Studies in Second Language Acquisition (2025), **47**, 361–387
doi:10.1017/S0272263124000482

STATE OF THE SCHOLARSHIP

Phonological processing and the L2 mental lexicon

Looking back and moving forward

Isabelle Darcy[1,7], Miquel Llompart[2], Rachel Hayes-Harb[3], Joan C. Mora[4], Miren Adrian[4], Svetlana Cook[5] and Mirjam Ernestus[6]

[1]Indiana University, Bloomington, IN, USA; [2]Universitat Pompeu Fabra, Barcelona, Spain; [3]University of Utah, Salt Lake City, UT, USA; [4]Universitat de Barcelona, Barcelona, Spain; [5]University of Michigan, Ann Arbor, MI, USA; [6]Radboud University, Nijmegen, The Netherlands and [7]Université Grenoble-Alpes, Grenoble, France
Corresponding author: Isabelle Darcy; E-mail: idarcy@iu.edu

(Received 29 September 2023; Revised 13 June 2024; Accepted 12 July 2024)

Abstract

Twenty-five years ago, the publication of an article by Pallier, Colomé, and Sebastián-Gallés (2001) launched a new and rapidly evolving research program on how second language (L2) learners represent the phonological forms of words in their mental lexicons. Many insights are starting to form an overall picture of the unique difficulties for establishing functional and precise phonolexical representations in L2; however, for the field to move forward it is pertinent to outline its major emerging research questions and existing challenges. Among significant obstacles for further research, the paper explores the current lack of theoretical agreement on the concept of phonolexical representations and the underlying mechanism involved in establishing them, as well as the variable use of the related terminology (e.g., fuzziness and target-likeness). Methodological challenges involved in investigating phonological processing and phonolexical representations as well as their theoretical implications are also discussed. To conclude, we explore the significance of L2-specific phonological representations for the bottom-up lexical access during casual, conversational speech and how our emerging knowledge of L2 lexical representations can be applied in an instructional setting as two potentially fruitful research avenues at the forefront of the current research agenda.

Keywords: Fuzzy phonolexical representations; L2 mental lexicon; phonological processing; speech perception; spoken word recognition

Introduction

We use language to talk about people, ideas, events, or objects that do not need to be physically present; language systems are symbolic. Symbols stand in for or *represent* something else. Therefore, an important property of language systems is that of a

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

cognitive or mental *representation*. A language system has representations for concepts through words, but other types are also used, such as representations of grammatical structures or phonological categories. These representations are connected and can be combined—for example, users of a given language have learned specific patterns to combine words into a sentence. Importantly, representations must be learned (that is, stored in memory), and *shared* by members of a group to enable interaction. They are acquired early by children through exposure and interaction (though some representations are acquired later than others). Representations also must be retrieved, or *accessed*, during interaction to enable understanding. The process of accessing representations for words, which involves mapping the speech signal onto available representations, is called *lexical access*.

Representations for words are complex and contain multiple kinds of information, such as a word's meaning, its grammatical category, its pronunciation (that is, its *phonolexical representation*) and, if the person is literate, its spelling, among others (Hulstijn, 2001). In the first language (L1), phonolexical representations are built essentially from what a child perceives in the input. Therefore, the form of representations closely mirrors the L1 phonological system. For example, the phonolexical representation of a rose in a French child's mental lexicon could resemble /ʁoz/, but in an English child's, it could be /ɹəʊz/. The phonological content of these representations uses units (such as vowels or consonants, but also other elements such as word stress) that are specific to French and English, respectively. What is important to note here is that these units, and the representations that contain them, are language–specific, that is, they refer to a particular language (usually the first language). This close overlap between input and representation helps to access and recognize spoken words in the L1 very efficiently.

In any language learned after the first one (L2), not only the perception of speech sounds but also the stored phonological representations of words and the mechanisms to access them, are influenced by the L1. To illustrate, it is possible that for an English learner of French, the representation of the color word *rouge* 'red' /ʁuʒ/ could be approximately /ɹuʒ/, showing an L1 influence in the initial consonant. This makes recognizing spoken words more complicated and slower in the L2 than in the L1 because the phonological form of L2 representations does not always match the actual input sufficiently closely. In some cases, the representation is fragmentary or imprecise, an issue we discuss below in the section on "Fuzziness." In addition, the very process of *accessing* the representations can also be influenced by the L1, making it less effective during L2 spoken word recognition, even in the presence of accurate, precise, and separate representations (Cook & Gor, 2015). First, the L2 listener's phonological processing may lead to more words or word fragments being activated due to confusable sounds (Broersma & Cutler, 2008). Second, listeners may use L1-specific routines when accessing the lexicon, for instance, they may rely on L1 phonotactic knowledge to find word boundaries in the speech stream (we return to the issue of lexical access in the section Lexical representations). This impacts learners' ability to decode the input and can lead to bottlenecks in the listening process (Goh, 2000), making it difficult to follow spoken conversations in everyday life. In this paper, we discuss the three interdependent aspects of L2 phonolexical behavior: perceptual processing of the input, the content and form of lexical representations, and accessing the lexicon. Discussing them together is necessary because our theoretical assumptions about one aspect may determine our views of the other two.

Besides the impact on spoken word recognition just evoked, this broad influence of L1 phonology in the L2 mental lexicon can have wide-ranging consequences for L2

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

learners and their language use in the realm of production—and therefore also one's intelligibility—as well as in reading and writing. Helping learners ascertain their representations and streamline the perceptual decoding and processing of their new language would be a valuable component of language teaching. Much remains to be done to better understand the mechanisms that help learners reduce the L1 influence over time in all three aspects: phonological processing, phonolexical representations, and lexical access.

Over the past two and a half decades, this area of research has grown and led to substantial discoveries about the shape of lexical representations in multilingual speakers, as well as about the way representations are accessed and acquired over time. Just about 25 years after the important starting point made by the publication of an article by Pallier, Colomé, and Sebastián-Gallés in 2001, now is an exciting time to outline the essential questions that remain to be explored, and to work collectively toward a long-term research agenda. Several methodological challenges are also best addressed together and across disciplines, for the field to move forward on a solid basis. In this paper, we summarize the most important progress made since then, before addressing some enduring challenges and outlining several promising avenues for future research.

Current state of the research

Where it all began

We have known for nearly a century that listeners do not perceive an L2 in the same way as native listeners do. L2 learners seem to speak the L2 "through" their native language; they also perceive the L2 "through" their L1. An early proponent of this phenomenon was Polivanov (Polivanov, 1931). Since then, the field has been collecting experimental evidence supporting his vision (e.g., Best & Strange, 1992; Best & Tyler, 2007; Bohn, 1995; Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997; Dupoux, Kakehi, Hirose, Pallier, & Mehler 1999; Flege & Bohn, 2021; Goto, 1971; Iverson et al., 2003; Kabak & Idsardi, 2007; McAllister, Flege, & Piske, 2002; Sebastián-Gallés, 2005; Segui, Frauenfelder, & Hallé, 2001; Strange, Akahane-Yamada, Kubo, Trent, & Nishi, 2001, and many more). Several decades of research have established that L2 listeners initially over-rely on their familiar L1 processing routines, using their L1 phonological system to decode and produce L2 speech. This phenomenon, which we may call "L1–based processing," interferes with an accurate perception of L2 in adulthood, revealing L1 influence in every dimension of the phonological system, including segmental, suprasegmental, phonotactic, and prosodic dimensions. It also impacts behavior across all skills, such as production, reading and listening comprehension, and the process of learning itself. Importantly, researchers also established that L1–based processing can diminish as proficiency and L2 use increases, although it is difficult to fully eliminate. Explicit instruction benefits learners by reducing the impact of L1–based processing, and by helping them develop a more robust L2 phonological system that is more closely aligned with the phonology of the target language. For instance, new phoneme categories can be created over time, or L2 phonotactic constraints can be acquired (Cabrelli, Luque, & Finestrat-Martínez, 2019).

Compared with the wealth of research on L2 perception, much less was known about the way phonological representations are implemented in the L2 mental lexicon until relatively recently: Pallier and colleagues (2001) were the first to explore the consequences of potential misperceptions for lexical representations. They asked whether the

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

fact that L2 listeners repeatedly perceive words with segments other than the ones these words actually contain would lead to ambiguity in how these words are represented. Their study revealed that L2 learners appear to represent many words the way they initially perceived them. The authors based their conclusions on a task known as an auditory lexical decision task with repetition priming. In this task, participants must decide whether each item they hear in a list is a word or a nonword. When a word is repeated further down the list, decision times are typically shorter because the item has been recently processed (repetition priming). Pallier et al. examined three Eastern Catalan contrasts absent in Spanish: /s, z/, and the vowel pairs /e, ɛ/ and /o, ɔ/, which map onto the Spanish mid-closed vowels /e/ and /o/, respectively. They presented minimal pair stimuli, such as *néta*, /netə/ 'granddaughter'—*neta*, /nɛtə/ 'clean,' in the lists to determine if processing the first would cause faster recognition of the second. While Catalan-dominant bilinguals displayed no repetition priming for minimal pairs (only for actual repetitions), Spanish-dominant bilinguals did. They responded faster to *néta* following itself, /netə/ (as expected), but *also* following *neta.* Pallier and colleagues interpreted these findings to suggest that the minimal pairs might in fact be stored lexically as homophones for some learners. Before concluding that these findings point to a lexical *representation* issue however, it matters to point out that in theory, another scenario could explain the results: It may be the case that the two representations exist separately and are not homophones, but one of them is never accessed by the percept, if participants do not perceive the phonetic difference between the sounds *during the task* (see Ota, Hartsuiker, & Haywood, 2009). This scenario would point to an *access* issue grounded in perception, not a representational one (although unlikely, it is conceivable that separate representations could be acquired without perceptual support, for instance, if acquired via metalinguistic or spelling information, although the lack of perceptual evidence might make this unviable in the long term.) In fact, the 2001 study did not separately examine whether participants were perceiving the contrasting sounds in the task as the same or not, and therefore cannot fully disambiguate between an access and a representational explanation. That is, the underlying forms of the lexical representations can only be investigated if we first establish how participants perceive the stimuli while participating in the task.

