orthographic prime that was a word that was orthographically similar but was not morphologically or semantically related to its target and was similar in length to the morphological prime (e.g., loose-LOOK, fill-FALL, and men-MEET), an unrelated prime that was a word that was orthographically, morphologically, and semantically unrelated to its target and was exactly the same length as the morphological prime (e.g., master-LOOK, bank-FALL, and lab-MEET). (See **Table 1** for information concerning prime and target characteristics.)

[2]Some previous studies with L1 readers have subdivided irregular verbs into those that are more irregular and less irregular, with results suggesting that inflectional regularity may have gradable effects on the processing of past-tense forms (e.g., Kielar et al., 2008; Kielar and Joanisse, 2010). We did not take this approach, because it requires the use of a much larger number of verbs (with greater variability in their orthographic similarity). We stuck to 2–3 categorization of verbs because we wanted to ensure that L2 readers are well familiar with the past-tense forms of irregular verbs, and this restriction forced us to select small sets of irregular verbs. There presentations of past-test verbs being robust in L2 readers' lexicons is critical because if they are not robust, morphological facilitation would be confounded with orthographic facilitation.

[3]The TOEIC and TOEFL are developed and administered by the Educational Testing Service (ETS) and assess the English abilities of non-native English speakers. The TOEIC test includes listening and reading comprehension questions with content related to daily communication and business. Its test scores range from 10 to 990. The TOEFL ITP is designed for administration at institutions, such as universities. It also includes listening and reading comprehension questions. The TOEFL test focuses on English used in academic settings, and its test scores range from 310 to 677. The EIKEN is a test for English communication administered by the Eiken Foundation of Japan and backed by Japan's Ministry of Education, Culture, Sports, Science, and Technology. It tests general English reading, writing, listening, and speaking skills. It has seven levels, with Grade 1 indicating the highest English proficiency. EIKEN, unlike TOEIC and TOEFL, adapts a pass/fail system. An EIKEN Grade 2 certificate holder is assumed to have the English proficiency level of a high-school graduate (6 years of English learning in an academic setting) or higher. In Experiment 1, of the 90 participants whose data were analyzed, 68 had taken the TOEIC, and their mean score was 798.43 (*SD* = 102.89, range: 485–970). Thirty-five participants had taken the TOEFL ITP, and their mean score was 554.74 (*SD* = 25.33, range: 520–647). Forty-five participants had taken the EIKEN with grades ranging from grade 3 to grade 1. Many participants had scores for multiple English proficiency tests. Individuals were invited to participate in the experiments as long as they satisfied one of the test score criteria.

TABLE 1 | Lexical characteristics and examples of prime-word target pairs used in Experiment 1.

		Prime type			Targets
		MORPH	**ORTH**	**UNREL**	
IRLP		**fell**	**fill**	**bank**	**FALL**
	Frequency	91 (152.2)	69 (138.3)	95 (143.1)	512 (1131.7)
	Length	4.3 (0.5)	4.3 (0.5)	4.3 (0.5)	4.3 (0.5)
	Neighbors	8.0 (4.9)	7.6 (4.1)	7.0 (5.7)	8.8 (4.5)
	% Overlap	60 (20.9)	56 (18.9)	8 (11.4)	
	LD	0.39 (0.19)	0.44 (0.19)	0.92 (0.12)	
IRLV		**paid**	**pair**	**jump**	**PAY**
	Frequency	139 (195.1)	63 (120.8)	104 (147.2)	444 (852.3)
	Length	4.5 (1.1)	4.4 (1.1)	4.5 (1.1)	4.3 (0.9)
	Neighbors	6.7 (4.7)	7.4 (5.3)	6.6 (4.6)	8.6 (5.2)
	% Overlap	54 (27.7)	50 (23.4)	6.0 (10.5)	
	LD	0.54 (0.43)	0.62 (0.41)	1.12 (0.31)	
REG		**looked**	**loose**	**master**	**LOOK**
	Frequency	97 (141.5)	40 (73.8)	97 (165.1)	450 (613.6)
	Length	6.2 (0.8)	5.9 (1.0)	6.2 (0.8)	4.2 (0.8)
	Neighbors	4.4 (2.4)	1.9 (2.3)	4.1 (2.5)	7.9 (4.2)
	% Overlap	67 (4.0)	52 (12.1)	5.0 (8.1)	
	LD	0.49 (0.09)	0.68 (0.26)	1.36 (0.15)	

Values in word frequencies (per million words) and the number of neighbors were according to the English Lexicon Project (Balota et al., 2007). LD refers to the Levenshtein Distance (Levenshtein, 1966).

Because we expected that it would be important to calculate our morphological priming effects based on using the orthographic primes as a control, an effort was made to select orthographic and morphological primes that were equally orthographically similar to their targets. As was done by Feldman et al. (2010), the proportion of letters repeated in the same position between the prime and target was used as a measure of orthographic similarity. We calculated this measure by dividing the number of identical characters in the same letter position between primes and targets by the letter length of the prime and then multiplying it by 100. Therefore, a value of 100 means that the prime and target are exactly the same, whereas a value of 0 means that not a single letter is shared in the same position between the prime and its target. In **Table 1**, we also report the Levenshtein distances (Levenshtein, 1966) between prime-target pairs as an additional reference of orthographic similarity. To take the differential letter lengths into account (i.e., IRLV and REG conditions), we calculated the normalized Levenshtein distance, where the distance between the prime and target was divided by the number of letters in the longer word stimulus. Hence, the value varies from 0 to 1 with smaller values indicating a higher degree of similarity.

REG, IRLP, and IRLV verb targets were matched in their mean word frequencies, word lengths, and numbers of neighbors (Coltheart et al., 1977), all Fs < 1. For primes, strict matches were difficult to achieve on some lexical characteristics in certain conditions, mainly due to the fact that the number of irregular verbs is relatively small in English. Further, because late-bilinguals would know a smaller number of words in English than L1 English readers, our stimulus selection had to be even more restrictive. An effort was made, however, to match the lexical characteristics of the primes as much as possible.

For the primes paired with REG targets, repeated measure ANOVAs confirmed that the morphological, orthographic, and unrelated primes were matched on their mean word frequencies [Ms = 97, 40, 97, respectively, $F(2, 52) = 1.75$, $p > 0.18$] and word lengths [Ms = 6.2., 5.9., and 6.2., $F(2, 52) = 2.75$, $p > 0.07$]. Despite our best efforts, the prime-target orthographic similarities were higher for morphological primes ($M = 67\%$, e.g., looked-LOOK) than for orthographic primes [$M = 52\%$, e.g., loose-LOOK, $t(26) = 6.06$, $p < 0.001$]. Unrelated primes had significantly lower prime-target orthographic similarity than both morphological and orthographic primes ($M = 5\%$, master-LOOK, ps < 0.001). Lastly, morphological and unrelated primes had a statistically equivalent number of neighbors [Ns = 4.4 and 4.1, $t(26) = 1.0$, $p = 0.33$]; however, orthographic primes had a significantly lower number of neighbors ($N = 1.9$, ps < 0.001), due to the fact that matching on prime-target orthographic similarity was made a priority over the primes' neighborhood sizes.

For the primes paired with IRLP targets, morphological, orthographic, and unrelated primes were statistically matched on their mean word frequencies (Ms = 91, 69, 95, $F < 1$), word lengths (Ms = all 4.3), and neighborhood sizes (Ms = 8.0, 7.6, 7.0, $F < 1$). The prime-target orthographic similarity was matched between morphological primes ($M = 60\%$, e.g., fell-FALL) and orthographic primes ($M = 56\%$, e.g., fill-FALL), $t(26) = 1.28$, $p > 0.21$. Unrelated primes had a significantly lower prime-target orthographic similarity than both morphological and orthographic primes ($M = 8\%$, e.g., bank-FALL, ps < 0.001).