Still, these initial findings have been supported by several subsequent studies that assessed the perceptual component separately and confirmed Pallier et al.'s interpretation: It is indeed possible that initial misperceptions of the input are mirrored in the content of lexical representations (Cutler & Otake, 2004; Ota et al., 2009; Trofimovich & John, 2011; Darcy, Dekydtspotter, Sprouse, et al., 2012). This argument was extended to areas beyond segments (e.g., phonotactics: Darcy & Thomas, 2019; lexical stress: Dupoux et al., 2008). Taken together, these studies suggest that lexical representations can indeed be thought to be constrained by perception at the time of learning: If two similar words (such as *néta* and *neta*) are perceived as the same at the time at which they are first learned, they may be initially stored with the same phonolexical representation. And, as we discuss below, these phonolexical representations can remain inaccurate for a long time, even after the perception of the relevant contrast has improved.

What have we learned?

The findings of Pallier et al. (2001) are crucial because they established a clear connection between L2 perception and lexical representation by showing that difficulties in the former can lead to ambiguity (underdifferentiation, such as merged or homophonous lexical entries) in the latter. This of course highlights the importance of

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

L2 perceptual abilities. On the surface, it could be taken to imply that perception and lexical representation develop in close correspondence: When perception improves, representations become less ambiguous and more target-like, and, without perceptual improvement, no favorable changes at the lexical level are to be observed. However, as we outline in the following sections, one of the main areas in which research on L2 lexical representations has made progress since 2001 is precisely in the characterization of this link. Mainly, the accumulated evidence now suggests that the relationship between perception and lexical representation is much more nuanced than previously thought.

Accurate perception does not guarantee accurate lexical representations, even if it facilitates it

Several studies have shown that even L2 listeners who are able to perceive non-native phonological contrasts quite accurately do not seem to have fully encoded these contrasts into the phonolexical representations of L2 words (e.g., Amengual, 2016; Darcy, Daidone, & Kojima, 2013; Sebastián-Gallés, Echeverría, & Bosch, 2005). In a similar vein, the few studies that have assessed directly whether perceptual scores (for example on tasks requiring categorization such as ABX) predict lexical scores (such as on lexical decision or word-picture matching) have rendered mixed results (Darcy & Holliday, 2019; Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Elvin, 2016; Simonchyk & Darcy, 2017; Melnik & Peperkamp, 2021), or shown that perceptual abilities may sometimes not predict lexical encoding at all for L2 learners of intermediate–high to high proficiency (Daidone & Darcy, 2021; Llompart, 2021a). Finally, eye-tracking studies focusing on the time course of L2 word recognition suggest that, in some cases, even listeners with unreliable perception of specific L2 contrasts can distinguish between the phones in these contrasts when accessing their stored representations for L2 words (Cutler, Weber, & Otake, 2006; Weber & Cutler, 2004).

Together, these studies suggest that even though "good enough" perceptual discriminability is needed, perfectly accurate perception is not necessary—nor is it a guarantee for accurate phonolexical encoding. While it cannot be said that perception and phonolexical encoding are fully dissociated, the common finding of the above studies is that perception, while important, is not the sole predictor of how accurately words may be represented (see section Lexical access can be difficult despite accurate lexical representations).

This conclusion prompted substantial research into what other potential factors are likely to play a role in how learners acquire accurate phonolexical representations, asking what other information learners rely on to store words with sufficient precision if perception does not explain it all.

Learners may use other sources of information to supplement their perception

Above and beyond the interference from the L1 phonological system, other factors are likely involved in determining the precision with which L2 words are lexically encoded, and the ease with which these representations can be adjusted over time. One of the factors under investigation is the presence of orthographic input when learning words. Findings so far paint a complex picture of the role played by spelling when learning new words, or when processing and accessing familiar words. What is clear is that orthographic input interacts with these processes in different ways. For instance, when L1/L2 phoneme–grapheme correspondences mismatch, exposure to spelling information can interfere with the memorization of word forms (Hayes-Harb, Nicol, & Barker, 2010).

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Evidence also suggests that orthographic input may be beneficial in some cases. In a study that involved novel word learning, Escudero, Hayes-Harb, and Mitterer (2008) found that learners potentially relied on congruent L1 phoneme–grapheme correspondences to distinguish a difficult L2 contrast. Taken together, the role played by orthographic input is likely dependent on a range of conditions, on the specific L1/L2 phoneme–grapheme correspondences, and on learners' awareness (Brakovec & Darcy, 2023)—all of which need to be further elucidated.

Other factors examined by researchers include the role of explicit instruction (Bailey & Brandl, 2013; Lee, Plonsky, & Saito, 2020; Zhang & Yuan, 2020), the role of visual gesture information (Chan, 2018; Li, Xi, Baills, & Prieto, 2021; Llompart & Reinisch, 2017; Xi, Li, Baills, & Prieto, 2020), or the awareness of the existence of minimal pairs. For instance, providing minimal pairs during learning can cue learners to a contrast that is perceptually challenging for them, thus helping them establish distinct lexical representations (Llompart & Reinisch, 2020). However, these effects are not stable across studies and their underlying mechanisms remain to be more clearly understood.

Another factor that could guide learners to more fine-grained lexical representations is how many words they already know. Two recent studies (Daidone & Darcy, 2021; Llompart, 2021a) found that a larger L2 vocabulary was predictive of more accurate phonolexical representations. A possible mechanism could be that knowing more words may increase learners' ability to notice whether or how a newly learned word is distinct from all its phonological neighbors. This may lead them to push the representation toward being more contrastive, leading to the refinement of new and existing phonolexical representations (Rocca, Llompart, & Darcy, accepted).

Lexical access can be difficult despite accurate lexical representations

An important finding that enhanced our understanding of the relationship between phonological processing and lexical representations revolves around lexical access in L2, which is overall less straightforward than in L1, independent of how specific the representations are.

L2 listeners typically experience processing disadvantages when accessing their L2 lexicon, compared with their L1 lexicon. One of the emerging explanations for this effect is the entrenchment hypothesis (e.g., Diependaele et al., 2013; Gollan, Montoya, Cera, & Sandoval, 2008). Entrenchment relates to the role of lexical frequency in lexical access: More frequently retrieved and accessed words are more entrenched, and require less effort to retrieve and process. Specifically, entrenched units are easier to process and manipulate because they require less effort to be combined with and integrated into other structures (Schmid, 2010). Conversely, less entrenchment can cause processing costs, with implications for lexical access at different stages. While more frequent words are overall more entrenched, speaker-specific differences can emerge where low-frequency words used often by certain speakers are more entrenched in their lexicons. L2 learners do not experience the language in the frequencies equivalent to those of the L1 speakers, therefore, the L2 lexicon is generally less entrenched compared with the L1 lexicon, which leads to less specific representations in all areas of linguistic L2 knowledge—not only in phonological representations. As a consequence of weaker entrenchment, L2 words are less integrated into the phonological and semantic networks, slowing down access to these representations during speech comprehension (Cook & Gor, 2015; Cook, Pandža, Lancaster, & Gor, 2016; Gor, Cook, Bordag, Chrabaszcz, & Opitz, 2022).

It is important to stress that even when the phonological ambiguities are resolved with more L2 experience and stronger entrenchment of the L2 word forms in the

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

lexicon, accurate and automatic lexical access may still not be achieved. For example, in a priming experiment in L2 Russian the auditory target (/malatok/, 'hammer') was primed with a word (/karova/ 'cow') semantically related to the target's phonological competitor (/malako/ 'milk'). Slower processing for L2 speakers (compared to an unrelated condition) indicates that the pseudoprime 'cow' activated 'milk,' leading to complications and to a delay due to reanalysis when retrieving 'hammer,' providing evidence to weaker form-to-meaning mappings and retrieval of incorrect semantic content (Cook et al., 2016).

Learners' representations can change over time

Finally, an important property of learners' lexical representations that emerged from recent research is their potential to change over time (e.g., Darcy et al., 2012; Darcy et al., 2013). Generally speaking, representations are very stable, but not immutable, and a substantial body of research is currently investigating how learners are able to modify existing representations for L2 words, including the time course of these updates, and the factors that facilitate them (e.g., Darcy & Holliday, 2019; Llompart & Reinisch, 2021; Rothgerber, 2020). This idea was recently formalized in the Ontogenesis Model of the L2 Lexical Representation (OM, Bordag, Gor, & Opitz, 2021), which suggests that fuzziness is a pervasive property of the L2 lexicon, and continually evolves until words reach a stable, *optimum* state. The model describes the lexical development of an L2 word from the perspective of its ontogenetic curve toward its optimum, or the word's optimal encoding, which can happen independently in any of the domains (orthographic, phonological, or semantic). The lexical entry is assumed to reach its optimal encoding when all domains are able to be activated with relative synchronicity and reliable word identification, and automatic retrieval is achieved. According to the OM, most L2 words will not reach their optimum, leaving most lexical domains to some degree underspecified.

Enduring challenges

Progress in this area of research has also revealed several enduring challenges. One of these relates to the variable use of terminology (see the section Terminology). For example, while learners' representations are often characterized as "fuzzy" and "not target-like," there does not appear to be an agreed-upon definition of these terms across studies. As pointed out by Hayes-Harb and Barrios (2019), and Barrios and Hayes-Harb (2021), many patterns of learner performance have been attributed to phonolexical "fuzziness," though they arguably result from very different characterizations of learners' phonolexical representations.

A second enduring challenge in this research relates to available methodologies and their limitations (see the Methods section). Because phonolexical representations for spoken language words cannot be observed directly and must be probed via learners' perception of speech (or writing), or their production, it is often unclear—as briefly evoked in the section Where it all began above—whether learners' performance can be attributed uniquely to their phonolexical representations, to their phonetic/perceptual representations, and/or to how representations are accessed. We discuss methods that have been deployed in an effort to tease apart speech perception (phonetic representations) and phonolexical representations, along with their limitations and what is still needed.