Finally, for the primes paired with IRLV targets, morphological, orthographic, and unrelated primes were statistically matched on their mean letter lengths [$M = 4.5$, 4.4, 4.5, $F(2, 52) = 2.85$, $p > 0.06$] and mean numbers of neighbors ($M = 6.7$, 7.4, 6.6, $F < 1$). The prime-target orthographic similarity was matched between morphological primes and their targets ($M = 54\%$, e.g., fell-FALL) and orthographic primes and those same targets ($M = 50\%$, e.g., fill-FALL, $t < 1$). Unrelated primes had significantly lower prime-target orthographic similarity than both morphological and orthographic primes ($M = 6\%$, ps < 0.001). A variable that we could not statistically match in this condition was the mean word frequencies; morphological primes had statistically higher mean word frequency ($M = 139$) than both orthographic primes ($M = 63$) and unrelated primes ($M = 104$), ps < 0.05, which were not statistically different, $t(26) = -1.55$, $p > 0.10$. The fact that the orthographic primes were lower-frequency was not likely problematic, as Nakayama and Lupker (2018) showed that the facilitation effect from orthographically similar primes for Japanese-English bilinguals is not affected by whether they are words or non-words (non-words have a frequency of 0). For word targets, three presentation lists (List A, List B, and List C) were created in such a way that within a list, a third of the word targets were primed by the morphological primes, a third by the orthographic primes, and a third by the unrelated primes. Across the lists, each word target was primed by each of the three prime types equally frequently.

A total of 81 non-word targets were also selected for "NO" responses in a lexical decision task. The non-word targets consisted of three sets of 27 non-words which served as

TABLE 2 | Mean response latencies and (error rates) of targets primed by morphological, orthographic, and unrelated primes for Experiment 1.

Verb type	Prime type			Priming effect	
	MORPH (M)	ORTH (O)	UNREL (U)	O-M	UR-O
IRLP	606 (9.6)	612 (11.4)	647 (10.9)	6 (1.8)	35 (−0.5)
IRLV	596 (6.3)	600 (8.0)	622 (8.3)	4 (1.7)	22 (0.2)
REG	590 (8.5)	604 (7.2)	633 (8.2)	14 (−1.4)	29 (1.0)

IRLP = Irregular Length Preserved Verbs; IRLV = Irregular Length Varied Verbs; REG = Regular Verbs.

counterparts to the REG, IRLP, and IRLV verb targets. Within each set of non-word targets, a third of the targets ($n = 9$) were primed by words that mimicked the relationship between the morphological prime-target word pairs (e.g., father-FATH, slam-SLOG, ticket-TIVE). A third of the targets were primed by words that were orthographically similar to their targets (e.g., carbon-CARN, box-BOP, and nag-NAGE). A third of the targets were primed by words that were orthographically and phonologically unrelated (e.g., corner-TOAK, carry-PONER, and team-TATCH). As non-words do not have morphological representations, the implication is that within the 81 non-words, two-thirds of the targets were primed by orthographically (and also phonologically) similar word primes, and one-third by unrelated word primes. Lexical characteristics of the primes (e.g., mean word frequencies, lengths, numbers of neighbors, and orthographic similarity) were similar to their counterparts in the word target conditions. The lexical characteristics for the stimuli in the non-word target conditions are available in the **Supplementary Material**. Prime Type was not manipulated for non-words, and, therefore, there was only one presentation list for non-word targets. None of the word primes preceding non-word targets was used as a critical stimulus (i.e., in the word prime-target pairs).

Apparatus and Procedure

The presentations of the stimuli and the recording of responses were controlled by DMDX (Forster and Forster, 2003). Participants were tested individually in a quiet room. The presentation sequence of a trial was identical to that of Feldman et al. (2010) and was as follows: a fixation point (i.e., "+") was first presented for 450 ms, which was followed by a 50 ms blank screen. Then, a string of number signs (i.e., "#"), matching the letter length of the prime, was presented for 500 ms as a forward mask. Immediately after the presentation of the forward mask, a prime was presented for 50 ms in lower-case letters, which was immediately replaced by a target in upper-case letters. Targets remained on the screen for 3,000 ms or until a response was made. The inter-trial interval was 1,000 ms. The stimuli were presented in 18 pt. Courier New font at the center of the display.

Participants were asked to decide whether each target stimulus is a real English word or not and indicate their decision by pressing the "YES" or "NO" button on a game pad (Tohoku University) or on a response box (Waseda University) as fast and accurately as possible. Prior to the presentation of the experimental trials, 36 practice trials were presented in order to familiarize participants with the task. Participants were asked to repeat the practice session until they felt comfortable with the task. The presentation lists were counterbalanced across participants and the order of trials within each list was randomized for each participant. Approval for the experiments was obtained from the ethics review board of the Graduate School of International Cultural Studies Tohoku University, and the ethics review committee on research with human subjects of Waseda University.

Results

Data from two participants were removed because they made more than 25% errors (one participant each from List A and List C). To equate the numbers of participants between the presentation lists, data from one additional participant (the last participant from List B) were removed. As a result, data from 90 participants were analyzed. Responses with latencies greater than 1,500 ms were considered to be outliers (0.32% of word data) and were also removed from the entire analyses.

In the analysis of the response latencies, we analyzed the data with linear mixed effect (LME) models (e.g., Baayen et al., 2008) using the lme4 package (Version 1.1–27.1, Bates et al., 2015) available in R (Version 4.1.1, R Core Team, 2021). For the response latency analyses, a reciprocal inverse transformation was applied to the raw RTs (i.e., −1,000/RT; hereafter, *invRT*) to meet the assumption of normality. In order to calculate the *p*-values with the degrees of freedom based on Satterthwaite's approximation, we used the anova function of the lmerTest package (Version 3.1–3, Kuznetsova et al., 2017). The model used was *invRT* ~ Verb Type*Prime Type + (1|Participants) + (1|Targets). In addition, *post-hoc* comparisons were carried out using the emmeans package (Version 1.7.2, Lenth, 2022) with Tukey's HSD adjustments when necessary. The error analysis was conducted with the same procedure except that we used a generalized linear mixed-effect model, assuming a binomial distribution, and the anova function in the car package (Version 3.0–12, Fox and Weisberg, 2019) was used to obtain the *p* values for the fixed effects. The model used was Error ~ Verb Type*Prime Type + (1|Participants) + (1|Targets). Mean response latencies and error rates of Experiment 1 are shown in **Table 2**.

Response Latencies

The main effect of Prime Type was significant, $F(2, 6458.8) = 91.29$, $p < 0.001$. The main effect of Verb Type was not significant, $F < 1$. The interaction between Verb Type and Prime Type was not significant, $F(4, 6458.9) = 1.22$, $p = 0.30$, meaning that the patterns of priming effects were not different for REG, IRLP, and IRLV targets.

Follow-up analyses of the significant main effect of Prime Type revealed that across Verb Type, targets primed by morphological primes were responded to significantly faster than the same targets primed by unrelated primes, *estimated coef.* = −0.112, $SE = 0.00861$, $t = -13.03$, $p < 0.001$. Targets primed by orthographic primes were also responded to significantly faster than the same targets primed by unrelated primes, *estimated coef.* = −0.083, $SE = 0.00863$, $t = -9.65$, $p < 0.001$. Critically, targets primed by morphological primes were responded to significantly faster than targets primed by orthographic primes, *estimated coef.* = −0.029, $SE = 0.00860$, $t = -3.36$, $p = 0.02$.

Error Rates

No effects were significant, all $ps > 0.25$.

Discussion

In Experiment 1, targets primed by orthographic primes were responded to significantly faster than targets primed by unrelated primes. This effect replicated the results of Nakayama and Lupker (2018), in which orthographically similar English word primes significantly facilitated lexical decision latencies to English targets for Japanese-English bilinguals and is also in line with what may be expected when encoding of orthographic form is fuzzy (e.g., time-TIDE < doll-TIDE). Although similar results have been found in L2 morphological priming experiments when the readers' L1 was alphabetic (e.g., Diependaele et al., 2011), this result does contrast sharply with findings observed in orthographic neighbor priming experiments for L1 readers, where the direction of the effect is typically inhibitory (e.g., time-TIDE > doll-TIDE; Segui and Grainger, 1990; Davis and Lupker, 2006; Nakayama et al., 2008).