A third challenge pertains to our understanding of lexical representations themselves (see the Lexical representations section). Some models assume a highly detailed,

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

exemplar-like type of representation, whereas others posit a more abstract format for each representation which would be more similar to a word citation form. These two extreme views are complemented by intermediate positions and hybrid models, which assume that representations are abstract but connected to detailed traces (for instance, as in Ramus, Peperkamp, Christophe, Jacquemot, Kouider, & Dupoux, 2010).

Terminology

Most research concerned with L2 phonolexical representations builds on the premise that they are in some way imprecise with regard to their phonological form, and that this imprecision leads to difficulties such as accepting nonwords as words (e.g., *summaly*, a nonword, being accepted as a real word alongside *summary*, e.g., Darcy et al., 2013) or delays in word identification (e.g., considering words like *rock* and *rocket* as candidates for recognition when hearing the word *locker*, e.g., Weber & Cutler, 2004). Two terms that are often used to characterize L2 phonolexical representations are "fuzzy" and "non-target-like." "Fuzzy" and "fuzziness" typically refer to how precise, or well-defined, L2 phonolexical representations are. "Target-likeness," on the other hand, refers to how well the representations of language learners match those of an idealized "target" speaker, who in most cases is presumed to be a native speaker of the language. While this sort of usage may seem intuitive and straightforward at first sight, a review of past work suggests that several issues should be clarified and taken into account in future research. In particular, the concept of "fuzziness" needs to be revisited both in terms of its meaning (e.g., is "fuzzy" just another way to say "imprecise"?) and its scope (i.e., are words fuzzy vs. are phonological units within words fuzzy?). Similarly, "target-likeness" should be problematized because (1) the idea that something is target-like requires a very clear definition of what such a target is, and this definition is often lacking, and (2) what could be conceived as the target is bound to change throughout the L2 learning process. These are the issues discussed in the following subsections.

"Fuzziness"

The terms "fuzziness" and "fuzzy" have been widely used in recent research on L2 lexical representations when describing the source of the difficulties that language learners experience in auditory word recognition when particular L2 phonological categories are involved. They are often used to highlight that representations are weak (Llompart & Reinisch, 2019b) or imprecise (Llompart, 2021a) from a phonological standpoint. However, there does not appear to be a generally agreed-upon definition of such "fuzziness" among researchers. At times, it is used to refer to phonetically or phonologically imprecise representations, while at other times it seems to be applied broadly to representations that are "non-target-like" in some unspecified way, for instance, Ota et al. (2009) or Cook et al. (2016).

Barrios and Hayes-Harb (2021) propose a typology of possible meanings of phonolexical fuzziness for difficult L2 contrasts. They elaborate eight scenarios by crossing two types of perceptual representations (neutralized, precise) with four types of phonolexical representations (neutralized, ambiguous, "not X," precise). For perception, **neutralized** vs. **precise** refers to the ability to distinguish phonological contrasts (precise), or the lack thereof (neutralized, sometimes also called *merged*). One example of perceptual neutralization/precision can be found in Højen and Flege (2006) where L1 Spanish listeners neutralized English vowel pairs in perception, whereas early learners distinguished the contrasts with high precision. For phonolexical representations,

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

precise is taken to mean something akin to "target-like"; that is, the functional opposition of two distinct and well-defined categories. For example, this would be the case when an L1-English learner encodes L2 Japanese singleton /t/ and geminate /tt/ separately in their lexical representations. In this case, /tt/ is the nondominant, unfamiliar category, whereas /t/ is the dominant category due to its similarity to English /t/. The other three terms describe three different possible characterizations of representational inaccuracy/fuzziness: (1) **neutralized**, where a nondominant (i.e., new) category is not distinguished from the dominant (i.e., familiar) category. In our example, a neutralization would happen if the learner encoded both /t/ and /tt/ as /t/; (2) **ambiguous**, where the nondominant category neither matches nor mismatches the dominant category. In our example, this would be the case if the learner had encoded /tt/ as /t?/ or /t*/ (Hayes-Harb & Masuda, 2008); and (3) "**not X**," where the nondominant category is differentiated from the dominant category (X) but is otherwise unspecified phonologically. In our example, this happens when /tt/ is encoded as "not /t/." Subsequently, Barrios and Hayes-Harb demonstrate that the type of phonolexical fuzziness that is assumed matters in that they make differential predictions for lexical decision and/or word-picture matching performance patterns, and that many of the eight scenarios have been documented in the literature as exemplifying "fuzziness," thus leading to a rather inconsistent use of this term.

It is useful to clarify at this stage what is the relationship between the percept (the result of categorization) and the phonolexical representation (which is stored in long-term memory representations for words). During spoken word recognition, the percept (created from the rapidly unfolding, complex acoustic signal) is mapped onto lexical representations; when the percept overlaps at least in part with the stored representation, lexical matches are activated on the fly, to select the most likely word that was heard (Jusczyk & Luce, 2002; Magnuson, Dixon, Tanenhaus, & Aslin, 2007). Three outcomes are typically possible during this mapping process: The percept can *match* the representation (contacting and activating the candidate), it can be a *mismatch* (inhibiting or reducing activation), and it can also be a *no-mismatch.* This third possibility—not an actual match but also not an actual mismatch—is thought to happen when representations are variable or underspecified (e.g., Fitzpatrick & Wheeldon, 2000; Lahiri & Reetz, 2010). In our case, if the representation is imprecise, the no-mismatch scenario still allows the activation of candidates and does not reduce the activation (see Darcy et al., 2013).

In this paper, we propose a slightly revised version of the terminology suggested by Barrios and Hayes-Harb (2021). We adopt the two-way distinction predicted for perception but resort to the terms **merged** and **distinct** instead of the labels of neutralized and precise they used, respectively. This is to avoid confusion with the labels used to characterize phonolexical representations. At the phonolexical level, we propose that two essential concepts must be teased apart to capture predicted patterns based on Barrios and Hayes-Harb's taxonomy. The first is the **precision** of the phonolexical representation, which we define here as the property of having phonological content such that it is activated by *specific* types of perceptual representations and not others (note that this is different from the use of "precision" by Barrios & Hayes-Harb 2021). The second concept is that of **contrastiveness**, which is a property of sets of phonolexical representations and has to do with whether representations are *differentially* activated by perceptual representations. To illustrate, a given percept ([t] or [tt]) will either match or mismatch a precise representation. In the case of imprecise representations, different scenarios are possible, depending on the representation's contrastiveness. If it is contrastive ("not /t/"), hearing the percept [t] will

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Table 1. A matrix of the relationship between precision and contrastiveness in lexical representations

	Precise	Imprecise
Contrastive	Functional distinction: /t/, /tt/	Not X: /t/, "not /t/"
Noncontrastive	Neutralized: /t/	Ambiguous: /t/, /t?/

Note: A functional distinction is not necessarily implemented in a target-like manner.

mismatch with the representation, but in the case of a noncontrastive imprecise (i.e., ambiguous) representation (/t?/), either percept ([t] or [tt]) will not mismatch the representation. In Table 1, we demonstrate how binary oppositions concerning precision and contrastiveness factorially generate relevant characterizations of phonolexical representations. For this reason, we propose that widespread adoption of the use of precision and contrastiveness as two different yet interrelated dimensions of what has routinely been called "fuzziness" would result in a more informative characterization of L2 lexical representations and would lead to more transparent and more readily testable predictions concerning their development.

Global and local representational "fuzziness"

Another issue with the use of "fuzzy" representations and phonolexical "fuzziness" in previous research concerns the scope of the representations these terms describe, or rather, the grain size of the units to which fuzziness applies. In its original conception within the Fuzzy Lexical Representations hypothesis (Cook, 2012; Cook & Gor, 2015; Cook et al., 2016), the term "fuzzy" was used to illustrate that the lexical representations of L2 learners "are not fully specified and lack details at both phonological and phonolexical levels of representation" (Cook et al., 2016). Crucially, fuzziness is conceptualized here as a general property of L2 lexical representations that leads to heightened lexical competition and imprecise word-meaning mappings during lexical access at a global level. Gradually, these terms have been adopted by researchers in L2 phonology acquisition to again refer to a lack of phonetic–phonological detail in L2 lexical representations, but only in relation to specific L2 phonological categories and contrasts that are known to trigger difficulties in speech perception and production (e.g., /r/-/l/ for native speakers of Japanese), so a more local ambiguity in this case. While we do not consider these two conceptualizations to be incompatible, and, in fact, we argue that more research is needed to delineate the extent to which both refer to the same underlying phenomenon, it would be appropriate for future research endeavors to provide an explicit statement about their scope in this regard. We suggest that an opposition of terms such as ***global*** vs. ***local*** fuzziness could be helpful in our attempts to clarify our use of this terminology.

The "target"

Finally, another pair of terms that is often used in research on L2 lexical representations is "target-like" and its contrary, "non-target-like" (Darcy et al., 2012; Simonchyk & Darcy, 2017). These terms primarily serve to make statements about the figurative distance between the representations of words by L2 learners and those that so-called "native" speakers of the language are assumed to have. A move away from the concepts of "native" and "non-native" (e.g., Cheng, Burgess, Vernooij, Solís-Barroso, McDermott, & Namboodiripad, 2021) requires us to also interrogate our conception of the language learner's target (which has long been advocated for), so simply replacing

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

"native-like" with "target-like" is inadequate without a more nuanced and inclusive description of the target. For this reason, it is important to define the target as unambiguously as possible. To do so, the two dimensions elaborated above, precision and contrastiveness, may provide a useful starting point for describing the language learner's target in functional and representational terms.