In Experiment 1, targets were also responded to significantly faster when they were primed by morphological primes than by unrelated primes. This facilitation from morphological primes can be orthographic, not necessarily morphological, in origin, because as was observed, orthographic similarity can facilitate bilinguals' L2 lexical decision latencies. Nevertheless, in Experiment 1, the *post-hoc* analysis showed that across target verb types, the size of the priming effect was significantly larger from morphological primes than from orthographic primes, although the difference was numerically small (7 ms). This additional facilitation for morphological prime-target pairs over orthographic prime-target pairs seems to suggest that representations reflecting morphological relationships *do develop in the L2 English lexicons of Japanese-English bilinguals*. However, we need to point out that although there was no Prime Type by Verb Type interaction, suggesting that the priming advantage for morphological over orthographic primes was not different for REG, IRLP, and IRLV verbs, REG targets produced a larger numerical advantage (14 ms) than the other two verb types (6 ms and 4 ms, respectively). This pattern is a bit difficult to interpret because in the REG condition, prime-target orthographic similarity was higher for morphological primes (67%) than for orthographic primes (52%). Thus, in the REG condition, the priming effect from morphological primes could involve additional facilitation due to orthographic similarity, meaning that the present experiment might overestimate the size of the morphological priming effect in that condition.

On the other hand, prime-target orthographic similarity between morphological and orthographic primes was matched in the IRLP and IRLV conditions and, therefore, any priming advantage for morphological primes in those conditions, would make a strong case for the impact of morphology. When the data in the IRLP and IRLV conditions were analyzed (removing the data from the REG condition), however, the priming advantage for morphological over orthographic primes was not quite statistically significant, $F(1, 2801.7) = 3.66$, $p = 0.056$. Although the morphological priming advantage over orthographic priming was nevertheless significant when data in the REG condition alone were analyzed, $F(1, 1375.2) = 9.35$, $p < 0.01$, as noted above, this difference could partly be due to morphological primes having orthographic similarity. Thus, although there was an indication that morphological level representations in L2 English lexicons may develop for Japanese-English bilinguals, the evidence is not robust. Therefore, Experiment 2 was an effort to investigate this issue further.

EXPERIMENT 2

One potential problem with the design of Experiment 1, which may have led to the somewhat ambiguous results, was that the number of the items selected was relatively small. In Experiment 1, for each of the three verb type conditions, 27 items were primed by three types of primes. Therefore, there were only 9 items per cell for any given participant. Although we attempted to address this problem by testing a large number of participants ($N = 90$), our results may, unfortunately, not have been as stable as we might have wished. Therefore, in Experiment 2, we selected a larger set of items as critical stimuli. In order to allow us to increase the item numbers, we dropped the IRLV condition. There are not many IRLV verbs that our bilinguals would be familiar with, and, thus, the inclusion of this condition in Experiment 1 made it difficult for us to have a large number of stimuli in the various conditions. Because we were not directly interested in investigating the effects of word length (varied or preserved) on the development of morphologically-based connections between past-tense verbs and their present-tense forms, removing the IRLV condition does not impede our research goal. Thus, in Experiment 2, only two verb type conditions were examined: regular verbs (REG; $n = 48$, e.g., look-LOOKED) and irregular verbs (IREG; $n = 48$, e.g., fell-FALL). We should acknowledge that, although equating orthographic similarity between morphological prime-target pairs and orthographic prime-target pairs was optimized, it was still not possible to fully equate the values for regular prime-target pairs given the limited vocabulary sizes of our bilinguals. To compensate, we conducted a *post-hoc* regression analysis to ascertain if greater orthographic overlap affected priming effects from morphological primes in the REG condition. In the IREG condition, orthographic similarity was fully matched between the morphological and orthographic prime-target pairs. Therefore, any priming effect observed relative to orthographic primes in this condition can be attributed to the prime and target's morphological relationship.

Method

Participants

Participants were 84 Japanese-English bilinguals recruited at Tohoku University ($n = 44$) and Waseda University ($n = 40$).

They spoke Japanese as their first language and were reasonably proficient in English (they all had a TOEIC score of 605 or a TOEFL ITP score of 510 or higher).[4] Thirty-eight of the participants were male and 46 were female. The age of participants (excluding one who did not report his/her age) at the time of the experiment was 21.65 ($SD = 3.20$). The age they started learning English was, on average, $M = 9.74$, ($SD = 3.37$). The time they had spent in an English-speaking region was, on average, 3.83 months ($SD = 15.78$). Participants each received a 1,000-yen gift card (roughly equivalent to US$9.00) for their participation.

Stimuli

The targets consisted of two types of verbs (Verb Type): irregular verbs (IREG) and regular verbs (REG). Irregular verbs (IREG, $n = 48$) were verbs which do not take the "-ed" ending in their past tense (i.e., their past tense is formed irregularly) and their past- and present-tense forms have the same letter length (e.g., fell-FALL). Regular verbs (REG, $n = 48$) were verbs which take the "-ed" ending to form the past tense (e.g., looked-LOOK). Each target was paired with three types of primes: morphological, orthographic, and unrelated primes. Examples and the lexical characteristics of the word targets are shown in **Table 3**.

Targets in the IREG and REG conditions were matched on their mean word frequencies, word lengths, and numbers of neighbors (all $ts < 1$). For primes paired with IREG targets, the morphological, orthographic, and unrelated primes were matched on their mean word frequencies ($M = 105$, 119, 97, $F < 1$), word lengths ($Ms =$ all 4.2), and number of neighbors ($Ms = 8.9$, 9.2, 8.4, $F < 1$). Prime-target orthographic overlap was statistically equivalent for morphological primes and orthographic primes ($Ms = 64$ and 66%), $t < 1$. Unrelated primes had no orthographic overlap with their targets ($M = 0\%$).

For the primes paired with REG targets, morphological, orthographic, and unrelated primes were matched in their mean word frequencies ($Ms = 76$, 61, 73, $F < 1$) and word lengths ($Ms =$ all 6.2, as all primes for a given target had the same length). As was the case in Experiment 1, prime-target orthographic overlap was inevitably significantly higher for morphological primes ($M = 67\%$) than for orthographic primes ($M = 46\%$), $t(47) = 13.03$, $p < 0.001$, $SEM = 1.62$. Unrelated primes had no orthographic overlap with their targets ($M = 0\%$). Morphological primes also had a statistically higher number of neighbors ($N = 5.3$) than orthographic ($N = 2.2$, $t(47) = 6.92$, $p < 0.001$, $SEM = 0.46$) or unrelated primes ($N = 2.1$, $t(47) = 6.44$, $p < 0.001$, $SEM = 0.50$), which did not differ from one another ($t < 1$).

A total of 96 non-word targets were also selected for "NO" responses. More than half of the non-words were generated with Wuggy (Keuleers and Brysbaert, 2010). The non-word targets consisted of two sets of 48 non-words which served as counterparts to the REG and the IREG verb targets. The non-word targets had similar mean word lengths and numbers of neighbors as those of the word targets. Non-words were paired with word primes in the same way as in Experiment 1. Lexical characteristics of the primes (e.g., mean word frequencies, lengths, numbers of neighbors, and orthographic similarity) were similar to those of their counterparts in the word target condition. The lexical characteristics for the non-word target condition are available in the **Supplementary Material**. None of the word primes preceding non-word targets were used in the critical stimuli (e.g., word prime-target pairs).

TABLE 3 | Lexical characteristics and examples of prime-target pairs used in Experiment 2.

		Prime type			Targets
		MORPH	**ORTH**	**UNREL**	
IREG		**fell**	**fill**	**joke**	**FALL**
	Frequency	105 (177.4)	120 (343.4)	97 (160.6)	371 (668.1)
	Length	4.2 (0.7)	4.2 (0.7)	4.2 (0.7)	4.2 (0.7)
	Neighbors	8.9 (5.3)	9.2 (5.6)	8.4 (4.5)	8.9 (4.5)
	% Overlap	64 (18.8)	66 (11.7)	0.0 (0.0)	
	LD	0.35 (0.17)	0.34 (0.12)	1.0 (0.00)	
REG		**looked**	**locker**	**rather**	**LOOK**
	Frequency	76 (112.2)	61 (234.3)	73 (102.3)	360 (545.7)
	Length	6.2 (0.6)	6.2 (0.6)	6.2 (0.6)	4.2 (0.6)
	Neighbors	5.3 (2.9)	2.2 (2.4)	2.2 (2.1)	9.3 (4.7)
	% Overlap	67 (2.8)	46 (10.2)	0.0 (0.0)	
	LD	0.49 (0.06)	0.78 (0.15)	1.47 (0.10)	

Values in word frequencies (per million words) and the number of neighbors were according to the English Lexicon Project (Balota et al., 2007). LD refers to the Levenshtein Distance (Levenshtein, 1966).