An alternative is to characterize the learner's target as phonological ***alignment*** of the learner's representations (particularly for challenging L2 categories and features) with the language that learners actually experience in their input. This approach to characterizing the target requires a close examination of the linguistic history of the learner. For example, beginning learners in an instructional nonimmersion setting may have limited exposure to the language, and much of their exposure is likely to come from student peers. To the extent we are interested in the development of language learners' phonological representations, we should focus on this development in relation to the input they are actually exposed to (see Eger & Reinisch, 2019 and Llompart & Reinisch, 2021). Whether or not learners achieve socially-defined language targets (such as "native-likeness") is an interrelated, but conceptually distinct, question. Further, a learner's target is not necessarily static but bound to change over time as a function of how the learner's personal circumstances, careers, relationships, and interests evolve. Because of this, in many cases, we as researchers may still lack the appropriate tools and the necessary information to properly assess this alignment, and thus call for caution regarding our underlying assumptions. Finally, all language users—even those characterized as "native speakers"—exhibit variability (e.g., Perry, Kelley, & Tucker, 2024) in perceptual and lexical behavior. To properly characterize a learner's achievement of some target, we must recognize that the target is inherently variable, and expect variability in language learners' performance.

In light of all this, we must discard the idea of "the target" as a monolithic entity and embrace a more flexible conception of a target that can vary depending on the learning environment and experience of the learner, how precision, distinctiveness, and alignment are weighted, and also as a function of the research questions at hand. For instance, spoken word recognition research assessing lexical activation and competition dynamics may lead to a definition of a target with *precision* and *contrastiveness* as key dimensions, as these tasks can shed light on the degree of specificity with which the percept activates phonolexical representations (i.e., precision) and on whether these representations show differential patterns of contact and activation (i.e., contrastiveness), but often not so much on whether the differentiation itself aligns with that of the predicted target. Other tasks and measures (such as elicited production tasks) may present a better opportunity to incorporate phonological alignment into their definition of the target, provided that there is a good understanding of who (and what) the subject of comparison for such alignment is. In sum, while there may not be a straightforward solution for the issues outlined above, we believe that a valuable first step for future work will be to clearly voice the assumptions made with regard to the intended target for any L2-learning populations, while also considering the extent to which these can be probed through the experimental method of choice.

Methods

Investigating the state of phonolexical representations in the learners' lexicon can be demanding from a methodological standpoint. As described above, a central challenge is to ensure that our methods uniquely probe the phonolexical representations

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

themselves, without artifacts of perceptual or production processes. Indeed, as mentioned above, several studies' findings do not unambiguously reflect an effect at the level of phonolexical representations (e.g., Pallier et al. 2001; Weber & Cutler 2004). It is thus essential that we reflect upon our current methods, and seek improvements to tease apart the contributions that perception (and production), and phonolexical representations have on word processing.

Concerns over confounding representational issues with processing difficulties at different levels are not new, and researchers have attempted to mitigate these potential confounds in a variety of ways. Over the last 20 years, several methods have emerged that permit such disambiguation to some extent. For example, Darcy et al. (2013) demonstrated that careful attention to participant performance across word/nonword and segment conditions can serve to distinguish between difficulty at the perceptual vs. phonolexical levels of representation. More recently, and as described above, Hayes-Harb and Barrios (2019) and Barrios and Hayes-Harb (2021) have further elaborated this line of thinking. Another approach to addressing this ambiguity has been to ensure that learners are in fact able to perceive the targeted contrasts (e.g., Amengual, 2016; Llompart, 2021b; Llompart & Reinisch, 2020), though this approach cannot eliminate the possibility that perceptual sensitivity to the contrast is depressed under the more demanding conditions of tasks requiring lexical access. For instance, Weber and Cutler (2004) and Escudero et al. (2008) employed eye-tracking and the visual world paradigm to observe word recognition dynamics over time, arguing that asymmetric-looking patterns can provide evidence of contrastive phonolexical representation of L2 contrasts, despite perceptual neutralization of the contrast. Another promising strategy for isolating the influence of phonolexical representations from perceptual representations is to bypass auditory perceptual altogether in favor of the visual presentation of words via written forms, as Ota et al. (2009) have done (see below).

Importantly, even though our current methods—including lexical decision, auditory word-picture matching, and eye tracking—have already brought about significant progress in our understanding of L2 phonolexical representations, their full potential has not been exhausted. In particular, if a greater emphasis is placed on manipulating these methods to target the main issue discussed above, mainly how perceptual routines and phonolexical representations interact in L2 word processing and how to capture their relative contributions with confidence, we can envisage at least two promising avenues for further research.

A first approach that has the potential to be transformative in how we think about representational issues at the word level is that of attempting to isolate them by prompting lexical access without providing an auditory or orthographic percept. Building on the use by Ota et al. (2009) of a semantic relatedness task with orthographic prompts (e.g., *LOCK-key* vs. *ROCK-key*), similar tasks could be used in which decisions are triggered by pictures instead. While this would add a layer of complexity to selecting the appropriate materials and would limit the researcher's choices in that regard, a robust design would be able to assess learners' phonological content for L2 words without exposure to any acoustic signal nor the involvement of potentially confounding orthographic representations. For example, judgments of phonological similarity, rather than semantic relatedness, between words presented through pictures (e.g., *LOCKER - ROCKET*) would be able to shed new light on phonolexical representations circumventing the recurrent concerns about lower-level processing.

Secondly, an alternative to minimizing the role of perceptual processes is to try to control for them as much as possible in the experimental design. This could be done by designing materials that are not only concerned with L2 phonological contrasts that are

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

known to be challenging for a given population but also contain carefully chosen control stimuli targeting distinctions that are not expected to be problematic from a perceptual standpoint (e.g., contrasts that also exist in the L1 and with similar phonetic implementation). While some studies have already used these "easier" L2 contrasts as a general baseline (e.g., John & Frasnelli, 2022; Llompart & Reinisch, 2019a, 2019b), we still lack proper comparisons between how learners perform in lexical tasks with vs. without the perceptual difficulties that are expected to contribute to learners' substantial uncertainty in these tasks. These comparisons would allow us to determine the extent to which variation in performance for the easier contrasts can capture the variation observed for the challenging L2 distinctions, which could shed much light on the perception–representation divide in non-native spoken word recognition. This is because a large amount of shared variance would mostly point to shared representational issues that go beyond particular sounds or features (i.e., to global fuzziness as conceived in the section Global and local representational "fuzziness") whereas divergences between the two could be used to refine our predictions regarding perceptual challenges and representational imprecisions (i.e., local fuzziness) for the contrast of interest.

Furthermore, the methods we are commonly using can also be fine-tuned to answer or attempt to answer other questions that extend beyond the duality perception–representation that has been the focus of this section thus far. For instance, very much related to the discussion in the previous paragraph, including materials targeting a variety of contrasts could also lead to a more comprehensive and informative account of the ways in which fuzziness permeates the L2 lexicon. As discussed in the Terminology section, fuzziness has been conceptualized at two different scales, but work assessing the way in which these two interact is lacking. Word and nonword responses in lexical decision tasks with different types of substitutions, or assessments of lexical competition dynamics for targets and competitors with different degrees of phonological overlap could very well further our understanding of the interplay between global and local fuzziness in L2 word recognition, as well as of the type of phonolexical fuzziness that we encounter in each case (i.e., the different scenarios in Table 1).

Likewise, care and creativity in designing materials and adapting experimental procedures should also allow us to approach the encoding of phonological categories in the lexicon as a dynamic process that takes place in an ever-developing system rather than as just a monolithic property of all the words or stimuli items to the same extent. To do so, more emphasis needs to be placed on researching how the item-level properties that situate words within the lexicon, such as lexical frequency, phonological neighborhood density, cognate, and loanword status (considering the learners' L1), or the presence or absence of particular minimal pairs for that word affect lexical competition and selection. Even though there is some preliminary, albeit promising, work in this direction (Darcy & Thomas, 2019; Llompart, 2021b; Rocca, Llompart, & Darcy, accepted), this area remains underexplored despite its potential, likely because of its complexity. Increasing the amount of research devoted to it can play a crucial role in continuing to make our field move forward in the following years.

Finally, while behavioral methods like those commonly found in the literature present the obvious advantages that they are well-understood, cost-effective, and relatively widely accessible, the incorporation of other methodologies to the study of L2 phonolexical representations can complement and extend current findings in important ways. In particular, neurophysiological methods may provide critical insights into the contributions of perceptual processes and imprecisions in lexical representation to the challenges that L2 learners experience in spoken word recognition.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Electroencephalography (EEG) can precisely measure *when* the brain responds, in the form of event-related potentials (ERPs), to an auditory stimulus (affording high temporal resolution) and how strong this reaction is. This way, early responses, which are likely to result from the perception system, can be disentangled from later responses, which reflect lexical access or semantic and syntactic integration (e.g., White, Titone, Genesee, & Steinhauer, 2017; Wagner, Shafer, Martin, & Steinschneider, 2012). The oddball paradigm (e.g., Näätänen, 2001; Dehaene-Lambertz, Dupoux, & Gout, 2000; Mah, Goad, & Steinhauer, 2016) is nowadays widely used to distinguish between perception and lexical access processes (by means of the mismatch negativity) and in which listeners are presented with a series of identical stimuli, followed by a slightly different stimulus. Event-related potentials can also be recorded while listeners are presented with words in full sentences or even in long stretches of speech (e.g., Bentum, ten Bosch, van den Bosch, & Ernestus, 2022). This way, we can approximate more natural listening conditions, in which listeners do not perform a metalinguistic task, and study lexical access processes vs. later processes, including semantic and syntactic integration. EEG can also indicate which brain region produces the responses, providing additional information about the underlying processes involved. EEG recordings are sometimes combined with other experimental paradigms, including the visual world paradigm, so that neural responses can be connected to behavioral measures, providing additional opportunities to study the mechanisms underlying speech comprehension (e.g., Mulder, Brand, Boves, & Ernestus, 2024). Finally, in addition to analyzing the EEG signal for event-related potentials, several studies focus on the energy in different oscillation bands (e.g., alpha, beta, gamma, etc.), which have been argued to reflect specific processes involved in speech comprehension (e.g., for an overview, see Meyer, 2018).