Apparatus and Procedure

The apparatus and procedure of Experiment 2 were identical to those in Experiment 1.

Results

Data from six participants were removed due to high error rates (25% or more). Data from one additional participant were removed due to noncompliance with the instructions. Data for these participants were replaced by those from additional participants while maintaining the counterbalancing of the lists. Responses with latencies greater than 1,500 ms were considered as outliers and were removed from the analyses (0.69% of word data). Six items (4 REG verbs, "lick," "jail," "sail," and "bust" and one IREG verb: "stink," as well as one verb in the IREG condition which also has a regular ending: "lie" which has past-tense forms of both "lay" and "lied" depending on which meaning of "lie" is intended) were also excluded from the entire analyses because those words produced more than 40% error rates. The remaining response latencies and error rates were analyzed as in Experiment 1 with LME and GLM models, respectively, except that the categorical factor Verb Type now only had two levels (REG, IREG). Mean response latencies and error rates of Experiment 2 are shown in **Table 4**.

[4] Of the 84 participants whose data were analyzed, 66 had taken the TOEIC. Their TOEIC mean score was 790.76 ($SD = 93.33$, range: 605–965). Thirty-six participants had taken the TOEFL ITP. Their mean score was 552.00 ($SD = 35.60$, range: 485–650).

TABLE 4 | Mean response latencies and (error rates) of targets primed by morphological, orthographic, and unrelated primes for Experiment 2.

Verb type	Prime type			Priming effect	
	MORPH (M)	ORTH (O)	UNREL (U)	O-M	UR-O
IREG	634 (9.2)	655 (11.9)	676 (12.1)	21 (2.7)	21 (0.2)
REG	627 (8.2)	656 (11.0)	675 (12.0)	29 (2.8)	19 (1.0)

IREG = Irregular verbs with the same character length between past and present forms; REG = Regular verbs.

Response Latencies

The main effect of Prime Type was significant, $F(2, 6525.7) = 98.27$, $p < 0.001$. The main effect of Verb Type was not significant, $F < 1$. The interaction between Verb Type and Prime Type was also not significant, $F(2, 6525.8) = 1.35$, $p > 0.25$. Therefore, as was the case in Experiment 1, patterns of priming effects were not significantly different for regular and irregular targets.

Follow-up analyses of the significant main effect of Prime Type revealed that across Verb Type, targets primed by morphological primes were responded to significantly faster than targets primed by unrelated primes, *estimated coef.* = −0.114, $SE = 0.00815$, $t = -14.02$, $p < 0.001$. Targets primed by orthographic primes were also responded to significantly faster than targets primed by unrelated primes, *estimated coef.* = −0.059, $SE = 0.00821$, $t = -7.14$, $p < 0.001$. Consistent with Experiment 1, targets primed by morphological primes were responded to significantly faster than targets primed by orthographic primes, *estimated coef.* = −0.056, $SE = 0.00813$, $t = -6.84$, $p < 0.001$. Although the interaction was not significant, we further analyzed the patterns of morphological priming effects separately for REG and IREG conditions. The results showed that in the IREG condition, targets primed by morphological primes were responded to 21 ms faster than targets primed by orthographic primes, and this advantage in processing speed was statistically significant, $F(1, 2154.8) = 12.02$, $p < 0.001$. In the REG condition, there was a significant 29 ms processing advantage for targets primed by morphological compared to orthographic primes, $F(1, 2091.4) = 39.10$, $p < 0.001$.

Error Rates

The only significant effect was the main effect of Prime Type, $X^2 = 8.10$, $p = 0.017$. As expected, across Verb Type, error rates were significantly smaller for targets primed by morphological primes ($M = 8.7\%$) than by unrelated primes ($M = 11.9\%$), *estimated coef.* = − 0.419, $SE = 0.0985$, $z = -4.26$, $p < 0.001$. Error rates were also significantly smaller for targets primed by morphologically related primes than targets primed by orthographic primes ($M = 11.4\%$), *estimated coef.* = − 0.343, $SE = 0.0993$, $z = -3.45$, $p < 0.01$. Error rates were not statistically different for targets primed by orthographic primes versus unrelated primes, $z < 1$. These patterns did not interact with Verb Type, $X^2 < 1$, $p > 0.75$. Thus, paralleling the RT data, there was a significant morphological priming effect relative to both orthographic and unrelated primes. Separate analyses for the IREG and REG conditions also confirmed that morphological priming effects, measured against orthographic primes, were significant in both the IREG condition, $X^2 = 5.99$, $p = 0.014$, and the REG condition, $X^2 = 6.08$, $p < 0.014$.[5]

Discussion

In Experiment 2, morphological primes facilitated lexical decisions to targets more than orthographic primes did. Importantly, this pattern did not appear to vary by verb type as indicated by the non-significant interaction. The results in Experiment 2, therefore, successfully replicated those in Experiment 1.

In the stimulus selection of Experiment 2, our priority was that our participants were reasonably familiar with English stimuli, especially the masked primes. As a result of this constraint, in the REG condition, morphological primes were more orthographically similar to their targets than the orthographic control primes were to their targets (67 and 46%, respectively). We were able to equate prime-target orthographic similarity between morphological and orthographic primes (64 and 66%, respectively) in the IREG condition, however. The difference in the orthographic overlap in the REG condition may be at least part of the reason why the morphological priming effect was larger than the orthographic priming effect in that condition. Thus, we decided to re-examine this issue. For individual targets, the size of the morphological priming effect (calculated from raw RTs) was regressed against the difference in the orthographic overlap between M-UR vs. O-UR pairs in the REG condition. Crucially, the size of the additional orthographic overlap had no association with the size of morphological priming effect observed in the REG condition ($t < 1$, $\beta = 0.12$, *n.s.*). Together with the fact that Verb Type did not significantly interact with the pattern of morphological priming effects, it seems safe to conclude that the morphological priming effect observed in the REG condition, when measured against orthographic primes, reflected mainly the impact of the primes' and targets' morphological similarity.

GENERAL DISCUSSION

In the present research, we investigated whether representations of past-tense verbs and their present-tense forms in the lexicons of L2 readers are organized in the same way as is assumed to be the case in the lexicons of English L1 readers. The results of the two experiments suggest that connections based on the morphological relationships of past- and present-tense verbs in the L2 English lexicon are similar to those of L1 English readers, at least in the population of Japanese-English bilinguals examined in the present research. Although the sensitivity of L2 readers to the morphological structure of complex words might be prone to fuzziness, our data seem to indicate that the morphological representations in our participants' lexicons were not organized in an imprecise manner (at least with respect to the representations of past-tense verbs and their present-tense forms). In Experiment 1, past-tense verb primes facilitated lexical decisions to their present-tense targets compared to both an orthographic and an

[5]We removed the random intercept for subjects in the error analyses as the model did not converge when it was included.

unrelated baseline. The results of Experiment 1 were, however, not entirely clear because it was difficult to dissociate the impact of prime-target orthographic relationships from those of morphological relationships. Nevertheless, the results of Experiment 2, which had a better controlled and larger set of stimuli, replicated the general data pattern of Experiment 1. That is, past-tense primes facilitated responses to their present-tense targets when compared to an orthographic baseline and such was the case for both regular and irregular verbs. Therefore, the way English past-tense verbs and their present-tense forms are connected in the lexicons of Japanese-English bilinguals appears to be reasonably similar to the way they are connected in the lexicons of L1 English readers. Fuzziness in encoding L2 word forms, manifesting itself as facilitory priming effects from orthographically related primes, did not lead to differences in the way that past-tense verbs and their present-tense forms were connected.

We should note, however, that the exact nature of such connections in L1 readers, especially when considering the representations of regular versus irregular verbs, is not yet agreed upon. Some researchers believe that the past- and present-tense forms of regular verbs and irregular verbs are connected in qualitatively different ways. One common view is that regular verbs share the same underlying representation through the common root (i.e., boiled - boil *via* the shared root "boil"), while irregular verbs have separate underlying representations with tight connections *via* semantic and formal association (i.e., "fell" and "fall" have separate and independent representations). Other researchers believe that there is no qualitative difference in how regular and irregular verbs are represented. One variant of the latter view further assumes that both types of past-tense and present-tense verbs are connected *via* their semantic and formal relationships with some models not assuming the presence of explicit morphological level representations. Another variant assumes that both types of past-tense and present-tense verbs have shared representations *via* the morphological root or lemma. Discussion of exactly how past-tense and present-tense verbs are connected in English L1 lexicons is beyond the scope of the present research (please see Feldman and Weber, 2012; Milin et al., 2018, for reviews). However, the main point here is that, whatever account of L1 connections is accurate, the similar patterns of morphological priming observed for regular and irregular verb targets in our experiments provide no evidence in support of a different morpho-syntactic process for L2 readers than for L1 readers (*cf.*, Silva and Clahsen, 2008; Clahsen et al., 2013). In that sense, our results are in line with the view that seems to be more widely accepted in the current L1 literature, that the two types of past-tense and present-tense verbs are represented similarly, with their word forms being connected in a way that goes beyond just a sum of form and semantic similarity (e.g., Stockall and Marantz, 2006; Kielar et al., 2008; Crepaldi et al., 2010; Morris and Stockall, 2012; Fruchter et al., 2013).

Effects of L2 Proficiency on Morphological and Orthographic Priming Effects

According to the fuzzy lexicon hypothesis, L2 lexical representations do not inevitably remain fuzzy but rather become more robust with more experience with the L2 input (which would lead to a higher lexical quality of words). In Experiment 2, we recruited a large number of participants who encountered a larger set of stimuli per condition than in Experiment 1. To gain some additional understanding of the development of the L2 lexicon, we examined how L2 proficiency affected the patterns of morphological priming effects for our Japanese-English participants. In these *post-hoc* analyses of L2 proficiency, TOEIC scores were used. The analyses were based on 66 participants who had their TOEIC scores available (79% of the participants in Experiment 2). Their mean TOEIC score was 790.76 ($SD = 93.33$, range: 605–965).

In the L2 proficiency analyses, TOEIC scores were first converted to z scores and entered as a factor. The model used was *invRT* ~ Verb Type*Prime Type*zTOEIC score + (1|subject) + (1|target). The results showed no main effect of L2 Proficiency (i.e., zTOEIC score), $F(1, 64.1) = 1.11$, $p = 0.30$, indicating that overall responding speed to targets was not modulated by TOEIC scores. L2 proficiency, however, significantly interacted with Prime Type, $F(2, 5152.0) = 3.02$, $p = 0.049$. L2 Proficiency did not interact with Verb Type, $F < 1$. The three-way interaction between Proficiency, Verb Type, and Prime type was also not significant, $F < 1$. **Figure 1** shows the pattern of the two-way interaction observed in the L2 proficiency analyses.[6]

Because the main issue here is the impact of L2 proficiency on morphological priming, we first investigated the difference between the morphological and orthographic priming conditions. L2 proficiency did not significantly interact with the pattern of morphological priming, $F(1, 3456.3) = 1.64$, $p = 0.20$. However, as can be seen in **Figure 1**, L2 proficiency appears to be affecting morphological priming effects quite differently at the two extreme levels of proficiency, with the effect being considerably smaller for less proficient bilinguals. To further assess the patterns of morphological priming effects, we statistically tested the priming effects for the most and the least proficient bilinguals, using the *emmeans* package (Version 1.7.2, Lenth, 2022). For participants with the highest L2 proficiency ($z = 1.87$; the raw TOEIC score = 965), the morphological priming effect defined this way was significant, *estimated coef.* = −0.072, $SE = 0.0207$, $t = -3.46$, $p < 0.001$. For participants with the lowest L2 proficiency ($z = -2.01$; the raw TOEIC score = 605), however, there was no clear sign of a morphological priming effect, $t = -1.05$, $p = 0.30$.

Next, we considered the impact of L2 proficiency on the orthographic priming effects (i.e., the difference between the orthographic and unrelated conditions). L2 proficiency significantly affected this priming pattern, $F(1, 3376.9) = 5.78$, $p = 0.016$. The modulation emerged because orthographic priming became smaller as L2 proficiency increased. When the priming patterns for the most and least proficient bilinguals were

[6]We also analyzed the effects of the age of acquisition (AoA) and the time they had spent in an English-speaking region (length of residence; LOR) using the LME models. For AoA, there was a significant main effect of AoA, with faster AoA leading to faster responses to targets, $F(1, 82) = 4.04$, $p = 0.048$. However, importantly, AoA did not interact with any of the priming effects, $F(2, 6{,}522{,}1) = 1.73$, $p = 0.18$. The three-way interaction between AoA, Prime Type, and Verb Type was also not significant, $F(2, 6522.2) = 1.04$, $p = 0.35$. For LOR, no effects relating to priming effects were significant, all $Fs < 1$.

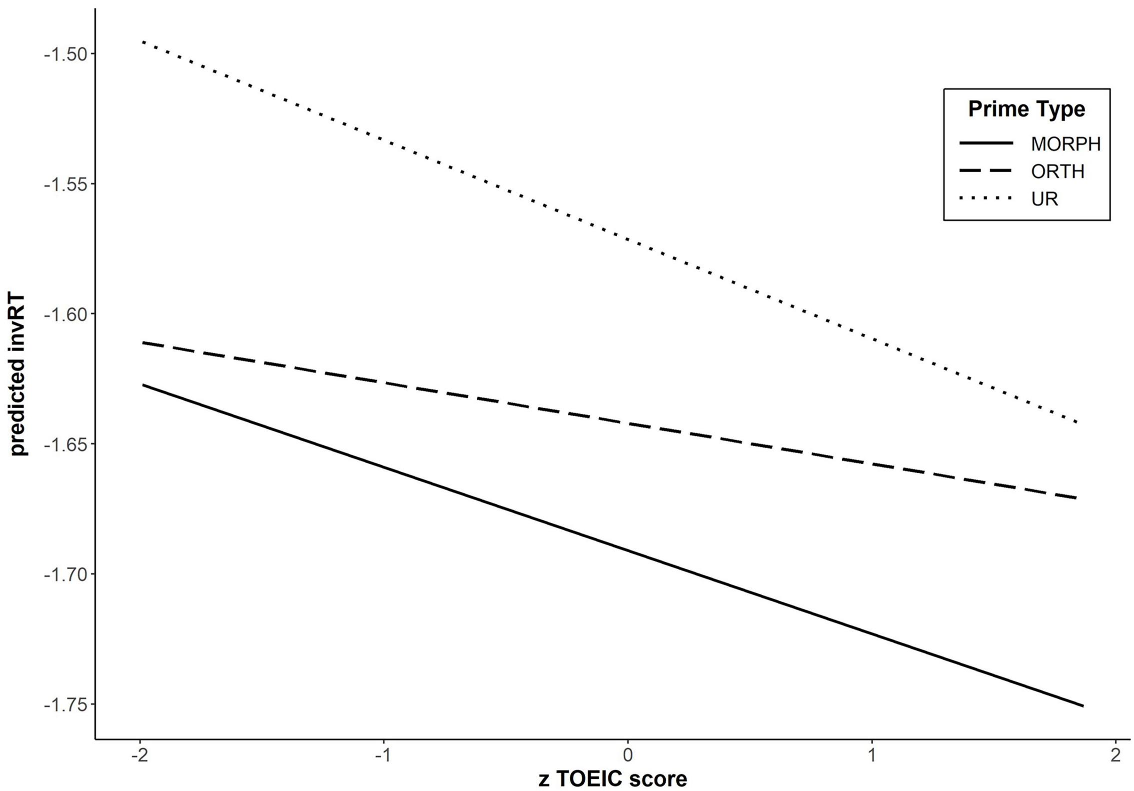


FIGURE 1 | Response latencies to L2 targets primed by morphological, orthographic, and unrelated primes and priming effects as a function of L2 proficiency in Experiment 2. Greater zTOEIC score indicates higher proficiency, and smaller invRT indicates faster responses to L2 targets.

separately analyzed, the most proficient bilinguals did not show a significant orthographic priming effect, *estimated coef.* = −0.026, $SE = 0.0208$, $t = -1.27$, $p = 0.20$. For the least proficient bilinguals, on the other hand, there was a significant orthographic priming effect, *estimated coef.* = −0.118, $SE = 0.0218$, $t = -5.41$, $p < 0.001$.