While many studies investigate non-native speech processing using electroencephalography, fewer studies make use of magnetic resonance imaging (MRI). This technique has the advantage that it can precisely locate the brain regions that are involved in processing the stimuli, which may help to disentangle perception processes from lexical access. Its main disadvantage, however, is that its temporal resolution is much lower than in EEG or ERPs. Moreover, participants have to lie still in a large noisy scanner, which makes the technique invasive, expensive, and the listening conditions suboptimal.

In addition to obtaining a clearer fundamental understanding of perceptual, representational, and access processes, thoroughly investigating how these neural responses change and interact as a function of the way in which *speaker*-related (e.g., L2 proficiency, age) and *item*-related factors (e.g., phonological contrast addressed, lexical frequency of the items) are manipulated may lead to significant breakthroughs in our knowledge (e.g., Song & Iverson, 2018; Mulder, Wloch, Boves, ten Bosch, & Ernestus, 2022).

Lexical representations

A third enduring challenge pertains to the format of lexical representations. Answers to our questions about the mechanisms underlying L2 word recognition depend on the assumptions that we make about how each word's pronunciation is stored in the mental lexicon. The nature of these representations is a hotly debated issue in the literature on L1 listening as well. The various frameworks and positions form a continuum along the dimension of the assumed abstractness (from most to least) of lexical representations: Positions at the most abstract end of the continuum assume that exactly one pronunciation is stored for each word, in the form of a string of abstract phonemes. For instance, the English word *police* is often assumed to be represented as /pəlis/, even though it is frequently pronounced like [plis], with a strongly reduced or absent vowel.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

A few exceptional, mostly highly frequent, words may be lexically represented with more than one pronunciation variant. This view is incorporated in most computational models of human word recognition, including TRACE (McClelland & Elman, 1986) and Shortlist B (Norris & McQueen, 2008). According to this view, there may be three reasons why (non-native) listeners may not recognize words.

First, they might misidentify the sounds in the speech signal, for instance by thinking they hear an /l/, whereas the lexical representation contains an /r/, not an /l/, thus creating a mismatch. Note that listeners may not be able to *identify* a sound (even though they can perceptually distinguish it from other sounds) for various reasons: For instance, the sound may be realized differently in their native language (e.g., always with prevoicing) than in the foreign language (seldom with prevoicing); or it does not occur in their native language (e.g., French /y/ for L1-English listeners). Second, as we extensively demonstrated above, the listener's lexical representation of the word's pronunciation may be imprecise, noncontrastive, or incomplete, for instance, because it is based on the spelling of the word rather than its pronunciation or because it is based on an earlier misperception of the sounds. This situation might also trigger a mismatch. Third, listeners may be unable to map the phonemes they have identified on the word's lexical representation during lexical access. This may occur if the pronunciation variant they hear is not the one that is stored in their mental lexicons. For instance, if the English word *police* is pronounced as [plis], non-native listeners who are not familiar with schwa reduction from their native language, may have problems mapping [plis] on /pəlis/ (see the section Comprehending speech in everyday situations).

Somewhat further along the continuum of representational abstractness is the assumption that for each word, several pronunciation variants may be stored in the mental lexicon, still in the form of strings of abstract phonemes. According to this view, the English word *police* would be stored (at least) as /pəlis/ and /plis/. This assumption can explain why native and non-native listeners are sensitive to the frequencies of occurrences of the pronunciation variants: The more frequent a variant, the better the pronunciation is entrenched in the mental lexicon, and the easier it can be accessed during word recognition (e.g., Ranbom & Connine, 2007; Brand & Ernestus, 2018). Non-native listeners may not recognize a word (pronunciation variant) for the same reasons as mentioned above but also because their lexicon does not yet contain a lexical representation for the pronunciation variant presented.

Other theories assume that a word's pronunciation is stored with more phonetic detail, and is less abstract. The lexical representation of a word would not just contain phonemes, but also information about how these phonemes are pronounced in the given word (including allophonic detail). For instance, the lexical representation of a word like *police* may indicate the initial plosive's typical voice onset time (VOT), or the VOT distribution. This assumption can thus explain why the same phoneme may be pronounced differently depending on the word in which it occurs (Pierrehumbert, 2002; Tang & Shaw, 2021). According to this theory, non-native lexical representations can be suboptimal not only because they contain the wrong phonemes but also because the phonetic detail specified is incorrect or not sufficiently specified. Phonetic specifications may be incorrect, for instance because they are not only based on the word's pronunciation in the L2, but also on how the phonemes are typically pronounced in the L1, or because the specifications are based on too few tokens of the words to be representative. As a result, lexical representations are misaligned with the input and will hinder the mapping of the acoustic signal.

At the most detailed (least abstract) end of the continuum, exemplar-based theories assume that a language user mentally stores every token of a word, produced or

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

perceived by the language user, in full phonetic detail (e.g., Craik & Kirsner, 1974; Palmeri, Goldinger, & Pisoni, 1993). These mental representations would be faithful representations of the acoustic realizations of the tokens. Thus, non-native listeners' exemplars would not be internally influenced by the properties of their native language (e.g., Morano, ten Bosch, & Ernestus, 2019). Clouds of exemplars form the word's phonolexical representation, on which word recognition is based. Exemplar-based theories can account for the experimental finding that both native and non-native listeners respond more quickly to the second token of a word when it is acoustically more similar to the first token. However, the extensive literature on native language influence when perceiving speech and within representations presents a challenge to these theories, which need to explain how the native language affects how these faithful lexical representations are accessed (see Goldrick & Cole, 2023 for a recent overview).

Several studies suggest that listeners may rely on both abstract lexical representations and exemplars, and researchers have proposed so-called hybrid models of the lexicon (e.g., Church & Schacter, 1994, McLennan & Luce, 2005). Whether listeners rely more on exemplars or abstract lexical representations for the recognition of a given word token may depend on the time that is needed to recognize the token (McLennan & Luce, 2005) or on the cognitive load involved in the task (Nijveld, ten Bosch, & Ernestus, 2022), which may differ between native and non-native listeners.

Possibly, auditory word recognition may even involve more memory systems, that is, collections of lexical representations that differ in their abstractness. Hawkins and Blakeslee (2004) developed a theory of the brain in which all positions along the abstractness continuum may be incorporated. They claim that, in addition to episodic memory, there are six layers in the neocortex in which information is stored. The layers differ from each other in the abstractness of the representations, with the top level being the most abstract. A theory of auditory word recognition based on the proposal by Hawkins and Blakeslee could assume that a word token, once it is recognized, is first stored in episodic memory, where it is fully specified (exemplar), and is incorporated later on in one or more of the neocortex layers, by contributing to the increasingly more abstract representations stored in each layer. During word recognition, listeners may have access to all layers, but the type of task may make them focus on some of them. Such a framework seems a promising way to unify the different abstraction levels reported across studies. Evidence from non-native listeners will certainly be crucial to help the precise formulation of this theory.

This latter theory of how the pronunciation of words is stored holds great promise because it incorporates all the hypotheses proposed so far while being firmly grounded in neurocognitive findings; however, more research is clearly needed to outline this new approach more precisely. Until then—or until it is clear that another theory of lexical representations has to be preferred—we urge researchers on non-native speech production and perception to be aware that the pronunciation of words may be simultaneously stored with varying degrees of abstractness. It is therefore important to explicitly describe the abstractness level we assume to be relevant for each non-native speech phenomenon that we study. Our results may be accounted for differently depending on the assumptions we make about the nature of the lexical representations involved.

Two forward-looking research avenues

We now turn to two fruitful research directions at the forefront of this line of research. The first addresses the specific challenges connected to bottom-up lexical access when listening to casual, conversational speech, and how our emerging knowledge of L2

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

lexical representations can be informed by a deeper understanding of these questions. The second one takes us to the language classroom: It explores how our current understanding of the L2 mental lexicon can inform instructional approaches, and in turn, to what extent these approaches have the potential to help learners optimize their lexical representations.

Comprehending speech in everyday situations

Most of the research outlined above on how well, and how, non-native listeners recognize words in L2 is based on highly controlled experiments performed under laboratory conditions, usually with speech presented in citation form. This level of control in the materials leaves open the question of how well, and how exactly, non-native listeners comprehend speech and recognize words in everyday situations, with background noise or while the listener is also performing another task (e.g., driving a car). We know little about non-native speech comprehension in real-life situations although the answer to this question is relevant, both for theories of L2 speech processing and for language teaching.

In addition to background noise and secondary tasks that may increase processing load, everyday conversations pose another challenge to non-native listeners. The single pronunciation that learners typically acquire in class for every word (i.e., the word's citation form; e.g., McCarthy & Carter, 1995; O'Connor Di Vito, 1991) does not prepare them for the wide range of variable pronunciations they will encounter in everyday situations. In any language, the pronunciation of words can vary substantially depending on a speaker's regiolect, social group, generational or educational background, gender, mood, physical health, emotions, affect, or situation (Moyer, 2013). In addition, in any spontaneous conversation, even a single speaker tends to vary their pronunciation of the same word, with these variants differing in how much they deviate from the word's citation form. Specifically, these variants might have segments that are weakly articulated or acoustically completely absent. For instance, the Dutch word *natuurlijk* ('of course') is rarely pronounced in full (/natyrlək/); rather, it appears with variable pronunciations anywhere between the citation form and strongly reduced [ty] or [dy] (including, among other forms, the frequent variant [tyk]). Corpus research in Dutch, English, and French has shown that, on average, at least 25% of the word tokens in conversational speech deviate from the words citation pronunciations in at least a single sound, whereas approximately 6% of word tokens contain at least one syllable less (Schuppler, Ernestus, Scharenborg, & Boves, 2011; Johnson, 2004; Adda-Decker, de Mareüil, Adda, & Lamel, 2005). Native listeners easily recognize reduced pronunciation variants when they are embedded in conversational speech (e.g., Ernestus, Baayen, & Schreuder, 2002). In contrast, many dictation tasks have shown that non-native listeners typically experience great difficulties recognizing such reduced variants, even if the deviation between the reduced variant and the citation pronunciation is small (just a schwa) and the learners are highly proficient (e.g., Brand & Ernestus, 2018). This raises the question of why this is the case.

Previous research suggests that non-native listener's difficulties with reduced variants are related to lower levels of experience with these variants. For instance, Brand and Ernestus (2018) suggest that the speed with which listeners perform lexical decisions for French words lacking a schwa (e.g., /rvy/, compared with the citation form /rəvy/ for *revue*) is correlated with how often they think this pronunciation variant occurs for the given word. The listener's *experience* with a reduced pronunciation variant appears to be more relevant than the phonetic or phonological distance between

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

the variant and the word's citation form (Brand & Ernestus, 2019). However, a higher frequency of occurrence can only facilitate recognition if listeners have lexical representations for these pronunciation variants in the first place, or know how to determine which citation form they are variants of. This is not the case for many reduced variants that non-native listeners do not recognize at all.

Possibly, non-native listeners have difficulties recognizing reduced words because the reduction pattern deviates from those in their L1. For instance, while French tends to reduce vowels, Spanish tends to reduce consonants (Torreira & Ernestus, 2011). As a consequence, native listeners of Spanish may have problems reconstructing the many reduced vowels in their French L2 (see Darcy, Peperkamp, & Dupoux, 2007 for evidence of language-specific reconstruction phenomena). Whether this is indeed the case for reductions, and how we could train non-native listeners on the reduction patterns of the foreign language (e.g., Kennedy & Blanchet, 2014) is an exciting task for future research.

The reduction of segments may leave subtle traces in the acoustic signal, which may also differ among languages. For instance, in German, the absence of the verbal affix /t/ (e.g., *fliehst* 'flee 2nd PERS. SG.' pronounced as /fli:s/) does not lead to a substantial reduction in the duration of the word. The word's duration may therefore be an acoustic cue to reconstruct an absent /t/. Zimmerer and colleagues (2014) showed that this is indeed the case for native listeners of German in simple psycholinguistic experiments (Zimmerer & Reetz, 2014). While the authors used authentic speech taken from a natural corpus, the question is still open to what extent native listeners may rely on subtle acoustic cues also in nonlaboratory, real-life situations. With respect to non-native listeners, we have to investigate whether they can perceive the subtle acoustic cues and, if so, to what extent they can then use these cues in understanding reduced speech.

Native listeners are able to rely on the semantic and syntactic context when interpreting reduced variants presented in sentences (Ernestus et al., 2002). Also, the meanings of reduced variants may facilitate the interpretation of upcoming speech, although semantic priming from reduced variants seems to take more time than from citation forms (van de Ven, Tucker, & Ernestus, 2011). Non-native listeners appear to benefit less from the context provided by a sentence, both for speech in noise (Bradlow & Alexander, 2007) and for recognizing reduced variants (van de Ven, Tucker, & Ernestus, 2010). Of course, this raises the question of *why* sentence contexts are less beneficial in L2, and how we can train learners to make better use of the available context.

In sum, previous research on the comprehension of non-native speech outlines what non-native listeners can do in ideal listening situations. To understand how they process L2 speech in real life outside of the laboratory, more research is needed. This research will have to take the fact into account that in conversational speech, words appear in many variable forms, and typically not (or rarely) in the citation form that we teach in the classroom.

The theoretical question of how non-native listeners process conversational speech is crucial to more fully understanding how pronunciation variants are lexically stored in general (see the previous section). Is there only a single, abstract representation for the word's citation form in the L2 mental lexicon, or are several pronunciation variants stored, and with how much detail? Non-native listeners' speech comprehension may thus open up a window to the general structure of the L2 mental lexicon.

Instructional approaches to representational precision

Given the manifold difficulties connected to lexical representations and access in L2 reviewed so far in this article, and in view of the potential impact that classroom

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

interventions could have on reducing these difficulties, it is crucial to develop research in this area. Learners may partially overcome L1-based processing and establish accurate L2 lexical representations through extended exposure to the L2 (e.g., Gorba & Cebrian, 2021), but doing so in foreign language instructional contexts remains a challenge for several reasons, such as fewer opportunities for meaningful L2 use in authentic contexts, or the lack of lexically–oriented pronunciation instruction (Tyler, 2019). Helping learners establish precise phonological representations, both for new and already known words, should become a central goal of pronunciation instruction (Darcy, 2018) because it may enhance L2 pronunciation development and lead to benefits in speech intelligibility and comprehensibility. Yet, these potential benefits first need to be investigated systematically, which requires identifying the types of phonetic training/pronunciation intervention that are most effective in targeting the precision of representations. This gap in our knowledge is partly due to the fact that (1) most pronunciation instruction methods operate mainly at the phonetic and phonological level, and (2) the lexical level has not been included in the measures of effectiveness. Take for example the high-variability phonetic training (HVPT) paradigm, which leads to well-attested gains in phonetic and phonological perception and production (Suzukida & Saito, 2021; Barriuso & Hayes-Harb, 2018; Sakai & Moorman, 2018; Thomson, 2011; 2018). This paradigm's effectiveness at improving the lexical encoding of phonological contrasts still needs broader empirical evaluation (see Melnik & Peperkamp, 2021). A systematic investigation of HVPT's effectiveness range would greatly advance our knowledge in this area, especially if studies account for differences in the characteristics of training materials (e.g., which specific targets; training with words vs. nonwords; Mora, Ortega, Mora-Plaza, & Aliaga-García, 2022), and for individual differences in cognition (such as attention or memory) across learners. The same is true of other techniques that draw learners' attention to phonetic form (often implicitly), such as shadowing (Foote & McDonough, 2017), foreign accent imitation (Henderson & Rojczyk, 2023; Mora, Rochdi, & Kivistö-de Souza, 2014), and exposure to audiovisual input through captioned video (Galimberti, Mora, & Gilabert, 2023; Wisniewska & Mora, 2020) or embodied pronunciation training (Bailly, Alazard-Guiu, & Prieto, 2022). While all show promise in enhancing learners' phonetic awareness and possibly their phonological processing, the extent to which these improvements can benefit phonolexical representations long term remains to be evaluated.

We propose here that the lexical level be included in more ways than "merely" the outcome measures, namely also in the instructional activities themselves. Integrating a variety of pronunciation learning tasks that tap into different processing levels (phonetic, phonological, lexical) could help learners gradually develop more precise phonolexical representations. Indeed, L2 learners' pronunciation development partly depends on their ability to achieve speech processing efficiency at the phonetic (level of perception of categories, sounds), phonological (being able to identify units such as syllables, phonemes, not necessarily with meaning), and lexical levels (words). Thus, pronunciation instructors might want to consider not only which target phonological features learners need to focus on, but also which pedagogical tasks might enhance L2 speech processing at each level—including the lexical level. For example, including a communicative component in the design of a task targeting a phonological contrast (for instance, in Task-Based Pronunciation Teaching or the Automaticity in Communicative Contexts of Essential Speech Segments (ACCESS) framework, Trofimovich & Gatbonton, 2006; Gatbonton & Segalowitz, 2005) may enhance learners' lexical processing compared with *non*communicative training contexts such as HVPT. Before this

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

can be fully implemented in instruction, however, there is an urgent need to investigate which pronunciation instruction methods are able to tap the lexical level and effectively impact the precision of developing phonolexical representations.

Another promising direction for research pertains to the sequencing of instructional activities (tasks), within and across the three levels, according to L2 learners' skills. Many studies suggest combining explicit and communicative instruction (see Saito, 2012; Saito, 2015; Darcy & Rocca, 2022), but much remains to be investigated in terms of how tasks interact with proficiency. For instance, phonetic–level tasks may need to be prioritized at beginner levels, before using lexical–level tasks, but systematic studies do not yet exist to support one or the other. In addition, it may be worthwhile to take into account individual differences in proficiency and vocabulary size. For example, in foreign language learning contexts, large vocabularies may develop with imprecise encoding of L2 contrasts (Tyler, 2019). How many words a learner knows could determine how malleable these lexical representations are because smaller vocabularies can function with less precision (Daidone & Darcy, 2021), and therefore, they exert less pressure to increase the precision of representations. This, in turn, could affect which interventions may benefit learners the most, depending on individual differences in vocabulary size.

To sum up, when designing a pronunciation intervention, the type of tasks, the processing level they tap, and their sequencing need to be considered, taking into account the learner's L1 background, pronunciation skills, and vocabulary size. More research examining how these aspects all interact is needed, and ideally should be conducted in classrooms and in integrated curricula (that is, not in separate pronunciation–specific courses), for instance, as was recently done by Mora-Plaza (2023).

Conclusion

In this paper, we have reviewed what the past 25 years have taught us about the form of lexical representations in the L2 mental lexicon, and in particular about the phonological form that is stored in these representations, and accessed during spoken word recognition. Much progress was made in uncovering the fragility of both access and representations and the part they play in the perception and production of speech. Many insights regarding the (dis)connections between perception, lexical encoding, and production are starting to form an overall picture of the unique difficulties that creating functional and precise phonolexical representations pose for learners. In addition, these findings challenge assumptions that the *content* of phonolexical representations and the *access processes* to activate them are the same. Finally, research has also begun to uncover the role played by lexical neighborhoods and the larger structure of the L2 mental lexicon in promoting the creation of precise and contrastive representations.

Our paper also discussed several methodological and theoretical issues, together with enduring challenges in this research area, for which increased research should soon provide some answers. An area that we hope this paper will contribute to clarify is the terminology researchers use to refer to imprecise phonolexical representations. Furthermore, our paper calls for the need to connect the L2 data more explicitly to findings on the structure of the lexicon emerging from L1 research.

Finally, two areas with a crucial need for additional research were outlined. In particular, the findings we reviewed clearly highlight an urgent need to extend the research done in the laboratory (using items presented in citation form under optimal listening conditions) to research outside the lab, in more natural conditions, and with

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

speech materials that are more authentic in all respects. Only then will we be able to fully understand the true extent of what learners in fact face in daily life. This should also enable us to devise training paradigms and pedagogical approaches that are more effective in helping learners modify their representations and streamline the processing of their new language, thus helping them communicate more easily in the second language.

Acknowledgments. This article emerged from a small workshop organized in May 2022 at the Institut Méditérranéen de Recherches Avancées (IMéRA, Aix-Marseille Université, France), which aimed to provide an overview of the most recent developments in the field. The authors thankfully acknowledge the support and funding from IMéRA in organizing the workshop, and from the following sources: I.D. was supported by a fellowship from the Institute of Language, Cognition and the Brain (France) "Langage et cerveau" (2021-2022); J.C.M.'s contribution to this work was partly supported by research grants PID2022-138129NB-I00 and 2023SGR00303 from the Spanish Ministry of Science and Innovation and AGAUR, respectively. M.L.'s contribution was supported by Spanish Ministry of Universities, (Beatriz Galindo Junior Research Fellowship, BG22/00161).

Competing interest. The authors declare none.

References

Adda-Decker, M., de Mareüil, P. B., Adda, G., & Lamel, L. (2005). Investigating syllabic structures and their variation in spontaneous French. *Speech Communication*, *46*(2), 119–139. https://doi.org/10.1016/j.specom.2005.03.006

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*(5), 1221–1251. https://doi.org/10.1017/S0142716415000557

Bailey, A. A., & Brandl, A. (2013). Incorporating pronunciation in the first-year Spanish classroom: An early intervention. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th pronunciation in second language learning and teaching conference* (pp. 207–223). Iowa State University Digital Press.

Bailly, F., Alazard-Guiu, C., & Prieto, P. (2022). Embodied prosodic training helps improve accentedness and suprasegmental accuracy. *Applied Linguistics*, *43*(4), 776–804. https://doi.org/10.1093/applin/amac010

Barrios, S., & Hayes-Harb, R. (2021). L2 processing of words containing English /æ/-/ɛ / and /l/-/ɹ / contrasts, and the uses and limits of the auditory lexical decision task for understanding the locus of difficulty. *Frontiers in Communication*, *6*(144). https://doi.org/10.3389/fcomm.2021.689470

Barriuso, T. A., & Hayes-Harb, R. (2018). High variability phonetic training as a bridge from research to practice. *CATESOL Journal*, *30*(1), 177–194.

Bentum, M., ten Bosch, L., van den Bosch, A., & Ernestus, M. (2022). Speech register influences listeners' word expectations. *Brain and Language*, *235*, 105197.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, *20*(3), 305–330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second language speech perception. Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 273–304). York Press.

Bordag, D., Gor, K., & Opitz, A. (2021). Ontogenesis model of the L2 lexical representation. *Bilingualism: Language and Cognition*, 1–17. https://doi.org/10.1017/S1366728921000250

Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2339–2349. https://doi.org/10.1121/1.2642103

Brakovec, H., & Darcy, I. (2023, September). *Portuguese vowel contrasts are more precisely lexically encoded for learners with greater orthographic awareness* [Conference presentation]. 14th Pronunciation in Second Language Learning and Teaching Conference, West Lafayette, IN, USA.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Brand, S., & Ernestus, M. (2018). Listeners' processing of a given reduced word pronunciation variant directly reflects their exposure to this variant: Evidence from native listeners and learners of French. *Quarterly Journal of Experimental Psychology*, *71*(5), 1240–1259.

Brand, S., & Ernestus, M. (2019). Understanding reduced words. The relevance of reduction degree and frequency of occurrence. In S. Calhoun & P. Escudero (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 97–81). Australasian Speech Science; Technology Association Inc. https://repository.ubn.ru.nl/handle/2066/208204

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, *36*, 22–34. https://doi.org/10.1016/j.system.2007.11.003

Cabrelli, J., Luque, A., & Finestrat-Martínez, I. (2019). Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *Journal of Phonetics*, *73*, 55–69. https://doi.org/10.1016/j.wocn.2018.10.006

Chan, M. J. (2018). Embodied pronunciation learning: Research and practice. *CATESOL Journal*, *30*(1), 47–68.

Cheng, L. S., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology*, *12*, 715843.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(3), 521–533.

Cook, S. V. (2012). *Phonological form in L2 lexical access: Friend or foe?* (Unpublished doctoral dissertation). University of Maryland, College Park, MD. http://hdl.handle.net/1903/12708

Cook, S. V., & Gor, K. (2015). Lexical access in L2: Representational deficit or processing constraint? *The Mental Lexicon*, *10*(2), 247–270. https://doi.org/10.1075/ml.10.2.04coo

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*, *7*, 1345. https://doi.org/10.3389/fpsyg.2016.01345

Craik, F. I., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *The Quarterly Journal of Experimental Psychology*, *26*(2), 274–284.

Cutler, A., & Otake, T. (2004). Pseudo-homophony in non-native listening. *The Journal of the Acoustical Society of America*, *115*(5), 2392–2392. https://doi.org/10.1121/1.4780547

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, *34*(2), 269–284. https://doi.org/10.1016/j.wocn.2005.06.002

Daidone, D., & Darcy, I. (2021). Vocabulary size is a key factor in predicting second language lexical encoding accuracy. *Frontiers in Psychology*, *12*, 2769. https://doi.org/10.3389/fpsyg.2021.688356

Darcy, I. (2018). Powerful and effective pronunciation instruction: How can we achieve it? *CATESOL Journal*, *30*(1), 13–45.

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

Darcy, I., Dekydtspotter, L., Sprouse, R. A., Glover, J., Kaden, C., McGuire, M., & Scott, J. H. G. (2012). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research*, *28*, 5–40. https://doi.org/10.1177/0267658311423455

Darcy, I., & Holliday, J. J. (2019). Teaching an old word new tricks: Phonological updates in the L2 mental lexicon. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 10–26). Iowa State University Digital Press.

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules: Perceptual compensation for assimilation in late L2-learners. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411–442). Mouton de Gruyter.

Darcy, I., & Rocca, B. (2022). Comprehensibility improvements in integrated pronunciation instruction: A comparison of instructional methods and task effects. *Journal of Second Language Pronunciation*, *8*(3), 328–362.

Darcy, I., & Thomas, T. (2019). When blue is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition*, *22*(5), 1141–1159. https://doi.org/10.1017/S1366728918001050

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*(4), 635–647.

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first-and second-language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, *66*(5), 843–863.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578. https://doi.org/10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, *36*, 406–421.

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*, 682–706. https://doi.org/10.1016/j.cognition.2007.04.001

Eger, N. A., & Reinisch, E. (2019). The role of acoustic cues and listener proficiency in the perception of accent in nonnative sounds. *Studies in Second Language Acquisition*, *41*(1), 179–200.

Elvin, J. D. (2016). *The role of the native language in non-native perception and spoken word recognition: English vs. Spanish learners of Portuguese*. (Unpublished doctoral dissertation). Western Sydney University, Sydney, NSW, Australia.

Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*(1-3), 162–173.

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360. https://doi.org/10.1016/j.wocn.2007.11.002

Fitzpatrick, J., & Wheeldon, L. (2000). Phonology and phonetics in psycholinguistic models of speech perception. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge: Conceptual and empirical issues* (pp. 131–160). Oxford University Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press.

Foote, J. A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1), 34–56.

Galimberti, V., Mora, J. C., & Gilabert, R. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System*, *116*, 103078.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review*, *61*(3), 325–353. https://doi.org/10.3138/cmlr.61.3.325

Goh, C. C. M. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, *28*(1), 55–75. https://doi.org/10.1016/S0346-251X(99)00060-3

Goldrick, M., & Cole, J. (2023). Advancement of phonetics in the 21st century: Exemplar models of speech production. *Journal of Phonetics*, *99*, 101254.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814.

Gor, K., Cook, S., Bordag, D., Chrabaszcz, A., & Opitz, A. (2022). Fuzzy lexical representations in the nonnative mental lexicon. *Frontiers in Communication*, *7*, 1027692.

Gorba, C., & Cebrian, J. (2021). The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal of Phonetics*, *88*, 101094. https://doi.org/10.1016/j.wocn.2021.101094

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

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. Macmillan.

Hayes-Harb, R., & Barrios, S. (2019). Investigating the phonological content of learners' "fuzzy" lexical representation for new L2 words. In J. M. Levis, C. L. Nagle, & E. Todey (Eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference* (pp. 55–69).

Hayes-Harb, R., & Masuda, K. (2008). Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research*, *24*(1), 5–33. https://doi.org/10.1177/0267658307082980

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381. https://doi.org/10.1177/0023830910371460

Henderson, A., & Rojczyk, A. (2023). Foreign-language accent imitation: Matching production with perception. In V. Sardegna & A. Jarosz (Eds.), *English pronunciation teaching: Theory, practice, and research findings* (pp. 102–117). Multilingual Matters. https://doi.org/10.21832/9781800410503-013

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *Journal of the Acoustical Society of America*, *119*(5), 3072–3084. https://doi.org/10.1121/1.2184289

Hulstijn, J. (2001). Intentional and incidental second-language vocabulary learning. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*(1), B47–B57. https://doi.org/10.1016/S0010-0277(02)00198-1

John, P., & Frasnelli, J. (2022). On the lexical source of variable L2 phoneme production. *The Mental Lexicon*, *17*(2), 239–276. https://doi.org/10.1075/ml.22002.joh

Johnson, K. (2004). Massive reduction in conversational American English. In K. Yoneyama & K. Maekawa (Eds.), Casual speech: Data and analysis. *Proceedings of the 1st session of the 10th International Symposium* (pp. 29–54). National International Institute for Japanese Language.

Jusczyk, P. W., & Luce, P. A. (2002). Speech perception and spoken word recognition: Past and present. *Ear and Hearing*, *23*(1), 2–40.

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, *50*(1), 23–52. https://doi.org/10.1177/00238309070500010201

Kennedy, S., & Blanchet, J. (2014). Language awareness and perception of connected speech in a second language. *Language Awareness*, *23*(1-2), 92–106. https://doi.org/10.1080/09658416.2013.863904

Lahiri, A., & Reetz, H. (2010). Distinctive features: Phonological underspecification in representation and processing. *Journal of Phonetics*, *38*(1), 44–59. https://doi.org/10.1016/j.wocn.2010.01.002

Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, *88*, 102185.

Li, P., Xi, X., Baills, F., & Prieto, P. (2021). Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters. *Language, Cognition and Neuroscience*, *36*(10), 1313–1328.

Llompart, M. (2021a). Lexical and phonetic influences on the phonolexical encoding of difficult second-language contrasts: Insights from nonword rejection. *Frontiers in Psychology*, *12*, 659852. https://doi.org/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. *Bilingualism: Language and Cognition*, *24*(3), 481–496. https://doi.org/10.1017/S1366728920000656

Llompart, M., & Reinisch, E. (2017). Articulatory information helps encode lexical contrasts in a second language. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(5), 1040–1056. https://doi.org/10.1037/xhp0000383

Llompart, M., & Reinisch, E. (2019a). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*, *62*(3), 594–622.

Llompart, M., & Reinisch, E. (2019b). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, *22*(5), 1085–1100.

Llompart, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1590–1610. https://doi.org/10.1037/xlm0000832

Llompart, M., & Reinisch, E. (2021). Lexical representations can rapidly be updated in the early stages of second-language word learning. *Journal of Phonetics*, *88*, 101080. https://doi.org/10.1016/j.wocn.2021.101080

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Mah, J., Goad, H., & Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]. *Frontiers in Psychology*, *7*, 1469.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

McAllister, R., Flege, J. E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*(2), 229–258.

McCarthy, M., & Carter, R. (1995). Spoken grammar: What is it and how can we teach it? *ELT Journal*, *49*(3), 207–218.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McLennan, C. T., & Luce, P. A. (2005). Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 306–321.

Melnik, G. A., & Peperkamp, S. (2021). High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition*, *24*(3), 497–506.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621.

Mora, J. C., Ortega, M., Mora-Plaza, I., & Aliaga-García, C. (2022). Training the pronunciation of L2 vowels under different conditions: The use of nonlexical materials and masking noise. *Phonetica*, *79*(1), 1–43.

Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness*, *23*(1-2), 57–75.

Morano, L., ten Bosch, L., & Ernestus, M. (2019). Looking for exemplar effects: Testing the comprehension and memory representations of r'duced words in Dutch learners of French. In S. Fuchs, J. Cleland, & A. Rochet-Capellan (Eds.), *Speech perception and production: Learning and memory* (pp. 245–277). Peter Lang.

Mora-Plaza, I. (2023). *Task-based pronunciation teaching and learning of L2 vowels in efl learners: Task complexity effects*. (Unpublished doctoral dissertation). University of Barcelona, Barcelona, Spain. http://hdl.handle.net/2445/201879

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge University Press. https://doi.org/10.1017/CBO9780511794407

Mulder, K., Brand, S., Boves, L., & Ernestus, M. (2024). Processing reduced speech in the L1 and L2: A combined eye-tracking and ERP study. *Language, Cognition and Neuroscience*, *39*(5), 527–551.

Mulder, K., Wloch, L., Boves, L., ten Bosch, L., & Ernestus, M. (2022). Cognate status modulates the comprehension of isolated reduced forms. *Language, Cognition and Neuroscience*, *37*(5), 576–614.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, *38*(1), 1–21.

Nijveld, A., ten Bosch, L., & Ernestus, M. (2022). The use of exemplars differs between native and non-native listening. *Bilingualism: Language and Cognition*, *25*(5), 841–855.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

O'Connor Di Vito, N. (1991). Incorporating native speaker norms in second language materials. *Applied linguistics*, *12*(4), 383–396.

Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition*, *111*(2), 263–269. https://doi.org/10.1016/j.cognition.2008.12.007

Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, *12*(6), 445–449. http://www.jstor.org/stable/40063668

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(2), 309–328.

Perry, S. J., Kelley, M. C., & Tucker, B. V. (2024). Documenting and modeling the acoustic variability of intervocalic alveolar taps in conversational Peninsular Spanish. *The Journal of the Acoustical Society of America*, *155*(1), 294–305.

Pierrehumbert, J. B. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101–140). Mouton de Gruyter.

Polivanov, E. (1931). La perception des sons d'une langue étrangère. *Travaux du Cercle Linguistique de Prague*, *4*, 79–96.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., & Dupoux, E. (2010). A psycholinguistic perspective on the acquisition of phonology. In C. Fougeron, B. Kühnert, M. d'Imperio, & N. Vallée (Eds.), *Laboratory phonology 10: Variation, phonetic detail and phonological representation* (pp. 311–340). Mouton de Gruyter.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273–298.

Rocca, B., Llompart, M., & Darcy, I. (accepted) The effects of phonological neighborhood density, perception, and vocabulary size on L1 Korean-L2 English learners phonolexical encoding of confusable segmental contrasts. *Bilingualism: Language and Cognition*.

Rothgerber, J. R. (2020). *The influence of native language phonotactics on second language lexical representation in Japanese learners of English*. (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly*, *46*(4), 842–854.

Saito, K. (2015). Communicative focus on second language phonetic form: Teaching Japanese learners to perceive and produce English /ɹ/ without explicit instruction. *Applied Psycholinguistics*, *36*(2), 377–409.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*(1), 187–224.

Schmid, H.-J. (2010). Entrenchment, salience, and basic levels. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford Handbook of Cognitive Linguistics* (pp. 117–138). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199738632.013.0005

Schuppler, B., Ernestus, M., Scharenborg, O., & Boves, L. (2011). Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions. *Journal of Phonetics*, *39*(1), 96–109.

Sebastián-Gallés, N. (2005). Cross-language speech perception. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Blackwell Publishing.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*(2), 240–255. https://doi.org/10.1016/j.jml.2004.11.001

Segui, J., Frauenfelder, U., & Hallé, P. (2001). Phonotactic constraints shape speech perception: Implications for sublexical and lexical processing. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 195–208). MIT Press.

Simonchyk, A., & Darcy, I. (2017). Lexical encoding and perception of palatalized consonants in L2 Russian. In M. G. O'Brien & J. Levis (Eds.), *Proceedings of the 8th pronunciation in second language learning and teaching conference* (pp. 121–132). Iowa State University Digital Press.

Song, J., & Iverson, P. (2018). Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*, *179*, 163–170.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., & Nishi, K. (2001). Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. *The Journal of the Acoustical Society of America*, *109*(4), 1691–1704. https://doi.org/10.1121/1.1353594

Suzukida, Y., & Saito, K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, *25*(3), 431–450.

Tang, K., & Shaw, J. A. (2021). Prosody leaks into the memories of words. *Cognition*, *210*, 104601. https://doi.org/10.1016/j.cognition.2021.104601

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744–765.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231.

Torreira, F., & Ernestus, M. (2011). Realization of voiceless stops and vowels in conversational French and Spanish. *Laboratory Phonology*, *2*(2), 331–353.

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *The Modern Language Journal*, *90*(4), 519–535. https://doi.org/10.1111/j.1540-4781.2006.00464.x

Trofimovich, P., & John, P. (2011). Chapter 5. When three equals tree. In P. Trofimovich & K. Mc-Donough (Eds.), *Applying priming methods to L2 learning, teaching and research: Insights from psycholinguistics* (pp. 105–129). John Benjamins.

Tyler, M. D. (2019). PAM-L2 and phonological category acquisition in the foreign language classroom. In A. Nyvad, M. Hejná, A. Højen, A. Jespersen, & M. Sørensen (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 607–630). AU Library Scholarly Publishing Services. https://doi.org/10.7146/aul.322.218

van de Ven, M., Tucker, B. V., & Ernestus, M. (2010). Semantic facilitation in bilingual everyday speech comprehension. *Proceedings of Interspeech 2010*, 1245–1248. https://repository.ubn.ru.nl/handle/2066/86613

van de Ven, M., Tucker, B. V., & Ernestus, M. (2011). Semantic context effects in the comprehension of reduced pronunciation variants. *Memory & Cognition*, *39*, 1301–1316.

Wagner, M., Shafer, V. L., Martin, B., & Steinschneider, M. (2012). The phonotactic influence on the perception of a consonant cluster /pt/ by native English and native Polish listeners: A behavioral and event related potential (ERP) study. *Brain and Language*, *123*(1), 30–41.

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

White, E. J., Titone, D., Genesee, F., & Steinhauer, K. (2017). Phonological processing in late second language learners: The effects of proficiency and task. *Bilingualism: Language and Cognition*, *20*(1), 162–183.

Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599–624.

Xi, X., Li, P., Bailly, F., & Prieto, P. (2020). Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research*, *63*(11), 3571–3585.

Zhang, R., & Yuan, Z.-m. (2020). Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition*, *42*(4), 905–918.

Zimmerer, F., & Reetz, H. (2014). Do listeners recover "deleted" final /t/ in German? *Frontiers in Psychology*, *5*, 735.

Cite this article: Darcy, I., Llompart, M., Hayes-Harb, R., Mora, J. C., Adrian, M., Cook, S., & Ernestus, M. (2025). Phonological processing and the L2 mental lexicon: Looking back and moving forward. *Studies in Second Language Acquisition,* 47: 361–387. https://doi.org/10.1017/S0272263124000482

https://doi.org/10.1017/S0272263124000482 Published online by Cambridge University Press