With regard to the pattern of orthographic priming effects, the observed interaction is a result that was different from that of Nakayama and Lupker (2018) who found no modulation by L2 proficiency (TOEIC score) on the magnitude of orthographic priming effects. We are not entirely sure why L2 proficiency differently affected the patterns of orthographic priming effects here. We could, however, speculate that the presence of morphologically related pairs, which were not presented in Nakayama and Lupker (2018), may have somehow affected how high proficient bilinguals deal with orthographic similarity. Although inconsistent with the results of Nakayama and Lupker, numerically, our data pattern does converge with Feldman et al.'s (2010) Serbian-English bilinguals' data in that facilitation from orthographic primes was smaller for their more proficient bilinguals (12 ms) than for their less proficient bilinguals (26 ms). This pattern could be accounted for by assuming that there is a high degree of fuzziness in the low-proficient bilinguals' orthographic lexicon which might then lead to weaker competition between lexical items in this group (thereby increasing facilitation from orthographically similar word primes).

Finally, L2 proficiency did not significantly modulate morphological priming effects relative to unrelated primes, $F(1, 3444.8) = 1.23$, $p = 0.27$. As shown in **Figure 1**, there was no significant difference in the sizes of the morphological priming effect between the most and least proficient bilinguals. Given the fact that the orthographic priming effect was significant only for the least proficient bilinguals, the apparent morphological priming effect for less proficient bilinguals seems to be driven mostly by the prime-target orthographic similarity.

As such, the overall picture is that, although L2 proficiency did not modulate the patterns of morphological priming effects when measured using an unrelated baseline, L2 proficiency affected the pattern of orthographic priming effects and, hence, the degree to which orthographic priming effects contribute to the overall morphological priming effects. A reasonable account of this difference is that for those with a higher level of L2 visual word recognition skills, lexical competition starts to operate in the recognition of L2 words (orthographic representations become less fuzzy). Therefore, orthographic similarity no longer is an effective source of facilitation from word primes. The *post-hoc* analyses suggest that weak lexical competition, a characteristic of the L2 lexicon, will likely strengthen with greater L2 exposure, at least when an experimental input is visual (as opposed to, for example, the auditory domain, in which representations of some difficult-to-distinguish phoneme contrasts may not inevitably become of high quality due to greater amount of exposure). In contrast, with greater proficiency, the morphological relationships become more stable in the L2 lexicon. As a result, priming based on those relationships rather than orthographic relationships becomes more potent.

Orthographic and Morphological Level Representation in the L2 Lexicon

In the early stages of visual word recognition, activation in the lower-level representations is assumed to flow to upper-level representations *via* feed-forward connections. An obvious implication of this assumption is that the correct activation of a higher-level representation must, in some way, depend on the correct activation of lower-level representations. Lexical competition is one mechanism that would, presumably, prevent the lexical system from erroneously activating semantic and other upper-level representations of words not being presented (by suppressing any competitors at the lexical level). Hence, lexical competition would seem to be a mechanism that would need to be developed somewhat early by L2 readers. Nonetheless, L2 readers appear to have weak lexical competition operations (Nakayama and Lupker, 2018). Therefore, for those individuals, co-activation of orthographically related words is likely to occur frequently, leading to incorrect activation of the upper-level representations (e.g., lexical/semantic representations of orthographically similar words to the target would be co-activated). When there are only weak lexical competition operations, as seems to be the case for many L2 readers, it seems that it would be difficult for upper-level representations to become firmly established unlike in the L1 lexicon. Indeed, in the present experiments, our results suggest something of this sort is occurring. That is, when L2 lexical representations are weak (i.e., for the participants with lower TOEIC scores and larger orthographic priming effects), it is difficult for morphological representations to produce priming. However, as L2 readers become more skilled, their orthographic representations become less fuzzy, their orthographic priming effects become smaller, and priming based on morphological relationships alone becomes larger.

All this is not to say, however, that lower-level representations need to be fully developed before higher-level representations start to develop. Recently, Bordag et al. (2021) proposed the Ontogenesis Model (OM) of L2 lexical representations, which may provide further insight with respect to this issue. The OM is a model that provides a blueprint of the nature of L2 lexical representations and their development at the differential phases of L2 attainment. The OM posits that a central characteristic of the L2 lexicon is fuzziness in how words are represented and connected, and a weak form-based competition is one result of this fuzziness in L2 orthographic representations. Importantly, the OM also has the ability to explain how the development of higher-level representations is not as strongly affected by the development of lower-level representations in the L2 lexicon as one might assume. Of particular relevance, the OM acknowledges that L2 lexical acquisition is different from L1 lexical acquisition in that, although lexical acquisition in L1 involves the simultaneous development of both form and semantic representations, "*lexical acquisition in L2 often involves the establishment of a new form representation and its mapping onto an existing semantic representation*" (p: 10). The connection between form-level representations and upper-level (semantic) representations is therefore somewhat weaker in L2. The OM also posits that the weaker form to meaning link in L2 makes it difficult for co-activated (i.e., incorrect) orthographic representations to activate their upper-level representations. However, a link between a given L2 word form and its meaning is typically established differently than would be the case in L1. That is, that link is established as a result of mapping the existing semantic knowledge to the word form. Under this view, co-activation of orthographically similar words at the orthographic level does not substantially hinder the correct semantic activation of a target word. The OM in its current form does not yet incorporate representations involving morphological relationships (see Footnote 4, Bordag et al., 2021). However, it seems possible to apply the model's principles to the mappings between L2 forms and morphological/lemma level representations. That is, the present-tense and past-tense morphological relationship in the L1 language, which has been established at the conceptual level (e.g., aruku-aruita, utau-utatta; Japanese), can be directly mapped onto present- and past-tense forms in L2 (e.g., walk-walked, sing-sang; English).

The Contrast Between Experiment 2 and the Previous Studies Testing Past-Tense Morphological Priming in L2

In the present experiments, we largely followed the design of Feldman et al. (2010) by employing both an orthographic and an unrelated baseline. In terms of priming effects relative to unrelated primes, our results were largely consistent with what Feldman et al. observed—priming effects were significant, and they did not interact with verb type. However, our results were different in terms of priming effects relative to orthographic primes (particularly in Experiment 2); while we observed significant morphological priming effects for both regular and irregular verbs, Feldman et al. observed significant effects for regular verbs only.

One possible reason for the discrepancy is that there may have been insufficient power in Feldman et al.'s (2010) design. Although the number of participants in Feldman et al.'s experiment was large ($N=90$), the number of items for each verb type was small ($n=21$), which meant that, with three prime types for each target verb type, there were only seven items per cell. Further, numerically, their morphological priming effects when measured against orthographic primes were non-zero (2-19 ms). For irregular verbs, for example, the priming effect was as large as 14 ms. Our Experiment 1, using a similar number of participants and number of items as used by Feldman et al., also suffered from some ambiguity in its data pattern as it produced only small priming effects that were not significant when considered individually. It is possible that Feldman et al.'s experiment, as well as the present Experiment 1, did not have enough power to detect small differences.

Another possible explanation for the observed difference between the results in our Experiment 2 and Feldman et al.'s (2010) is that there was a difference in the word frequencies of targets used in the two experiments. In the present experiments, we selected our stimuli in such a way that our bilinguals would be quite familiar with the words (both primes and targets). As a result, our targets were very high frequency words (>350 occurrences per million), which we assumed that bilinguals would have had abundant experience with. On the other hand, target words in Feldman et al.'s experiments were of medium frequency (60–85 occurrences per million). A reflection of this difference can be seen in the speed of overall responses. In Feldman et al.'s experiments, the bilinguals' mean RTs ranged from 725 to 800 ms. In the present experiments, they ranged from 591 to 676 ms. In fact, the RT ranges observed for our Japanese-English bilinguals were quite similar to those observed for the L1 English speakers in Feldman et al.'s experiments (606–664 ms). Hence, it is likely that the representations of words used in our experiments had a higher overall level of entrenchment for our bilinguals. If L1-like connections do become established in the L2 lexicon with experience, such connections would be more likely to exist for words bilinguals encounter and process more often. The discrepancy between their experiments and ours may be that our high-frequency English stimuli allowed us to look at the connections that had become more entrenched than the connections that Feldman et al.'s stimuli allowed them to examine. In effect, word familiarity would be working in the same way as participant proficiency (see **Figure 1**).

As noted, in terms of priming effects from morphological primes relative to unrelated primes, our results were consistent with Feldman et al. (2010): both regular and irregular past-tense forms primed their present-tense targets. This result, however, is inconsistent with the findings for L2 readers reported by Silva and Clahsen (2008) and Clahsen et al. (2013), who found no priming for regular past-tense forms relative to unrelated primes. Although we do not know why Silva and Clahsen and Clahsen et al. did not observe priming for L2 readers, some possible sources of the discrepancy may include variables relating to list context such as the number of filler trials and type of non-word targets in the stimuli, as well as the frequency of word items in the experiment.

Participant variables such as the participants' first language background and L2 proficiency could also be at play with respect to this discrepancy. First, we consider the possibility of proficiency differences. Our results with Japanese-English bilinguals suggest that the morphological condition should facilitate target responses relative to unrelated controls for both more proficient and less proficient bilinguals, albeit for slightly different reasons. The facilitation is likely to be mainly driven by prime-target orthographic overlap for less proficient participants, whereas it is likely more due to the prime-target morphological relationship for proficient participants. Therefore, a lack of priming based on morphological relationships could be explained if the participants of Silva and Clahsen (2008) and Clahsen et al. (2013) were overall less proficient than ours. Unfortunately, we cannot compare the proficiency measure used in our *post-hoc* analysis (i.e., the TOEIC) to Silva and Clahsen and Clahsen et al.'s analysis (i.e., the Oxford Placement Test).

What is also puzzling is why the regular past-tense primes in Silva and Clahsen (2008) and Clahsen et al. (2013) did not facilitate recognition of present-tense targets based on orthographic similarity, because the orthographic overlap between prime-target pairs has consistently been observed to facilitate target recognition for L2 readers (e.g., the present experiments; Nakayama and Lupker, 2018; Jiang, 2021; Kida et al., 2022). One possibility is that the nature of the bilinguals' L1 versus L2 script could have affected lexical competition, as discussed by Nakayama and Lupker (2018): lexical competition operates more readily for same-script bilinguals than different-script bilinguals. Thus, in the case of same-script bilinguals, the orthographic similarity of L2 words would inhibit (or facilitate more weakly) target recognition. If those same-script bilinguals are also less proficient in the L2 and do not (yet) possess sufficient morphological processing skills, morphological primes will not facilitate target recognition relative to unrelated primes.

CONCLUSION

Previous studies have shown that connections between past-tense verbs and their present-tense forms exist in native (L1) English readers' lexicons. Our aim was to examine whether such connections are established in L2 English readers' lexicons. The results of two masked priming lexical decision experiments with Japanese-English bilinguals showed that morphologically related primes facilitated lexical decisions to targets more than orthographically related primes. This pattern suggests that facilitation occurs due to the primes' morphological relationship with their targets. Hence, our results demonstrate that connections in terms of past-tense morphology develop for L2 readers for both regular and irregular verbs. This result also indicates that at the morphological level, there is little imprecise (or fuzzy) encoding with respect to representations of morphological relationships. The present experiments also demonstrated that orthographic similarity of word primes and their targets facilitates target recognition for L2 readers, and thus underlines the importance of using orthographic primes in order to correctly gauge the impact of the underlying connections based on morphological relationships.

We would, therefore, like to advocate that subsequent studies should continue to use orthographic, in addition to, unrelated primes when attempting to gauge the impact of morphological priming effects in L2. With regard to the effects of orthographic similarity for L2 readers, significant facilitation effects have been observed with different-script bilinguals (Li et al., 2017a,b; Nakayama and Lupker, 2018; Jiang, 2021). Interestingly, the current models of bilingual visual word recognition (e.g., Bilingual Interactive Activation Model, Dijkstra et al., 1998; Bilingual Interactive Activation Model+, Dijkstra and van Heuven, 2002) do not assume that orthographically similar primes produce facilitation effects for same-script bilinguals (e.g., French-English bilinguals, Dutch-English bilinguals).

Empirically, however, orthographic primes do appear to facilitate L2 lexical decision latencies for same-script L2 readers based on data from morphological priming experiments (e.g., Diependaele et al., 2011; Heyer and Clahsen, 2015). Therefore, the models cannot be precisely correct. The notion of form-prominence resulting from fuzziness in L2 encoding of orthographic form, however, would offer an explanation for these empirical findings (See Footnote 1). The presence and degree of orthographic facilitation appears to be affected by a number of factors, such as L1 orthography, L2 proficiency, stimulus compositions, etc. Thus, until we have a clearer understanding of orthographic priming effects on L2 visual word recognition, it would be advisable to use an orthographic baseline in addition to an unrelated baseline in order to make sure that the effects of orthographic (form) and morphological relationships are fully dissociated.

DATA AVAILABILITY STATEMENT

Readers interested in the data supporting the conclusions of this article may contact the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Review Committee on Research with Human Subjects, Waseda University and the Ethical Review Board of the Graduate School of International Cultural Studies, Tohoku University. The participants provided their written informed consent to participate in these studies.

AUTHOR CONTRIBUTIONS

MN contributed to the conception and design of the study. JW-K and MN prepared the experimental stimuli. JW-K and MY collected data and prepared it for analysis. MN, MY, and JW-K performed statistical analysis. JW-K wrote the initial draft of the manuscript. MN, MY, SL, RV, and JW-K wrote sections of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by JSPS KAKENHI Grant Number JP19K14468.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.742965/full#supplementary-material

REFERENCES

Andrews, S., and Lo, S. (2012). Not all skilled readers have cracked the code: individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* 38, 152–163. doi: 10.1037/a0024953

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

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bordag, D., Gor, K., and Opitz, A. (2021). Ontogenesis model of the L2 lexical representation. *Biling. Lang. Congn.* 25, 185–201. doi: 10.1017/S1366728921000250

Brysbaert, M., Lagrou, E., and Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Biling. Lang. Congn.* 20, 530–548. doi: 10.1017/S1366728916000353

Clahsen, H., Balkhair, L., Schutter, J.-S., and Cunnings, I. (2013). The time course of morphological processing in a second language. *Second. Lang. Res.* 29, 7–31. doi: 10.1177/0267658312464970

Coltheart, M., Davelaar, E., Jonasson, T. J., and Besner, D. (1977). "Access to the internal lexicon," in *Attention and Performance VI*. ed. S. Dornič (New Jersey: Lawrence Erlbaum Associates), 535–555.

Cook, S., and Gor, K. (2015). Lexical access in L2: representational deficit or processing constraint? *The Mental Lexicon* 10, 247–270. doi: 10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. (2016). Fuzzy nonnative Phonolexical representations Lead to fuzzy form-to-meaning mappings. *Front. Psychol.* 7:1345. doi: 10.3389/fpsyg.2016.01345

Crepaldi, D., Rastle, K., Coltheart, M., and Nickels, L. (2010). 'Fell' primes 'fall', but does 'bell' prime 'ball'? Masked priming with irregularly-inflected primes. *J. Mem. Lang.* 63, 83–99. doi: 10.1016/j.jml.2010.03.002

Davis, C. J., and Lupker, S. J. (2006). Masked inhibitory priming in English: evidence for lexical inhibition. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 668–687. doi: 10.1037/0096-1523.32.3.668

Diependaele, K., Duñabeitia, J. A., Morris, J., and Keuleers, E. (2011). Fast morphological effects in first and second language word recognition. *J. Mem. Lang.* 64, 344–358. doi: 10.1016/j.jml.2011.01.003

Dijkstra, T. (2005). "Bilingual visual word Recognition and lexical Access" in *Handbook of Bilingualism: Psycholinguistic Approaches*. eds. J. F. Kroll and A. M. B. De Groot (New York, NY, US: Oxford University Press), 179–201.

Dijkstra, T., and van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Biling. Lang. Congn.* 5, 175–197. doi: 10.1017/S1366728902003012

Dijkstra, T., Van Heuven, W. J. B., and Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica* 38, 177–196. doi: 10.5334/pb.933

Duyck, W., Vanderelst, D., Desmet, T., and Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychon. Bull. Rev.* 15, 850–855. doi: 10.3758/PBR.15.4.850

Eiken Foundation of Japan (n.d.). Overview of the EIKEN tests. Available at: https://www.eiken.or.jp/eiken/en/eiken-tests/overview/ (Accessed June 27, 2022).

Feldman, L. B., Kostić, A., Basnight-Brown, D. M., Đurđević, D. F., and Pastizzo, M. J. (2010). Morphological facilitation for regular and irregular verb formations in native and non-native speakers: little evidence for two distinct mechanisms. *Biling. Lang. Congn.* 13, 119–135. doi: 10.1017/S1366728909990459

Feldman, L. B., and Weber, K. (2012). "Morphological Processing: A Comparison of Graded and Categorical Accounts," in *Visual word Recognition: Meaning and context, Individuals and Development*. *Vol. 2*. ed. J. Adelman (New York, NY, US: Psychology Press), 3–23.

Forster, K. I., Davis, C., Schoknecht, C., and Carter, R. (1987). Masked priming with graphemically related forms: repetition or partial activation? *The Qua. J. Experimental Psychol. Sec. A* 39, 211–251. doi: 10.1080/14640748708401785

Forster, K. I., and Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behav. Res. Methods Instrum. Comput.* 35, 116–124. doi: 10.3758/BF03195503

Fox, J., and Weisberg, S. (2019). *An R companion to applied regression*. Thousand Oaks: Sage publications.

Fruchter, J., Stockall, L., and Marantz, A. (2013). MEG masked priming evidence for form-based decomposition of irregular verbs. *Front. Hum. Neurosci.* 7:798. doi: 10.3389/fnhum.2013.00798

Gor, K., and Cook, S. V. (2020). A mare in a pub? Nonnative facilitation in phonological priming. *Second. Lang. Res.* 36, 123–140. doi: 10.1177/0267658318769962

Heyer, V., and Clahsen, H. (2015). Late bilinguals see a scan in scanner AND in scandal: dissecting formal overlap from morphological priming in the processing of derived words. *Biling. Lang. Congn.* 18, 543–550. doi: 10.1017/S1366728914000662

Jiang, N. (2015). "Six decades of research on lexical representation and processing in bilinguals," in *The Cambridge Handbook of Bilingual Processing*. ed. J. W. Schwieter (Cambridge: Cambridge University Press), 29–84.

Jiang, N. (2021). Orthographic friends and lexical strangers in the L2 lexicon. *J. Second Lang. Stud.* 4, 224–244. doi: 10.1075/jsls.21012.jia

Keuleers, E., and Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behav. Res. Methods* 42, 627–633. doi: 10.3758/BRM.42.3.627

Kida, S., Barcroft, J., and Sommers, M. (2022). Word learning and lexicalization in a second language: evidence from the prime lexicality effect in masked form priming. *Mem. Cogn.*, 1–18. doi: 10.3758/s13421-022-01274-6

Kielar, A., and Joanisse, M. F. (2010). Graded effects of regularity in language revealed by N400 indices of morphological priming. *J. Cogn. Neurosci..* 22, 1373–1398. doi: 10.1162/jocn.2009.21353

Kielar, A., Joanisse, M. F., and Hare, M. L. (2008). Priming English past tense verbs: rules or statistics? *J. Mem. Lang.* 58, 327–346. doi: 10.1016/j.jml.2007.10.002

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lenth, V. R. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.7.2) [Software]. Available at: https://CRAN.R-project.org/package=emmeans

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10, 707–710.

Li, M., Jiang, N., and Gor, K. (2017b). L1 and L2 processing of compound words: evidence from masked priming experiments in English. *Biling. Lang. Congn.* 20, 384–402. doi: 10.1017/S1366728915000681

Li, J., Taft, M., and Xu, J. (2017a). The processing of English derived words by Chinese-English bilinguals. *Lang. Learn.* 67, 858–884. doi: 10.1111/lang.12247

Milin, P., Smolka, E., and Feldman, L. B. (2018). "Models of lexical access and morphological processing," in *The Handbook of Psycholinguistics*. eds. E. M. Fernández and H. S. Cairns (Hoboken, NJ: John Wiley & Sons), 240–268.

Morris, J., and Stockall, L. (2012). Early, equivalent ERP masked priming effects for regular and irregular morphology. *Brain Lang.* 123, 81–93. doi: 10.1016/j.bandl.2012.07.001

Nakayama, M., and Lupker, S. J. (2018). Is there lexical competition in the recognition of L2 words for different-script bilinguals? An examination using masked priming with Japanese-English bilinguals. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1168–1185. doi: 10.1037/xhp0000525

Nakayama, M., Sears, C. R., Hino, Y., and Lupker, S. J. (2013). Masked translation priming with Japanese–English bilinguals: Interactions between cognate status, target frequency and L2 proficiency. *J. Cogn. Psychol.* 25, 949–981. doi: 10.1080/20445911.2013.839560

Nakayama, M., Sears, C. R., and Lupker, S. J. (2008). Masked priming with orthographic neighbors: A test of the lexical competition assumption. *J. Exp. Psychol. Hum. Percept. Perform.* 34, 1236–1260. doi: 10.1037/0096-1523.34.5.1236

Ota, M., Hartsuiker, R. J., and Haywood, S. L. (2009). The KEY to the ROCK: near-homophony in nonnative visual word recognition. *Cognition* 111, 263–269. doi: 10.1016/j.cognition.2008.12.007

Pastizzo, M. J., and Feldman, L. B. (2002). Discrepancies between orthographic and unrelated baselines in masked priming undermine a decompositional account of morphological facilitation. *J. Exp. Psychol. Learn. Mem. Cogn.* 28, 244–249. doi: 10.1037/0278-7393.28.1.244

Pinker, S., and Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn. Sci.* 6, 456–463. doi: 10.1016/S1364-6613(02)01990-3

Qiao, X., and Forster, K. I. (2017). Is the L2 lexicon different from the L1 lexicon? Evidence from novel word lexicalization. *Cognition* 158, 147–152. doi: 10.1016/j.cognition.2016.10.026

R Core Team (2021). *R: A language and environment for statistical computing* (Version 4.1.1) [Software]. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Segui, J., and Grainger, J. (1990). Priming word recognition with orthographic neighbors: effects of relative prime-target frequency. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 65–76. doi: 10.1037/0096-1523.16.1.65

Silva, R., and Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: evidence from masked priming experiments in English. *Biling. Lang. Congn.* 11, 245–260. doi: 10.1017/S1366728908003404

Stockall, L., and Marantz, A. (2006). A single route, full decomposition model of morphological complexity: MEG evidence. *The Mental Lexicon* 1, 85–123. doi: 10.1075/ml.1.1.07sto

Veríssimo, J., Heyer, V., Jacob, G., and Clahsen, H. (2018). Selective effects of age of acquisition on morphological priming: evidence for a sensitive period. *Lang. Acquis.* 25, 315–326. doi: 10.1080/10489223.2017.1346104

Voga, M., Anastassiadis-Syméonidis, A., and Giraudo, H. (2014). Does morphology play a role in L2 processing?: two masked priming experiments with Greek speakers of ESL. *Lingvisticæ Investigationes* 37, 338–352. doi: 10.1075/li.37.2.10vog

Witzel, N., Qiao, X., and Forster, K. (2011). Transposed letter priming with horizontal and vertical text in Japanese and English readers. *J. Exp. Psychol. Hum. Percept. Perform.* 37, 914–920. doi: 10.1037/a0022194

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership