Frontiers | Research Topics

Formal approaches to multilingual phonology

Edited by
Jennifer Cabrelli, Baris Kabak and
John Archibald

Published in
Frontiers in Language Sciences

ISSN 1664-8714
ISBN 978-2-8325-6749-4
DOI 10.3389/978-2-8325-6749-4

Generative AI statement
Any alternative text (Alt text) provided alongside figures in the articles in this ebook has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

About Frontiers

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

Frontiers journal series

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

Dedication to quality

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

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Formal approaches to multilingual phonology

Topic editors

Jennifer Cabrelli — University of Illinois Chicago, United States
Baris Kabak — Julius Maximilian University of Würzburg, Germany
John Archibald — University of Victoria, Canada

Citation

Cabrelli, J., Kabak, B., Archibald, J., eds. (2025). *Formal approaches to multilingual phonology*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6749-4

Table of contents

frontiers | Frontiers in Language Sciences

TYPE Editorial
PUBLISHED 12 June 2025
DOI 10.3389/flang.2025.1622057

OPEN ACCESS

EDITED AND REVIEWED BY
Guillaume Thierry,
Bangor University, United Kingdom

*CORRESPONDENCE
Barış Kabak
baris.kabak@uni-wuerzburg.de

RECEIVED 02 May 2025
ACCEPTED 12 May 2025
PUBLISHED 12 June 2025

CITATION
Archibald J, Cabrelli J and Kabak B (2025) Editorial: Formal approaches to multilingual phonology. *Front. Lang. Sci.* 4:1622057. doi: 10.3389/flang.2025.1622057

Editorial: Formal approaches to multilingual phonology

John Archibald[1], Jennifer Cabrelli[2] and Barış Kabak[3]*

[1]University of Victoria, Victoria, BC, Canada, [2]University of Illinois Chicago, Chicago, IL, United States, [3]Julius Maximilian University of Würzburg, Würzburg, Germany

KEYWORDS

multilingual phonology, phonological representations, phonological features, post-lexical phonological processes, second language phonology

Editorial on the Research Topic
Formal approaches to multilingual phonology

Examining the nature of multilingual speech has been a vibrant practice in the field of language acquisition, represented by specialist journals, conferences, and prolific research programs around the world. While research on its phonetic and sociocultural properties is abundant, abstract representational issues pertaining to different sound systems coming into contact in the multilingual mind call for special research informed by theories of phonology, typology and learning. We contend that phonological models should be able to also map multilinguals' phonological knowledge and thus shed light on the dynamic nature of crosslinguistic influence across different phonological systems. We have thus defined this Research Topic as including *formal approaches* toward the description and explanation of *multilingual phonology*, which is a cover term we chose in order to include the fields of both L2 and L3 phonology, as well as contact phonology.

In this Research Topic, we have articles that adopt such theories as the contrastive hierarchy, feature geometry, underspecification, and different brands of constraint-based theories such as Stochastic Optimality Theory (StOT) and Harmonic Grammar (HG) with weighted constraints, and test the empirical and theoretical coverage of these at the intersection of language acquisition and crosslinguistic influence (CLI) in a variety of L1-Lx constellations.

Accordingly, our Research Topic addresses such different phonological concepts as features, segments, metrical feet, and post-lexical phonological processes as well as the interaction between segmental and suprasegmental phenomena. Some contributions examine the nature and dynamics of interlanguage phonological grammars by visiting assumptions surrounding the initial state of L2 learning, unlearning categorical processes, full copying of L1 grammar, redeployment of features, and the relationship between perception and production in L2 development, etc. Below, we draw out the implications of each contribution for our Research Topic in three groups: (i) representation and redeployment of features, (ii) post-lexical processes at/and the phonological interfaces, and (iii) methodological implications.

Representation and redeployment of features

Archibald, Flynn, and Nelson all demonstrate that abstract features organized in a dependency hierarchy explain sometimes surprising surface facts. Archibald tackles differential substitution to propose that feature ranking provides an explanation of why in languages which possess both /t/ and /s/ phonemes some would choose /s/ and some /t/ as the "best" substitute for English /θ/. Flynn shows that the presence or absence of the feature [RTR] in a given language is a robust predictor of whether the language can adopt innovative emphatic consonants in language contact situations. Nelson demonstrates that in two languages (English and Spanish) which both lack uvular consonants we see differential performance in their ability to acquire uvular consonants in an additional language (Kaqchikel). The English speakers are able to redeploy their vocalic [RTR] feature to acquire the Kaqchikel uvular consonant which also is represented with an [RTR] feature.

In a similar vein, Yazawa et al. test the empirical and theoretical validity of using phonological features to model L2 perceptual behavior by providing a striking case from a L2 English vowel categorization experiment with L1 Japanese listeners. Employing simulations implemented on the premises of the StOT and the Gradual Learning Algorithm (GLA), they compare a segmental model that maps acoustic cues to segments, and a featural model that maps acoustic cues to features. The featural model correctly accounts for the L1 Japanese listeners' perceptual behavior such that a new category can be formed for an L2 vowel that comprises a structurally ill-formed combination of relevant features in the L1 but not for those neighboring vowels that map to a well-formed L1 feature bundle. While the latter type of L2 vowels are prone to perceptual assimilation to the existing L1 vowel categories, the former vowel is perceived to be a deviant of a similar vowel in the L1 vowel space. The segmental model, however, is not only inadequate to capture this difference, but also performs unrealistically native-like, with the implication that the degree to which a distinct L2 segmental category can be formed depends on the listeners' noticing of the perceptual distinctness of the familiar acoustic cues through copied L1 phonological features.

Finally, Barrientos investigates feature redeployment in L1 Spanish learners of German, focusing on the acquisition of front rounded vowels and tense/lax contrasts. Barrientos finds that learners perform better in discrimination tasks when acquiring a new feature ([+/-tense] than when redeploying an existing one ([+/-round]. However, identification tasks did not show a clear advantage for either contrast. The findings suggest that learners may rely more on salient acoustic cues than on abstract feature restructuring in L2 phonological development.

Post-lexical processes and phonological interfaces

Phonemic categories are subject to context-dependent distinct realizations, whereby contrasts may be neutralized in some languages while maintained in others in the same phonological context. Such conflicting demands on sound alternations may create a learning problem in different L1-Lx pairings. Bárkányi and Kiss approach this understudied aspect of phonological acquisition with a focus on regressive voicing assimilation (RVA), which is categorical in only adjacent obstruents in Hungarian (the L1 of participants), is present in Spanish (L2/L3), where it also extends to sonorant triggers (in the form of presonorant voicing, PSV), and inoperant in English (L2/L3). Production results suggest that Hungarian L1 learners show a strong effect of RVA in both of the non-native languages but do not apply PSV in their Spanish (non-target-like) or English (target-like). Furthermore, the multilingual learners are unable to perceptually distinguish the non-target-like (i.e., the lack of) application of PSV from its target-like application in Spanish. While these results are possibly due to the interlingual classification of Hungarian and Spanish laryngeal systems as identical, the variable nature of PSV in Spanish (thus lack of sufficient and salient input for this process) as well as PSV's typological rarity may be offered as potential explanations. The effect of RVA on both Spanish and English can be construed as an inability to block a dynamic and typologically common L1 post-lexical process as RVA in their L2/L3 productions, while perceptual results suggest that the same proficient multilingual learners are capable of detecting the non-target-like realizations of RVA in English. Altogether these point to a lack of direct correlation between multilingual perception and multilingual production as far as such dynamic phonological processes as RVA and PSV are concerned, intriguingly interacting with a multitude of such other factors as cognate and frequency effects.

Adding to the exploration of phonological interfaces, Schuhmann and Smith focus on the role of metrical feet in the acquisition of German plurals by L2 learners. Their study shows that L1 English learners gradually adopt the trochaic stress pattern typical of German plurals as their proficiency increases. This highlights how suprasegmental structures like stress patterns interact with morphological processes in L2 acquisition. As learners become more proficient, they internalize not only the morphophonological rules but also the prosodic patterns that define native-like production in German.

These studies illustrate the complexity of phonological learning with a focus on how phonological processes—particularly those that occur beyond the level of the segment—interact with other linguistic domains, such as morphology and suprasegmental features, in multilingual acquisition.

Variable surface realizations of sound sequences are also the focus of Zhang and Tessier, who investigate the anticipatory nasalization of low vowels preceding underlying nasal codas (loV-N), where N may be fully realized, lenited, or completely deleted on the surface in Beijing Mandarin. Despite the variable absence of coda Ns, the nasalized loVs carry the place of articulation feature of the following Ns such that they must agree for [+/– back] with following coronal or dorsal N codas while no labial N codas are allowed. None of these restrictions, however, holds for English. Similar to Yazawa et al. and Zhang and Tessier apply a computational simulation and use a GLA learner implemented in HG with weighted constraints, assuming that the "initial state of L2 grammar = the end state of the L1 grammar" in an attempt to explore how the fully copied L1 Mandarin grammar treats the range of loV-N sequences in L2 English. Evidence from L1 Mandarin speakers' perception is used to postulate various assumptions about the initial state of the grammar, which deviate from previous treatments of loV-N sequences. This grammar is then implemented in L1 and L2 simulations (the acquisition of English loV-N sequences). Independent evidence from L2 English production data and loanword phonology is then employed to test the validity of these simulations. These bring about an instructive methodology, where the cross-fertilization between theories of phonological grammars with inherent variability and learning simulations can inform L2 processes and be informed by them.

John and Rigoulot raise the question of how representational accounts can handle performance variation when looking at French speakers' acquisition of English /h/ focussing on the deletion of /h/ in production. They propose that the representations might be *fuzzy* or *murky*, and perhaps include diacritic markings which raises the question of whether the developmental grammars are constrained by UG.

Adding to the discussion of variability in surface realization, Cabrelli, Cruz, Escalante Martínez, Finestrat, and Luque examine the production of coda stops—phonotactically illicit in the L1 (Brazilian Portuguese) but permitted in the L2 (English)—by bilinguals immersed in an L2 environment. As with Bárkányi and Kiss, their data reveal that target-like perception does not guarantee target-like production; instead, production patterns often diverge through a variety of repair strategies. These asymmetries between modalities are formalized within the Bidirectional Phonetics and Phonology (BiPhon) framework (Boersma, 2011), which models perception and production within a single constraint-based grammar. As in their earlier work on perception with the same participant sample, the authors find that L2 production accuracy predicts L1 production patterns, suggesting L2 influence on L1 perception and production alike. While BiPhon captures the observed modality-specific asymmetries, the mechanisms that render L1 grammars permeable to influence from the L2 remain an open question for future research.

Methodological implications

Scott raises some methodological concerns in experimentation which may have led at times to contradictory behavioral results and proposes that experiments should control for orthographic and phonemic confounds. In particular, he reports on a phoneme detection task in which the object of the listener's attention is a sound adjacent to the phoneme of interest. Results of this task can be diagnostic of representational status.

Yazawa et al. suggest that perceptual behavior, as far as crosslinguistic categorical assimilation is concerned, may vary depending on the experimental setup. Depending on the task, perceived goodness of a vowel category in one language as another one in another language may be "fair" despite the considerable acoustic distance between the two since perceived cues may be defined relatively within each language rather than between two languages. As such and also compatible with the full copying hypothesis, they propose language-specific feature identification rather than a direct comparison of raw acoustic values between the two languages.

We hope that the breadth of this Research Topic in terms of theoretical approaches, empirical issues, as well as questions raised and answered will be appealing to a wide readership. In our view, the papers in this Research Topic unequivocally demonstrate that formal approaches to language acquisition embrace the integration of representation, interlanguage processes, input factors, learner variation, and psycholinguistic methodology. A field as complex, diverse (and fascinating) as multilingual phonology demands nothing less.

Author contributions

JA: Writing – original draft, Writing – review & editing. JC: Writing – original draft, Writing – review & editing. BK: Writing – original draft, Writing – review & editing.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Boersma, P. (2011). "A programme for bidirectional phonology and phonetics and their acquisition and evolution," in *Bidirectional Optimality Theory, Vol. 180*, eds A. Benz, and J. Mattausch (Linguistik Aktuell/Linguistics Today; John Benjamins Publishing Company), 33–72. doi: 10.1075/la.180.02boe

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 06 October 2023
DOI 10.3389/flang.2023.1242905

OPEN ACCESS

EDITED BY
Alicia Luque,
Nebrija University, Spain

REVIEWED BY
Joseph Salmons,
University of Wisconsin-Madison, United States
Ulrike Domahs,
University of Marburg, Germany

*CORRESPONDENCE
John Archibald
✉ johnarch@uvic.ca

RECEIVED 19 June 2023
ACCEPTED 12 September 2023
PUBLISHED 06 October 2023

CITATION
Archibald J (2023) Differential substitution: a contrastive hierarchy account.
Front. Lang. Sci. 2:1242905.
doi: 10.3389/flang.2023.1242905

Differential substitution: a contrastive hierarchy account

John Archibald*

Department of Linguistics, University of Victoria, Victoria, BC, Canada

In this article, I tackle the question of differential substitution in L2 phonology. A classic example of the phenomenon is learners from different L1s attempting to acquire the L2 English interdental fricative /θ/. Speakers of some languages (e.g., Japanese) tend to pronounce the /θ/ as [s] while speakers of other languages (e.g., Russian) tend to pronounce the /θ/ as [t]). Since both Japanese and Russian have both /s/ and /t/ in their phonemic inventories, it is interesting to ask why one language would choose [s] and the other [t]. What I argue in this article is that it is not a local comparison of two sounds, two features, or two phonemes that will determine why one segment rather than another is substituted. Rather, I argue that we must consider the formal representation of the entire segmental inventory (represented as a contrastive hierarchy) in order to understand why the Japanese pick the [s] but the Russian the [t] as the "best" substitute for the English /θ/. What I will demonstrate is that in the languages that substitute [s], [continuant] is the highest-ranked feature that has scope over the place and voice features in the contrastive hierarchy of phonological features. In the languages that substitute [t], the place and voice features rank above [continuant].

KEYWORDS

contrastive hierarchy, L2 phonology, differential substitution, phonological parsing, restructuring

1. Introduction

In this article, I tackle the question of *differential substitution* in L2 phonology. The term was coined by Weinberger (1997), though reference to the phenomenon goes back to at least (Weinreich, 1953). A classic example of the phenomenon is learners from different L1s attempting to acquire the L2 English interdental fricative /θ/. Speakers of some languages (e.g., Japanese) tend to pronounce the /θ/ as [s] (Lombardi, 2003) while speakers of other languages (e.g., Russian) tend to pronounce the /θ/ as [t] (Weinreich, 1953; Lombardi, 2003). Since both Japanese and Russian have both /s/ and /t/ in their phonemic inventories, it is interesting to ask why one language would choose [s] and the other [t]. What I am going to argue in this article is that it is not a local comparison of two sounds, two features, or two phonemes that will determine why one segment rather than another is substituted. Rather, I will argue that we must consider the formal representation of the entire segmental inventory (either consonantal or vocalic) in order to understand why the Japanese pick the [s] but the Russian the [t] as the "best" substitute for the English /θ/. Kabak (2019) reminds us that an even broader explanatory model would need to take both external factors, such as historical patterns and language contact, and internal linguistic factors into account in looking at the difficulties that are seen in the acquisition of interdental fricatives, but in this article, I restrict myself to a strictly phonological account.

What I will argue here is that in the languages that substitute [s], [continuant] is the highest-ranked feature that has scope over the place and voice features in the contrastive hierarchy of phonological features. In the languages that substitute [t], the place and voice features rank above [continuant].

1.1. Earlier approaches

Traditional approaches (Weinreich, 1957; Stockwell and Bowen, 1965; Nemser, 1971) would say something along the lines of "learners substitute the *closest* sound in their inventory." One can see the broad strokes of the argument when one considers how an L1 English speaker might substitute a [k] for a /q/. In the absence of a voiceless uvular stop [q] it seems unsurprising that a voiceless velar stop would be chosen as the "closest" match. However, previous explanations (Flege and Bohn, 2021) have struggled with independent measures of what might make one sound "closer" to another. Phonetic models such as the SLM (Flege, 1995) or PAM-L2 (Best and Tyler, 2007) have described how sounds that are *similar* in the L1 and the L2 tend to be assimilated (which makes it more difficult to acquire a new L2 phonetic category), while sounds that are *different* have been found to form new L2 phonetic categories more easily. But why is Russian [t] more like [θ] than Japanese [t] is? Conversely, why is Russian [s] less like [θ] than Japanese [s]? One can imagine detailed phonetic comparisons being the empirical data that would be drawn on to try to answer such questions. However, there is good reason to believe that substitution mechanisms (whatever they are) are not just motoric, articulatory problems because if they were we would not expect to see analogous problems of misperception [and we do; see John and Frasnelli (2023) for a discussion]. Oft times, learners who cannot produce a particular distinction also fail to perceive the contrast. Looking at phonological representation provides an explanation for both production and perception phenomena.

Brown (2000) took a more phonological view of the problem. She begins with the notion of categorical perception in phonology. Listeners hear an acoustically varied input stream and yet can assign the diverse phones from the environment to abstract phonemic categories. Let us take an example of a listener hearing a particular vowel, say [ɪ], spoken by an older male, a young adult female, and a child. The acoustics of those three vowels will be quite distinct, and yet the listener assigns them all to the category /ɪ/. Brown gives the example of English ears assigning non-English phones to English phonemes. An English speaker will have the voiceless stop phonemes /p/ and /t/ and /k/. When hearing a [q] in the input, the listener (or perhaps more accurately the parser) must assign the [q] to a phoneme. On an articulatory basis, one can argue that a uvular stop is "closer" to a velar stop than to an alveolar stop.

Rochet (1995) demonstrated that L1 English speakers substitute [u] for the L2 French /y/ while L1 Brazilian Portuguese speakers substitute [i], even though both languages have /i/ and /u/. He argues that acoustically there is more overlap with the English [u] and the French [y] (perhaps because of the fronted English [u]) than with the English [i] and the French [y] while, conversely, there is more overlap with the Brazilian Portuguese [i] and the French [y] than with the Brazilian Portuguese [u] and the French [y]. For Rochet, this means that the English speakers *hear* the French /y/ as [u] while the Brazilian Portuguese speakers *hear* the French /y/ as [i] because of the acoustic properties of the L1 and the L2 vowels. These perception issues are argued to underlie the analogous production patterns. Similarly, Brannen (2002) proposed that the explanation for differential substitution lies in phonetic rather than phonological patterns, but her model ran into problems related to this particular question in that it predicted that, for both European and Quebec French (QF), [f] should be the substitute of choice for the English /θ/ phoneme. Teasdale (1997) sought an articulatory phonetic approach to this problem by arguing that the relevant linguistic difference is in the articulation of the /s/. She noted that European French has a dental [s] and Japanese has a "flat" (or slit-type) [s] and that both of these languages substitute [s] for /θ/ while QF has an alveolar [s] so it substitutes [t]. While the correlation here is intriguing, the explanatory mechanism remains elusive.

TABLE 1 Phonological features of interdental fricative and close comparators.

	[continuant]	[coronal]	[strident]	[voice]
[θ]	+	+	–	–
[f]	+	–	–	–
[t]	–	+	–	–
[s]	+	+	+	–

Interestingly, James (1988) presents some data (which I will not explore in this article) that show that Dutch speakers substitute stops word-initially but fricatives word-finally (see also Collins and Mees, 1981). All I will say is that (a) contextual variation is different from differential substitution, and (b) it is not uncommon in the world's languages to have stops in onsets but fricatives in codas, but we would need to delve into Dutch syllable structure more closely before presenting a detailed analysis.

Brown's view was that the L1 phonemes act as a kind of categorical shoehorn to squeeze L2 sounds that do not quite fit into L1 categories. Brown argued that this sort of phonological funneling was impossible to overcome and that if the L1 lacked the relevant phonological feature to represent a particular contrast, that feature would never be triggered by the L2 input. González Poot (2011, 2014) demonstrated that such a position was too strong when he showed that L1 Spanish learners of L2 Yucatec Mayan had acquired a [constricted glottis] feature that is not found in L1 Spanish. The [constricted glottis] is a property of segments known as *ejectives* (e.g., [p'], [t'], [k]). González Poot's participants were indistinguishable from native speakers of Yucatec Mayan in their performance on a discrimination task with sound pairs like [p'a]/[pa] even though Spanish lacks ejectives and lacks the [constricted glottis] feature anywhere in its phonological inventory.

We can see that phonological features might well be involved in such a metric of *closeness* or *similarity*. Consider the substitution patterns commonly witnessed for *both* English interdentals /θ/ and /ð/. Crosslinguistically, we note that a /θ/ might be replaced by a [f] or a [t] or a [s]. What might lead to such variation? If we consider phonological features as the building blocks of phonemes, then Table 1 shows the relevant similarities.

Each of the "closest" sounds matches /θ/ on three features and differs on the other; all are 75% accurate. However, such a comparison does not tell us why Russian gives precedence to the mismatch on [continuant] (to change /θ/ → [t]) while Japanese gives precedence to the mismatch on [strident] (to change /θ/ → [s]). Is there something in Russian that prioritizes a match in place of articulation while something in Japanese values a match in

TABLE 2 Hancin-Bhatt's (1994) feature matrix.

	p	ph	b	bh	f	t	th	d	dh	s	z	ʈ	ɖ
[anterior]						+	+	+	+	+	+		
[coronal]						+	+	+	+	+	+	+	+
[back]													
[continuant]					+					+	+		
[strident]										+	+		
[voice]			+	+				+	+		+		+
[spread glottis]		+		+			+		+				

TABLE 3 Feature prominence rankings.

Feature	Prominence
[voice]; [spread glottis]	0.44
[coronal]	0.40
[strident]	0.32
[anterior]	0.24
[continuant]; [back]	0.20

manner of articulation? Under this approach, the answer seems to be "no."

Hancin-Bhatt (1994) tried to find a metric to establish a more nuanced comparison of features by looking at how many inventory contrasts each feature was involved in, and, in this way, arriving at a measure of functional load of each feature, which she called feature *prominence*. She adopted a model of radical underspecification (Archangeli, 1988), and then looked to see the weighting that each feature had. The first step was to generate a feature matrix showing by way of example (a selection) of Hindi phonemes in Table 2.

She proposed a prominence score that was calculated per feature by counting the number of phonemes for which that feature was specified and dividing by the number of phonemes in the inventory. For Table 2, we could calculate the prominence values of some of the features shown in (1), where [anterior] is more prominent than [spread glottis], which in turn is more prominent than [strident].

1. Anterior = 6/13 = 0.46
   Strident = 2/13 = 0.15
   Spread Glottis = 4/13 = 0.31

Hancin-Bhatt reported the feature prominence ranking for Hindi features based on the full inventory shown in Table 3.

Her prediction was that features that were more prominent would be more easily perceived in the L2 input. In the above case [continuant] is of low prominence, so the listeners are predicted not to be sensitive to [continuant] cues in the input and, thus, would be predicted to substitute the stop (i.e., non-continuant) [t] for English /θ/ in a perception task. Bansal (1969) claims that this is the preferred Hindi substitution pattern. However, ultimately, Hancin-Bhatt's predictions were not always borne out. For example, she predicted that L1 Turkish speakers should substitute [s] for /θ/, but in her experiment they misperceived /θ/ as [t].

Hanuliková and Weber (2012) show that differential substitution affects perception in that when German (an [s]-language) learners of English are compared to Dutch (a [t]-language) learners of English and participants hear English words spoken with either an [s] or a [t], respectively, substituted in an English word, they fixate on the correct English [θ] word longer when the accented version matches their own substitution preference. For example, L1 German participants fixate on *thick* when they hear *sick* longer than when they hear *tick*, as opposed to L1 Dutch participants who would fixate more on *thick* when they hear *tick* than when they hear *sick*.

1.2. Possible explanations

Weinberger (1997) adopted underspecification theory to account for why L1 Japanese subjects pick [s] while L1 Russian subjects pick [t] as the substitute for English /θ/. Basically, Weinberger argues that Russian and Japanese treat the feature [continuant] differently. Japanese has /s/ unspecified for [continuant] while Russian has /t/ unspecified for [continuant]. Weinberger does not provide detailed evidence for these feature values in the respective L1s. Picard (2002, p. 88) argued that this method could not account for the difference between European and Quebec French since "both dialects have *exactly* the same underlying consonant inventory," so he proposes a perception-based account. Picard acknowledges a key difference between the two French dialects as being that Canadian French has an assibilation rule that turns /t/ into [t^s] before high, front vowels. He also acknowledges that with [strident] playing a role in Quebec French allophony, it seems counterintuitive (given the L1 import of [strident]) that the QF speakers do not pick the strident [s] as the substitute. When we look at contrastive hierarchy (CH) and the activity principle, we will see how the relationship works.

Smith (1997) looks at L1 Japanese (s/z), German (s/z), Turkish (t/d), and European (s/z) and Quebec (t/d) French. She focuses on the lack of the feature [distributed] in these languages. Thus, she seems to argue that they all have the same wrong phonological representation for /θ/ but the differential substitutes are the result of phonetics. Her goal is to account for the phonological *cause* of the misperceptions rather than the nature of which substitute is chosen.

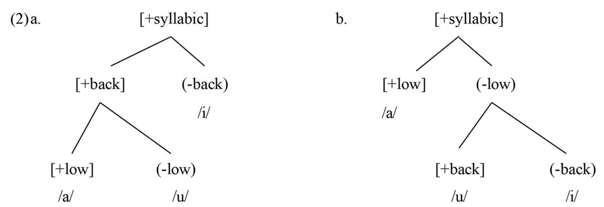
Lombardi (2003) proposes a quite different class of solution in that she argues that the [t]-substitution occurs as the result of a high-ranked markedness constraint, thus resulting in the production of a universally less-marked stop segment, while the [s]-substitution pattern results from a high-ranked faithfulness constraint, thus resulting in the production of a form that is faithful to an L1 underlying form. As we mentioned earlier, such an output-based account has difficulty explaining the parallel perception difficulties the learners demonstrate (given the assumption of richness of the base).

Kwon (2021) presents a line of work that tackles the notion of formalizing perceptual similarity and, hence, potentially being able to predict which L1 sound is closer to the L2 interdental. She adopts an integration of the Lahiri and Reetz (2002) featurally underspecified lexicon model and Dresher's (2009) Contrastive Hierarchy (CH). Her approach generates perceptual similarity scores that argue that phonological representation, not phonetic representation, has a stronger influence on L2 perception.

Note that, with the exception of Kwon (2021), none of the previous approaches recognized constituent structure in phonological features. Some invoked feature geometry, but as Cowper and Hall (2019) remind us, "contrastive hierarchies are paradigmatic (defining systems of oppositions), and feature geometries are syntagmatic, structuring combinations of features in phonological representations."

2. The contrastive hierarchy

I propose that using a CH approach (Dresher, 2009, 2018; Hall, 2011; Archibald, 2022a,b; Chandlee, 2023) provides just the mechanism we need to account for such differential substitution patterns. I present cross-linguistic patterns found in the existing literature and will propose a phonological explanation. I am going to focus on the interdentals [θ/ð] and their substitutes in Russian, Japanese, European French, and QF. Russian and QF tend to substitute [t] for /θ/ and [d] for /ð/ while European French and Japanese tend to substitute [s] for /θ/ and [z] for /ð/ (Gatbonton, 1978; Smith, 1997; Teasdale, 1997; Picard, 2002). However, this same machinery could be used to account for other examples of differential substitution. The CH model is a theory of ranked contrastive underspecification. Let us consider two partial phonemic inventories to see how the machinery works. Dresher (2018) gives the example of the two three-vowel systems that each use the same features but differ in feature ranking, shown in (2).



The difference between (2a) and (2b) is the ranking of features. In (2a) the inventory is first divided by the feature [back] while in (2b) the inventory is first divided by the feature [low]. In (2a) /i/ is uniquely and unambiguously represented by (non-back) so it requires no further features. In (2b), however, /i/ is represented by the feature (non-low) *and* the feature (non-back). Therefore, even though the two systems have the same phonemes, they may well behave differently. In the words of Dresher (2018, p. 21), "We predict that these differences in organization will be reflected in patterns of merger and neutralization." Imagine that there was a vowel harmony process in the language; in (2a) both /a/ and /u/ would be involved in [back] harmony while in (2b) only /u/ would be involved in [back] harmony. Key elements of the model are given in Table 4 (adapted from Dresher, 2018).

Each language, therefore, will have a different ranking of features; there is no universal invariant hierarchy that governs all human languages. The features chosen by each language will also vary. Only the features represented in the CH are available for allophonic computation. For example, if a language does not represent a [round] feature in the vowel in the CH then we could not see [round] spreading occurring in adjacent segments. If features do play a role in the computational system, they are said to be *active*. Note that (a) this activity can be a cue to the learner as to which features to include in the CH [following what Dresher (2009) calls the successive division algorithm], and (b) not all contrastive features are active; some are there to represent phonemic contrast alone. For Dresher (2018), features are emergent from contrast and not drawn from a universal inventory. Nothing in this article hinges on this stance, so I will remain agnostic at this time. Finally, for the features that are represented, I am assuming that features are binary with a notational convention that the marked value is shown in square brackets (e.g., [+voice]) while the unmarked value is shown in parentheses [e.g., (–voice)][1]. Again, this is not critical for the data and arguments in this article, but as Natvig and Salmons (2021) have shown, there is a principled difference in phonetic behavior that falls out from the markedness status (with more variation being found in the realization of the

TABLE 4 Key properties of contrastive hierarchy theory.

Variability of feature ordering	Contrastive feature hierarchies are language particular
The Contrastivist Hypothesis (Hall, 2017)	The phonological component of a language L operates only on those features that are necessary to distinguish the phonemes of L from one another
Feature activity	A feature can be said to be active if it plays a role in the phonological computation; that is, if it is required for the expression of phonological regularities in a language, including both static phonotactic patterns and patterns of alternation
Phonological primes	Features are binary; every feature has a marked value, designated [+F], and an unmarked value, designated (–F)

1 A reviewer notes that this is reminiscent of the inheritance strategy adopted in Purnell et al. (2019).

unmarked values). Archibald (in press) has shown how this input variation can be a relevant cue for the learner when setting up a new CH. Before moving on to discuss the relevance of this model to multilingualism, let me just expand slightly on the nature of the successive division algorithm (SDA). The SDA sets out the procedure by which a child or a phonologist discovers the contrastive inventory of the ambient language. At the initial stage, there are no phonemic contrasts. Then, the inventory is divided in two by introducing one feature. For example, when discovering contrastive vowels, a child may first divide the vowel inventory into [+high] and (−high) vowels. At this point they may be making a distinction between /i/ and /a/, but not between /i/ and /u/ (as they are both [+high]). Then, the SDA dictates that another feature is introduced to create a further binary distinction. For example, the [+high] vowels might be further subdivided into [+front] and (−front) vowels. At this stage, the child could represent a three-way phonemic contrast between /i/, /u/, and /a/. The division will continue successively until all the phonemes of the target language are uniquely specified.

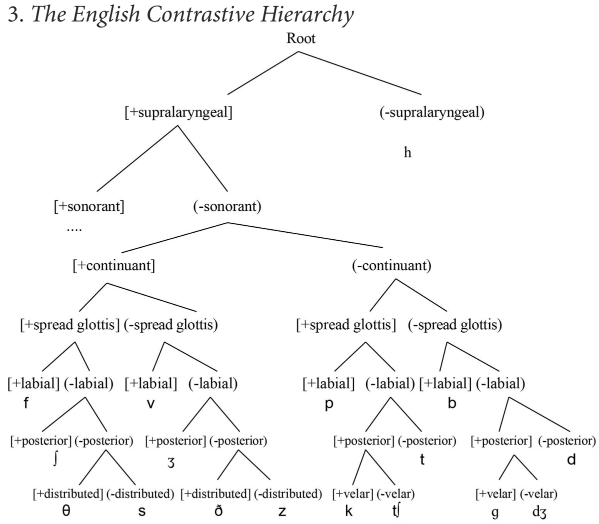
The resulting contrastive hierarchies can also be thought of as parsing the input to which the *Lx* learner is exposed (where *x* stands for a natural number such as 1, 2, or 3). By this, I mean to indicate that this architecture can account for L1, L2, or L3 (etc.) acquisition. If we return to the previous examples discussed when introducing Brown, we can see that the L*x* listener might be faced with [i/ɪ] in the input, or [k/q]. An English speaker would have a featural contrast that would allow the [i] to be assigned a different structure than the [ɪ], but would lack the structure necessary to disambiguate [k] from [q]. Consistent with the behavioral facts of Brown, the [q] would be assigned to the /k/ category, and the [ɪ] would be assigned to the /ɪ/ category. Certainly, at the initial stage of L2 learning, this explains why we speak and hear with an accent (see Archibald, 2021). The parser can also help us to understand differential substitution. Under a CH model, differential substitution is the by-product of phonological parsing of the L*x* input by means of the L1 feature hierarchy. The L*x* segments that are undifferentiated by an L1 parse explain both production and perception substitutions.

Let us now turn to the specific consonantal hierarchies (of English, Russian, Japanese, and European French and QF) that form the basis of my arguments. Since I am focusing on the differential substitution of the English interdentals, for reasons of space, I will limit my presentation of the features to the obstruents.

2.1. The English contrastive hierarchy

Working in the learnability tradition (Wexler and Culicover, 1983), let us begin by looking at the phonological representation of what is to be acquired in the L*x* in terms of obstruent phonemes. That is to say, what are the L1 French, Japanese, and Russian speakers trying to acquire when they acquire an English /θ/ or an English /ð/ð? The English tree shown in (3) is based on the feature ranking of [supralaryngeal] > [sonorant] > [continuant] > [spread glottis] > [labial] > [posterior] > [velar]/[distributed].

3. *The English Contrastive Hierarchy*



I will not go into much detail as to the activity of the English features (and the role this could play in determining the acquisition path) as the focus of this article is on the initial stages of the acquisition of English interdental fricatives, and not on the restructuring stages of subsequent acquisition stages (see Archibald, 2023).

The non-supralaryngeal analysis of the English /h/ is partially motivated by such things as its exceptional behavior in a variety of phonological domains. It has special phonotactics in that it does not occur in codas or before another consonant. It can be deleted in weak metrical positions in words (*Birming(h)am* with an unstressed final syllable [əm]), or phrases (*give it to (h)im*). However, Goad and Mah (2007) argue that English /h/ is not placeless (see Rose, 1996) in that in many languages with placeless /h/, the /h/ occurs in codas but English does not allow /h/ in codas. Goad and Mah suggest that the English /h/ has a [spread glottis] feature (under laryngeal place) and this is what leads to the fact that the distribution of /h/ mirrors the distribution of aspiration (Jensen, 1993) as seen in words like (*vehicular/ve(h)icle* and *atom/at^{h}omic*).

The feature [continuant] is involved in the phonological computation system as we see allophony in forms such as *delete/deletion* where /t/ → [ʃ], and *invade/invasion* where /d/ → [ʒ][2]. The feature [spread glottis] is active as we see voicing assimilation of the plural morpheme (in forms such as *dog+s* [z] and *cat+s* [s]). The feature [labial] is active as we see place assimilation of the negative prefix in forms such as *im+possible*. The feature [posterior] is active as we see palatalization in forms such as *education* where /dj/ → [dʒ]. The feature [velar] is active as we see it in cross-word assimilation in such phrases as *bad guy* → *bag guy* where /d/ → [g].

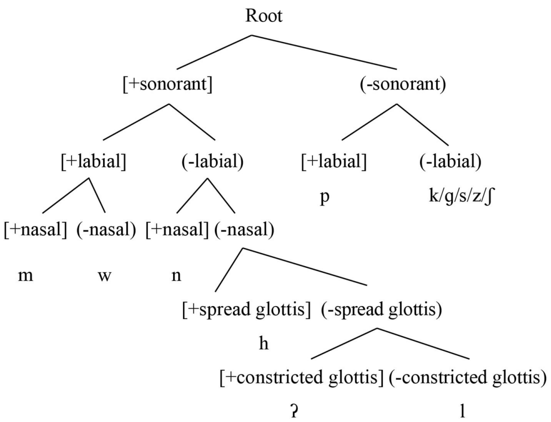
In terms of a transition theory (Cummins, 1983), L*x* learners will need to adjust their L1 CH to achieve this L2 CH. The listeners would be exposed to the English phones, and would assign structure to them by parsing via the L1 CH. Ultimately, I will demonstrate that it is the variation in this parsing process that explains the phenomenon of differential substitution.

2 A reviewer notes that these are largely historical patterns, and may not be representative of the same type of phonological activity discussed elsewhere in the article.

2.2. Loanword phonology

Before explicating the CH treatment of differential substitution, let me give some background in loanword phonology that will be useful. Herd (2005) shows convincingly how the CH model can account for loanword phonology. The hierarchy shown in (4) [adapted from Herd (2005)] summarizes how Hawaiian replaces English coronal obstruents and [g] with /k/ (e.g., "lettuce" → /lekuke/; "brush" → /palaki/). As one might imagine, it is difficult to come up with a rule that would turn [s] → /k/ *and* turn [z] → /k/ *and* turn [ʃ] → /k/.

4.



The key to understanding this approach is to think of the L2 input being parsed through the filter of the L1 CH. However, note that, unlike the approach of Brown (2000), this is not a segment-by-segment treatment but rather a reflection that the parsing takes place at an inventory level. Archibald (2022a,b) showed that inventory effects are clearly observed in L2 and L3 phonology. In the above example, we see that L2 [s], [z], [ʃ], and [g] are all parsed in the L1 hierarchy as [-sonorant, -labial]. Described in another way, the sounds [s], [z], [ʃ], and [g] cannot be unambiguously represented in the Hawaiian CH.

With this background, let us now turn to how the CH can provide insight into the phenomenon of differential substitution.

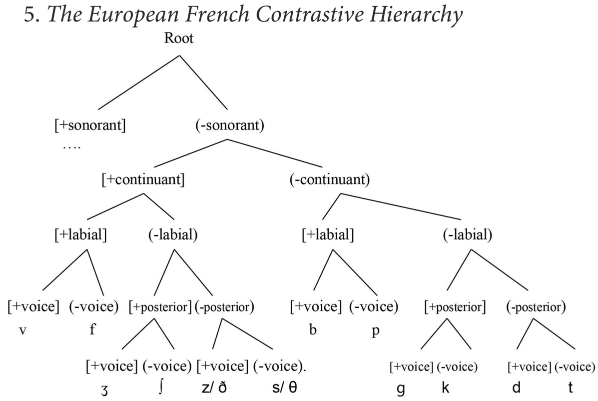
2.3. The European French contrastive hierarchy

First, we consider the parsing of the L*x* English interdentals by speakers of European French (EF). EF speakers tend to substitute [s] for the English /θ/. The obstruent inventory of French is uncontroversial (see Walker, 1984). There is no /h/ in EF so the [supralaryngeal] contrast is not necessary. Daniel Hall (private communication) raises the issue of h-aspiré (which suggests that French *does* have some way of marking /h/) and that perhaps this could influence how learners represent the English /h/. However, I would suggest that the h-aspiré would have to be represented differently than the English /h/ because otherwise, the sound would not be problematic for French learners (and it is). Walker (1984, p. 41) says that h-apiré is a "morphological and lexical" issue, not a phonological one.

There is a uvular /ʁ/ in the sonorants, but that is not relevant to the question at hand. In documenting the consonantal allophonic processes, there are no processes that make use of the [strident] feature so we can assume that [strident] is inactive and non-contrastive in EF.

There is evidence for the [posterior] feature being active in EF. Niebuhr et al. (2011, p. 430) state that, "contrary to the predominant view in the literature—assimilation of place of articulation does exist in French, at least in sequences of alveolar and postalveolar sibilants." They give the example (Kohler, 2002; Bertrand et al., 2007) of /s/ → [ʃ] in *Scotch sur la bouche* ("Scotch tape across the mouth"). There is also evidence that [voice] is active given the voicing assimilation (i.e., devoicing) noted in Walker (1984); for example, we see /pje/ → [pje̥]. The proposed CH for EF obstruents is given in (5).

5. *The European French Contrastive Hierarchy*



There is evidence that [continuant] is high-ranked in EF that emerges when we compare it with QF concerning the process of spirantization. The QF stop /d/ and /t/ lenite intervocalically, but they do not in EF (Colantoni et al., 2021). This differential behavior results from a difference in the ranking, as the feature does not seem to be active in either EF or QF. Having a high-ranking specified target for the stops under (-continuant) means that the articulatory targets are quite precise and would not brook much variation. Conversely, QF (which we will argue has low-ranked [continuant]) allows much more phonetic variation in the duration of the stops /d/ and /t/ because the relevant stops are unspecified for [continuant]. There is also evidence from loanword phonology in the two varieties that supports the activity of [continuant] in EF. Coté (2021) examines how English loanwords are repaired in both EF and QF. The examples that are relevant to our question include the affricates /dʒ/ and /tʃ/. The relevant borrowings and repairs are shown in Table 5.

Note that the English non-continuant affricates remain as affricates in QF but the value of [continuant] changes in EF thus indicating that [continuant] is active in the phonological computation.

TABLE 5 English loanword adaptations in European and Quebec French.

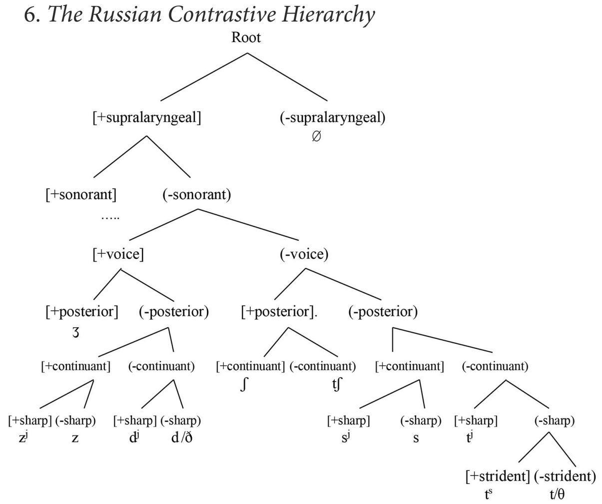
English source	EF repair	QF repair
jockey [dʒɑki]	[ʒɔkɛ]	[dʒɔke]
chips [tʃIps]	[ʃips]	[tʃips]
jamboree [dʒæmbəri]	[ʒãmbɔRe]	[dʒãmbɔʁi]

For ease of comparison in (4), I have included the English interdentals [θ/ð] to show how they would be parsed by the L1 EF CH. Note that the EF CH cannot disambiguate either [z/ ð] or [s/ θ]. I will not explore here how the EF hierarchy has to be changed (incrementally) to arrive at the English CH; differential substitution is our concern here. Note that the [s] substitution is driven by the highest-ranked [continuant] feature.

2.4. The Russian contrastive hierarchy

In tackling the question of differential substitution in Russian, I am limiting myself to presenting the CH for a subset of the Russian consonants. For reasons of space (Russian has a very complex consonantal inventory), I will only be presenting an analysis of the coronal consonants (which, after all, are the main candidates for /θ/ substitution). I base my analysis on Dresher and Hall (2020). They provide arguments for the ranking of [voice] above [continuant] in Russian; /ts/ and /tʃ/ trigger assimilatory devoicing even though they do not contrast with underlying voiced counterparts (/dz/, /dʒ/). This can be accounted for if [voice] is above [continuant] in the hierarchy. These patterns of regressive voicing assimilation evidence that [voice] is an active feature in Russian. Processes of palatalization also suggest that [sharp] is an active feature. The Russian CH is given in (6).

6. *The Russian Contrastive Hierarchy*



So, in Russian, [strident] is low-ranked, and is really only needed for the /tˢ/ vs. /t/ contrast. Crucially, [voice] is ranked above [continuant].

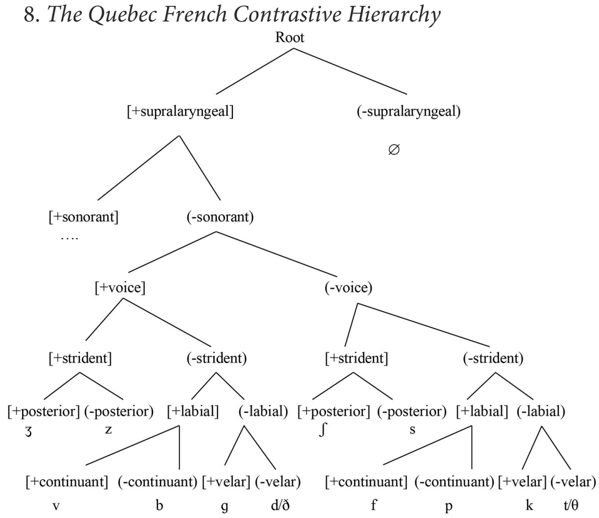
2.5. The Quebec French contrastive hierarchy

Walker (1984) notes that EF and QF are virtually identical in their consonantal systems (there are arguably differences in the vocalic system and certain properties of stress placement in the two varieties but those are not our concern here), and at the phonemic level this is certainly the case. There is, however, an allophonic process found in QF that is not found in EF that is directly relevant to the issue of differential substitution: assibilation. I will begin by talking about assibilation in QF and then proceed to a more general discussion. In QF there is an allophonic process that changes the phonemes /t/ and /d/ to [tˢ] and [dᶻ], respectively, before the high, front vowels /i/ and /y/ (but not before /u/). Examples are given in (7).

7.	'petit' /pəti/ → [pətsi] *small*	'battu' /baty/ → [batˢy] *beaten*	'tout' /tu/ → [tu] *all*
	'dites' [dit] → [dᶻit] *say-2sg*	'dupe' /dyp/ → [dᶻyp] *dupe*	'doute' /dut/ → [dut] *doubt*

Following LaCharité (1993), I assume that affricates are strident stops. My argument is that the change from a non-strident to a strident stop indicates that the [strident] feature is active in QF. I will not get into possible spreading analyses that can motivate the phonetic change. Telfer (2006) provides an interesting account of how the following high, front vowel environment can trigger the representation of a [strident] feature. Baker and Smith (2010), I would argue, provide some indirect evidence for the activity of the [strident] feature in QF. Their article concerns the acquisition of the /y/ phoneme in L2 French by speakers from an L1 that lacks the phoneme (English). They show a correlation between people who assibilate and people who have acquired /y/. This is evidence that assibilation is an active cue (and, hence, an active feature) in QF that helps the learners to acquire the new vowel /y/. The QF CH is given in (8).

8. *The Quebec French Contrastive Hierarchy*



Walker (1984, p. 100) also refers to the "mellowing" of [ʃ] and [ʒ] to [x] and [ɦ] in some QF dialects (Beauce; Ottawa/Hull). Thus, the target sounds for this phonological process are [+strident, +posterior], which is consistent with both [strident] and [posterior] being active in QF.

Colantoni et al. (2021) talk about coronal stop lenition using electropalatography data on two EF and two QF speakers. Interestingly, there was a significant difference between the two dialect groups in which the QF speakers showed significantly more lenition (i.e., less contact) for both /d/ and /t/, though the effect was greater for /d/ than for the EF speakers. Note that Colantoni et al.

TABLE 6 Quebec French loanword adaptation.

English loanword	French adaptation	Process
“thrill” [θɹɪl]	[tRI1]	θ → t
“tip” [tIp]	[tˢIp]	t → tˢ
“directory” [dəɹɛktɹi]	[dᶻiRɛktɔRe]	d → dᶻ

(2021) controlled for the presence of assibilation in their data in that the lenition effects were found regardless of the following vowel (i.e., not just in the case of a following high, front vowel)[3]. This L1 property provides an additional reason why the QF speakers would parse the English [θ] as /t/ and the English [ð] as /d/. I would suggest that since [continuant] is low-ranked in QF (and, in fact, unspecified for the non-labials), that is what causes the lenition of the QF /t,d/ in comparison to the EF phonemic inventory, which has [continuant] highly-ranked.

2.5.1. Quebec French loanword adaptation

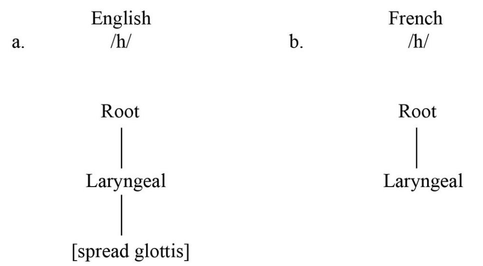
The discussion of loanword adaptations of English words that have been borrowed into QF is also relevant to our discussion. Hsu and Jesney (2017) provide a theoretical account of the patterns noted in Roy (1992). Many studies (Itô and Mester, 1995; Paradis and LaCharité, 1997) have revealed that loanwords often pattern differently phonologically than native vocabulary. One way to describe this is to say that there are different lexical *strata* and that the different strata obey different markedness restrictions. Among the patterns that Hsu and Jesney (2017) address are the following, shown in Table 6[4].

Some adaptation processes apply quite uniformly, while others are more variable. Of particular interest to our discussion here, though, is the fact that words with three segments were *always* repaired. Those three segments were [θ], [ð], and [h]. Phonologically, for Hsu and Jesney (2017), this suggests that [θ], [ð], and [h] are the most robust markedness constraints that must be respected at all lexical strata. I would argue that the CH gives us a potential explanation of what makes these three sounds problematic. The [θ] and [ð] English inputs cannot be disambiguated by the French feature hierarchy. The [h] sound is perhaps a different story, and let me probe that question in a little more detail.

2.6. The problem of [h]

There is a considerable body of evidence that francophone learners of English have difficulty acquiring the /h/ phoneme (Janda and Auger, 1992; Brannen, 2011; White et al., 2015; Mah et al., 2016). The English /h/ tends to be deleted/omitted in production (e.g., “harm” → [arm]). Many researchers (LaCharité and Prévost, 1999; Trofimovich and John, 2011) argue that this is related to inaccurate perception. Jackson and Cardoso (2023) argue that it is the variability in the English grapheme/phoneme correspondence that accounts for the difficulty. It can be pronounced in a stressed syllable (*history, inherent*) but silent in an unstressed syllable (*vehicle, I’d (h)ave*).

Mah et al. (2016) also address the question of francophone acquisition of English /h/ in an ERP study. They begin by noting that English [h] is acoustically weak; it is made with no significant oral or pharyngeal constriction, and it has no vocal fold vibration. They suggest that the articulation of /h/ is more like a voiceless vowel. This can be supported by looking at how differently the vocal tract is positioned in the production of the /h/ in /hɑt/ compared to the /h/ in /hit/ or /hu/; a narrow phonetic transcription could well be [ɑ̥ɑt], [i̥it], and [u̥u]. Mah et al. further suggest that English /h/ would have the laryngeal feature [spread glottis]. French /h/ they argue is realized by just a bare laryngeal node. They are working within a feature geometry model. The relevant structures are given in (9a) for English and (9b) for French.

9.


In their ERP study, they found no mismatched negativity (MMN) effect in a linguistic task but they did find an MMN effect in a non-linguistic task. This suggests that the difficulties L1 French participants are having are not at the acoustic level but rather, they argue, at the representational level. Shortly, I will provide a representational account adopting a CH model. Interestingly, production data show that francophone learners also occasionally insert an [h] in a non-target-like environment (e.g., “arm” → [h]arm). While the /h/ question may seem to be a bit removed from the differential substitution question, I hope to show that the CH approach can (a) handle the /h/ data too, (b) account for why some languages (e.g., Arabic) substitute a segment ([h]) for English /h/ rather than the French strategy of substituting Ø, and (c) show why the acquisition of an L2 phoneme that contrasts with Ø in the L1 is more difficult to acquire compared to an L2 phoneme that is realized with a different overt phone in the L1.

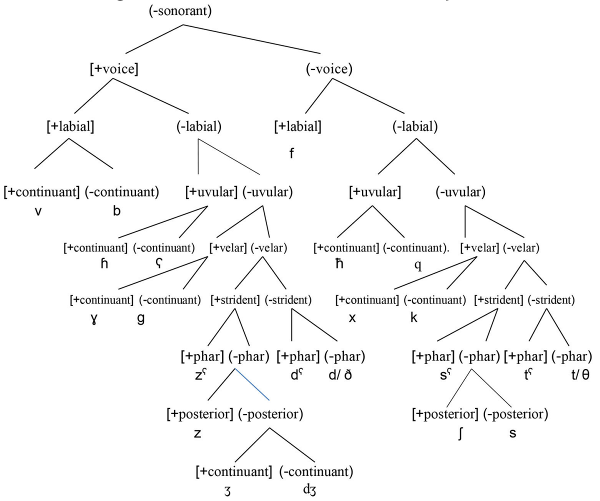
In contrast with the French problems with English /h/, let us look at the CH for Algerian Arabic. This discussion draws on Benrabah (1991) and Archibald (2022a,b), who look at the L3 English of L1 Arabic/L2 French individuals. We can see two things from this hierarchy: (1) following McCarthy (1994), Arabic gutturals (including /h/) are not placeless, and (2) the English [θ] would be parsed under the Arabic /t/. The Algerian Arabic CH is given in (10)[5].

3 They also discuss Spanish speakers in their article and the French lenition was less than the Spanish lenition, but nevertheless present.

4 I will not address the sonorants, only the obstruents in this article.

5 There might be more complexity required for the representations of /ɦ/ and /ħ/ than I am showing. The key point for the arguments in this article is that they are under a place feature node (unlike English).

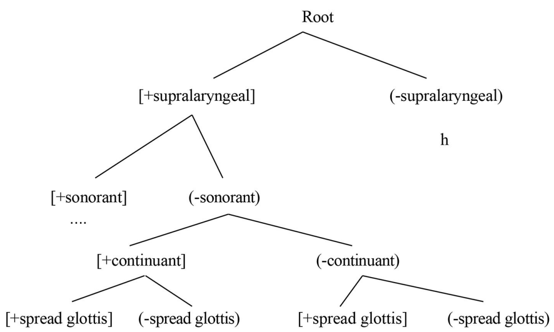
10. *The Algerian Arabic Contrastive Hierarchy*



It is worth noting that in this northern African cultural context, the variety of French would tend to substitute a [s] for the English /θ/ so this substitution pattern of [t] is consistent with the transfer of the Arabic CH to represent the English obstruents. In this CH as well, we note that [voice] > [place] > [continuant], which is consistent with the [t] substitution we have already seen in Russian and QF.

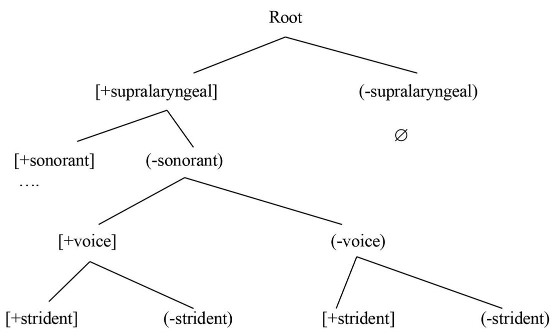
However, now let us compare the learning task of Arabic learners of English to the task of French learners of English when it comes to the learning of /h/. Let us reiterate what the structure of English /h/ looks like. I will repeat the top part of the CH of (3) in (11).

11.



Note that in English there is a top-ranked distinction between the supralaryngeal sounds and /h/. Note too that QF lacks such a [±supralaryngeal] contrast. I repeat the top part of the CH of (8) in (12).

12.



Acquiring such a high-ranked contrast, which is absent from the L1, appears to be difficult. This is consistent with what Archibald (in press) proposes in terms of a transition theory of acquiring an L2 CH where learning is conservative and incremental and begins by positing changes at the bottom of the hierarchy.

We can now contrast the English hierarchy with the high-ranked [supralaryngeal] feature with the French hierarchy, which lacks this high-ranked feature (leading to difficulty), with the Arabic hierarchy which can parse the English [h] under a place node.

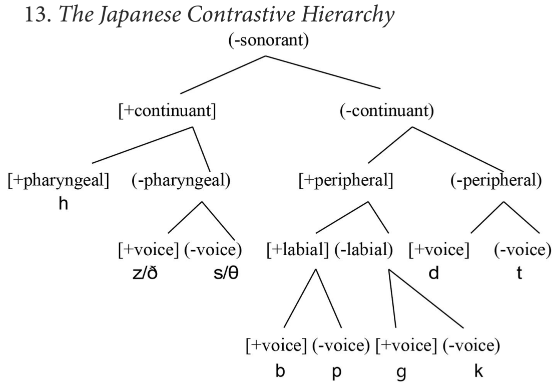
The question of how L2 or L3 grammars are restructured to become more targetlike is, of course, fascinating and complex. Space precludes me from exploring it in depth here (though see Archibald, 2023 for further discussion). I will note that the work of Oxford (2015) on historical change (i.e., grammar restructuring) reveals many potential similarities in how contrastive hierarchies change over time historically and in multilinguals.

Having added this relatively brief discussion of the /h/ phenomena to the discussion of interdental differential substitution, the following general point can be made. We do not need to invoke multiple explanations for /θ/ becoming [t] in some languages and [s] in others, nor do we need to invoke special machinery to account for the difficulty of L1 French producing (and perceiving) and English [h] compared with the L1 Arabic ability to produce (and perceive) the English [h]. It can all be explained, more parsimoniously, via the machinery of a CH.

2.7. The Japanese contrastive hierarchy

Let us conclude the discussion of cross-linguistic data by looking at the interdental patterns in L1 Japanese speakers. In (13) we can see the CH for Japanese.

13. *The Japanese Contrastive Hierarchy*



The process of spirantization in which /t/ → [s] and /d/ → [z] (Akamatsu, 1997, 2000; Labrune, 2012) indicates that [continuant] is an active feature in Japanese as shown in (14).

(14) /mikadɯki/ → [mikazɯki] "increasing moon"

Further evidence of its activity comes from Vance (1987, 2008), who notes that /b/ and /g/ spirantize to [β] and [ɣ], respectively, intervocalically. Labrune (2012, p. 64) comments that this spirantization occurs in "familiar register or fast tempo."

There is considerable evidence that from voicing assimilation (i.e., rendaku) that [voice] is an active feature (Rice, 2005; Kubozono, 2015). For example, the [s] in *sake* "salmon" becomes a [z] in the phrase *ʃio+zake* "salted salmon." Under this process, we see the following alternations:

/t/ → [d]
/k/ → [g]

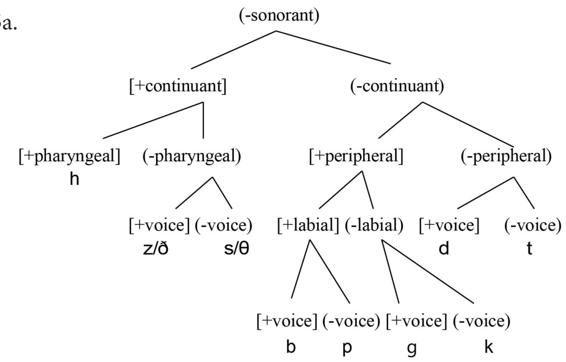
/h/ → [b]
/s/ → [z]

Consistent with what we saw with the EF participants, having [continuant] as the highest-ranked feature leads to the substitution of the fricatives for the English interdentals. Once again, we see [continuant] > [place] > [voice]; [continuant] has scope over [place].

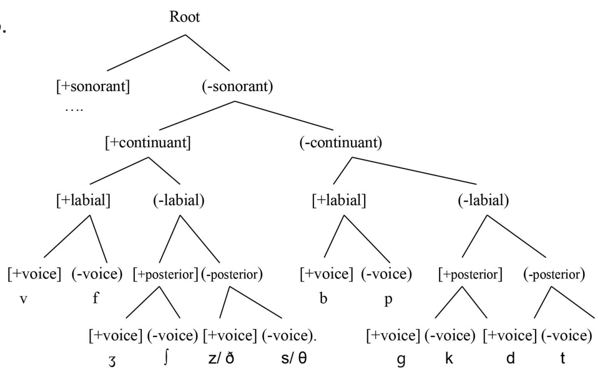
2.8. Summary

To summarize, let us recap the following three pairwise comparisons to illustrate the analyses. In (15a) and (15b), we see two [s]-substitution languages where [continuant] > [place].

15a.

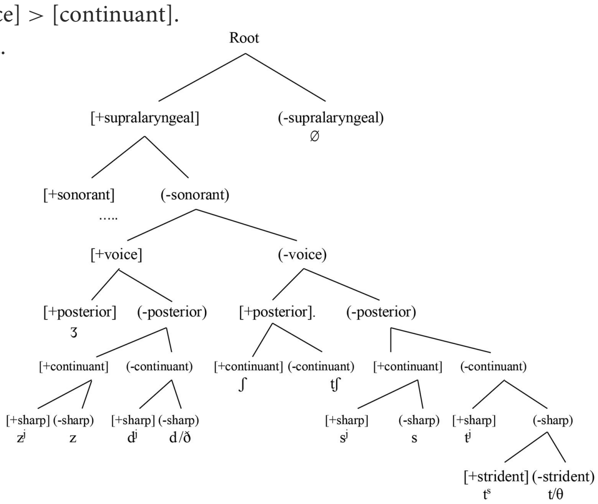


15b.

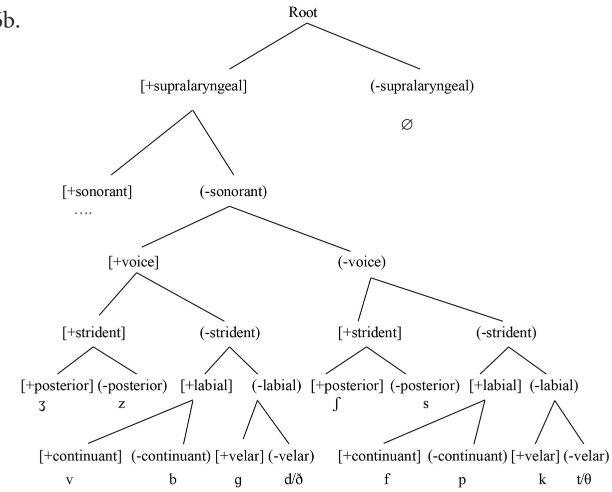


In (16a) and (16b), we see two [t]-substitution languages where [place] > [continuant].

16a.

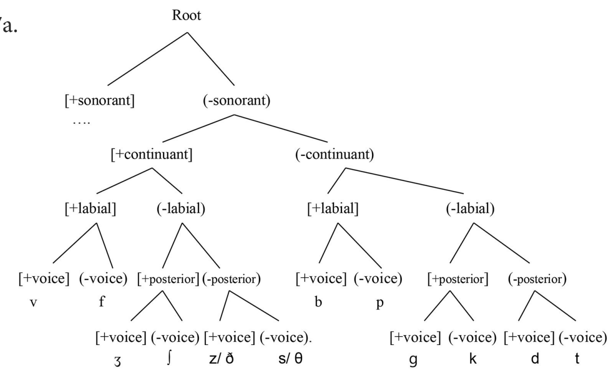


16b.

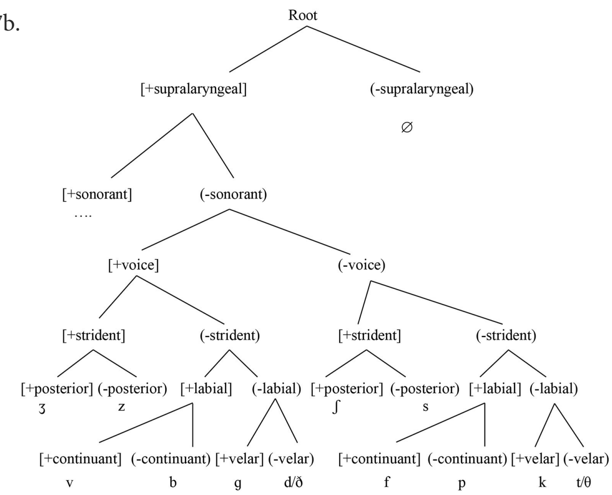


In Figures (17a) and (17b), we compare the two varieties of French where we see how one variety (EF) has [continuant] > [place] resulting in a [t] substitution while the other variety (QF) has the ranking [place] > [continuant] resulting in a [s] substitution.

17a.



17b.



3. Conclusion

In this article, I have argued that the phenomenon of differential substitution can be explained in a principled fashion under the architectural assumption of the CH model. We see an [s] substitution for /θ/ when the [continuant] feature is the highest-ranked feature in the hierarchy (above [place] and [voice]) (e.g., Japanese and EF). In languages where [place] and [voice] features are ranked above [continuant], we see [t] as it is the completely unmarked category into which the /θ/ can be parsed (e.g., Russian and QF).

We have also proposed that parsing failures at the lowest level lead to minor ambiguities that are resolved incrementally in the learning process, while parsing failures that require the addition of a high-ranked feature are more problematic for the learners (e.g., French /h/).

The CH approach differs from Lombardi's (2003) markedness vs. faithfulness analysis in that there is a unified analysis of the two substitution options: it is all transfer of parsing procedures. Input segments that are undifferentiated by an L1-feature parse are assigned to the same phonological category. Unlike an SLM approach to equivalence classification, this is not based on surface phonetic features but rather the phonological grammar. Furthermore, unlike Lombardi (2003), this accounts for why there are perception substitutions as well as production substitutions. We must acknowledge that this is the starting point of the learning path that will require the learner to restructure the L1 CH to move incrementally closer to the L2 CH using such operations as (1) merger of an L1 contrast that is not required in the L2, (2) redeployment of an L1 contrast from one part of the L1 CH to another part of the L2 CH, or (3) triggering a new feature not found in the L1. This approach also allows for individual variation to be accounted for in a principled way (Archibald, in press) given that different learners may restructure different portions of the hierarchy in different sequences.

The CH approach recognizes that the differential substitutions fall out from inventory effects, not local surface comparisons, and further shows that the machinery of CH that has been productively used to account for L1A (Bohn and Santos, 2018), historical change (Oxford, 2015), sociolinguistics (Natvig and Salmons, 2021; Hunt Gardner and Roeder, 2022), and L3A (Archibald, 2022a,b) can also be used productively for an explanatory account of one of the oldest questions in L2 phonology: differential substitution.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This study was funded by SSHRCC grant 435-2022-1097.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Akamatsu, T. (1997). *Japanese Phonetics: Theory and Practice*. Munich: Lincom Europa.

Akamatsu, T. (2000). *Japanese Phonology: A Functional Approach. Vol. 38. Lincom Studies in Asian Linguistics*. Munich: Lincom Europa.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology* 5, 183–207. doi: 10.1017/S0952675700002268

Archibald, J. (2021). Speaking and hearing with an accent. *Front. Young Minds*. 9:581824. doi: 10.3389/frym.2021.581824

Archibald, J. (2022a). Segmental and prosodic evidence for property-by-property phonological transfer in L3 English in northern Africa. *Languages* 7, 28. doi: 10.3390/languages7010028

Archibald, J. (2022b). Phonological parsing via an integrated I-language: the emergence of property-by-property transfer effects in L3 phonology. *Linguist. Approaches Biling*. doi: 10.1075/lab.21017.arc

Archibald, J. (2023). "Using a contrastive hierarchy to formalize structural similarity as I- proximity in L3 phonology," in *Linguistic Approaches to Bilingualism*, eds N. Kolb, N. Mitrofanova, and M. Westergaard. Special Issue. doi: 10.1075/lab.22051.arc

Archibald, J. (in press). "A transition theory of L3 segmental phonology: phonological features and phonetic variation in multilingual grammars," in *Multilingual Acquisition and Learning: Towards an Eco-systemic View of Diversity*, ed. E. Babatsouli (Amsterdam: John Benjamins).

Baker, W., and Smith, C. (2010). "Acquiring the high vowel contrast in Quebec French: how assibilation helps," in *Achievements and Perspectives in SLA of Speech: New Sounds 2010*, eds M. Wrembel, M. Kul, and K. Dziubalska-Kołaczyk (Brussels: Peter Lang), 279–288.

Bansal, R. K. (1969). *The Intelligibility of Indian English: Measurements of the Intelligibility of Connected Speech, and Sentence and Word Material, Presented to Listeners of Different Nationalities*. Hyderabad: Central Institute of English.

Benrabah, M. (1991). "Learning English segments with two languages," in *Proceedings of the International Congress of Phonetic Sciences*, ed P. Romeas (Aix-en-Provence: Université de Aix-en-Provence), 334–337.

Bertrand, R., Blache, P., and R. Espesser, R. (2007). CID: corpus of interaction data. *Trav. Interdiscip. Lab. Parole. Lang. d'Aix-en-Provence* 25, 25–55.

Best, C., and Tyler, M. (2007). "Nonnative and second-language speech perception," in *Language Experience in Second Language Speech Learning*, eds O.-S. Bohn and M. Munro (Amsterdam: John Benjamins), 13–34. doi: 10.1075/lllt.17.07bes

Bohn, G., and Santos, R. (2018). The acquisition of pre-tonic vowels in Brazilian Portuguese. *Alfa* 62, 191–221. doi: 10.5334/jpl.184

Brannen, K. (2002). The role of perception in differential substitution. *Can. J. Linguist.* 47, 1–46.

Brannen, K. (2011). *The Perception and Production of Interdental Fricatives in Second Language Acquisition* [Unpublished doctoral dissertation]. Montréal, QC: McGill University.

Brown, C. (2000). "The interrelation between speech perception and phonological acquisition from infant to adult," in Second Language Acquisition and Linguistic Theory, ed. J. Archibald (Oxford: Blackwell), 4–63.

Chandlee, J. (2023). Decision trees, entropy, and the contrastive feature hierarchy. *Poster at the LSA*. Denver Colorado. doi: 10.3765/plsa.v8i1.5465

Colantoni, L., Kochetov, A., and Steele, J. (2021). Coronal stop lenition in French and Spanish: electropalatographic evidence. *Loquens* 8, 1–15. doi: 10.3989/loquens.2021.080

Collins, B., and Mees, I. (1981). *The Sounds of English and Dutch*. Leiden: Leiden University Press.

Coté, M.-H. (2021). "L'adaptation de emprants de l'anglais en français: variation dialectale, phonologie, lexicographie," in *L'adaptation de Manlleus en Catalàien Altres Llengües Romàniques*, eds M.-R. Lloret, and C. Pons-Moll (Barcelona: Publicaciones i Edicions de la Universitat de Barcelona).

Cowper, E., and Hall, D. C. (2019). "Scope variation in contrastive hierarchies of morphosyntactic features," in *Variable Properties in Language: Their Nature and Acquisition*, ed. D. W. Lightfoot, and J. Havenhill (Washington, DC: Georgetown University Press), 27–41. doi: 10.2307/j.ctvfxv99p.7

Cummins, R. (1983). *The Nature of Psychological Explanation*. Cambridge: MIT Press.

Dresher, B. E. (2009). *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511642005

Dresher, B. E. (2018). "Contrastive hierarchy theory and the nature of features," in *Proceedings of the 35th West Coast Conference on Formal Linguistics*, eds W. Bennett, L. Hracs, and D. Storoshenko (Somerville, MA: Cascadilla Proceedings Project), 18–29.

Dresher, B. E., and Hall, D. C. (2020). The road not taken: the sound pattern of Russian and the history of contrast in phonology. *J. Linguist.* 57, 405–444. doi: 10.1017/S0022226720000377

Flege, J., and Bohn, O. (2021). "The revised Speech Learning Model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed. R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002

Flege, J. E. (1995). "Second language speech learning: theory, findings and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-language Research*, ed. W. Strange (Baltimore, MD: York Press), 233–277.

Gatbonton, E. (1978). Patterned phonetic variability in second-language speech: a gradual diffusion model. *Can. Mod. Lang. Rev.* 34, 335–347. doi: 10.3138/cmlr.34.3.335

Goad, H., and Mah, J. (2007). "Rethinking English /h/," in *Paper presented at Montreal-Ottawa-Toronto (MOT) Phonology Workshop*. University of Ottawa.

Gonález Poot, A. (2014). "Conflict resolution in the Spanish L2 acquisition of Yucatec ejectives: L1, L2, and Universal constraints," in *Proceedings of the 2014 Annual Conference of the Canadian Linguistic Association*, ed. L. Teddiman (Ottawa, ON: Canadian Linguistic Association).

González Poot, A. (2011). *Conflict Resolution in the Spanish SLA of Yucatec Ejectives: L1, L2, and Universal Constraints* [Ph.D. dissertation]. Calgary, AB: University of Calgary.

Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: dispersedness without dispersion. *Phonology* 28, 1–54. doi: 10.1017/S0952675711000029

Hall, D. C. (2017). "Contrastive specification in phonology," in *Oxford Research Encyclopedia of Linguistics*, ed M. Aronoff (Oxford: Oxford University Press). doi: 10.1093/acrefore/9780199384655.013.26

Hancin-Bhatt, B. (1994). Segmental transfer: a consequence of a dynamic system. *Second Lang. Res.* 10, 241–269. doi: 10.1177/026765839401000304

Hanuliková, A., and Weber, A. (2012). Sink positive: linguistic experience with the substitutions influences nonnative word recognition. *Atten. Percept. Psychophys.* 74, 613–629. doi: 10.3758/s13414-011-0259-7

Herd, J. (2005). "Loanword adaptation and the evaluation of similarity," in *Toronto Working Papers in Linguistics 24*, 65–116. Available online at: https://twpl.library.utoronto.ca/index.php/twpl/article/view/6195 (accessed July 11, 2022).

Hsu, B., and Jesney, K. (2017). "Loanword adaptation in Québec French: evidence for weighted scalar constraints," in *Proceedings of the 34th West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project*, eds A. Kaplan, A. Kaplan, M. K. McCarvel, and E. J. Rubin (Somerville, MA: Cascadilla Proceedings Project), 249–258.

Hunt Gardner, M., and Roeder, R. (2022). Phonological mergers have systemic phonetic consequences: PALM, trees, and the low back merger shift. *Lang. Var. Change* 34, 29–52. doi: 10.1017/S0954394522000059

Itô, J., and Mester, A. (1995). "The core-periphery structure of the lexicon and constraints on reranking," in *Papers in Optimality Theory*, eds J. Beckman, S. Urbanczyk, and L. Walsh Dickey (Amherst, MA: GLSA), 181–209.

Jackson, S., and Cardoso, W. (2023). "Orthographic interference in the acquisition of English /h/ by francophones," in *Second Language Pronunciation: Different Approaches to Teaching and Training*, eds U. Alves and J. Alcanatara de Albuquerque (Berlin: De Gruyter). doi: 10.1515/9783110736120-009

James, A. (1988). *The Acquisition of a Second-Language Phonology*. Gunter Narr Verlag.

Janda, R., and Auger, J. (1992). Quantitative evidence, qualitative hypercorrection, sociolinguistic variables – and French speakers' 'eadaches with English h/ø. *Lang. Commun.* 12, 195–236. doi: 10.1016/0271-5309(92)90015-2

Jensen, J. T. (1993). *English Phonology*. Amsterdam: Benjamins. doi: 10.1075/cilt.99

John, P., and Frasnelli, J. (2023). On the lexical source of variable L2 phoneme production. *Mental Lexicon* 17, 239–276. doi: 10.1075/ml.22002.joh

Kabak, B. (2019). "A dynamic equational approach to sound patterns in language change and second-language acquisition: the (un)stability of English dental fricatives illustrated," in *Patterns in Language and Linguistics: New Perspectives on a Ubiquitous Concept*, eds B. Busse, and R. Moehlig-Falke (Berlin: Mouton De Gruyter), 221–254. doi: 10.1515/9783110596656-009

Kohler, K. J. (2002). "Segmental reduction in connected speech in German: phonological facts and phonetic explanations," in *Speech Production and speech Modelling*, eds A. Hardcastle, and A. Marchal (Dordrecht: Springer), 69–92. doi: 10.1007/978-94-009-2037-8_4

Kubozono, H. (2015). "Introduction to Japanese phonetics and phonology," in Handbook of Japanese Phonetics and Phonology, ed. H. Kubozono (Berlin: De Gruyter), 1–40. doi: 10.1515/9781614511984.1

Kwon, J. (2021). Defining Perceptual Similarity with Phonological Levels of Representation*: Feature (Mis)match in Korean and English* [Ph.D. dissertation]. Madison, WI: University of Wisconsin-Madison.

Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199545834.001.0001

LaCharité, D. (1993). *The Internal Structure of Affricates* [Ph.D. Dissertation]. University of Ottawa, ON: University of Ottawa.

LaCharité, D., and Prévost, P. (1999). "The role of L1 and of teaching in the acquisition of English sounds by francophones," in *Proceedings of the 23rd Annual Boston University Conference on Language Development*, eds A. Greenhill, H. Littlefield, and C. Taro (Somerville, MA: Cascadilla Press), 373–385.

Lahiri, A., and Reetz, H. (2002). "Underspecified recognition," in *Laboratory Phonology 7*, eds C. Gussenhoven, and N. Warner (Berlin: MoutonDe Gruyter), 637–675. doi: 10.1515/9783110197105.2.637

Lombardi, L. (2003). Second language data and constraints on Manner: explaining substitutions for the English interdentals. *Second Lang. Res.* 18, 225–250. doi: 10.1177/026765830301900304

Mah, J., Goad, H., and Steinhauer, K. (2016). Using event-related. Brain potentials to assess perceptibility: the case of French speakers and English [h]. Front. Psychol. 7, 1469. doi: 10.3389/fpsyg.2016.01469 doi: 10.3389/fpsyg.2016.01469

McCarthy, J. (1994). "The phonetics and phonology of Semitic pharyngeals," in *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, ed. P. Keating (Cambridge: Cambridge University Press), 191–233. doi: 10.1017/CBO9780511659461.012

Natvig, D., and Salmons, J. (2021). Connecting structure and variation in sound change. *Cad. Pesqui.* 2, 1–20. doi: 10.25189/2675-4916.2021.v2.n1.id314

Nemser, W. (1971). *An Experimental Study of Phonological Interference in the English of Hungarians*. Bloomington, IN: Indiana University Press.

Niebuhr, O., Clayards, M., Meunier, C. and Lancia, L. (2011). On place assimilation in sibilant sequences: comparing English and French. *J. Phon.* 39, 429–451. doi: 10.1016/j.wocn.2011.04.003

Oxford, W. (2015). Patterns of contrast in phonological change: evidence form Algonquian vowel systems. *Language* 91, 308–357. doi: 10.1353/lan.2015.0028

Paradis, C., and LaCharité, D. (1997). Preservation and minimality in loanword adaptation. *J. Linguist.* 33, 379–430. doi: 10.1017/S0022226797006786

Picard, M. (2002). The differential substitution of English /θ ð*/ in French. *Lingvisticae Investig.* 25, 87–96. doi: 10.1075/li.25.1.07pic

Purnell, T., Raimy, E., and Salmons, J. (2019). Old English vowels: diachrony, privativity, and phonological representations. *Language* 95, e447–e473. doi: 10.1353/lan.2019.0083

Rice, K. (2005). "Sequential voicing, postnasal voicing, and Lyman's Law revisited," in *Voicing in Japanese*, eds J. van de Weijer, K. Nanjo, and T. Nishihara (Berlin: Mouton de Gruyter), 25–45. doi: 10.1515/9783110197686.1.25

Rochet, B. (1995). "Perception and production of second-language speech sounds by adults," in *Speech Perception and Linguistic Experience*, ed. W. Strange (Baltimore, MD: York Press), 379–410.

Rose, S. (1996). Variable laryngeals and vowel lowering. *Phonology* 13, 73–117. doi: 10.1017/S0952675700000191

Roy, M.-J. (1992). *Le role des Constraints Phonologiques dan l'adaptation d'emprunts Anglaise en Français Québécois* [MA Thesis]. Quebec City: Université Laval.

Smith, L. C. (1997). "The role of L1 feature geometry in the acquisition of L2 segmental phonology: acquiring /θ/ and /ð*/ in English," in *Calgary Working Papers in Linguistics, Volume 19* (Calgary: University of Calgary), 45–70.

Stockwell, R., and Bowen, J. (1965). *The Sounds of English and Spanish*. Chicago, IL: University of Chicago Press.

Teasdale, A. (1997). "On the differential substitution of English [θ]: a phonetic approach," in *Calgary Working Papers in Linguistics, Volume 19* (Calgary: University of Calgary), 71–91.

Telfer, C. (2006). *Coronalization as Assibilation* [MA thesis]. Calgary, AB: University of Calgary.

Trofimovich, P., and John, P. (2011). "When three equals tree: examining the nature of phonological entries in L2 lexicons of Quebec speakers of English," in *Applying Priming Methods to L2 learning, Teaching and Research: Insights from Psycholinguistics*, eds P. Trofimovich, and K. McDonough (Amsterdam: John Benjamins), 105–129. doi: 10.1075/lllt.30.09tro

Vance, T. (1987). *An Introduction to Japanese Phonology*. Albany, NY: SUNY Press.

Vance, T. (2008). *The Sounds of Japanese*. Cambridge: Cambridge University Press.

Walker, D. C. (1984). *The Pronunciation of Canadian French*. Ottawa, ON: University of Ottawa Press.

Weinberger, S. H. (1997). "Minimal segments in second language phonology," in *Second-Language Speech: Structure and Process*, eds A. James, and J. Leather (Berlin: Mouton de Gruyter), 263–311. doi: 10.1515/9783110882933.263

Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. The Hague: Mouton.

Weinreich, U. (1957). On the description of phonic interference. *Word* 13, 1–11. doi: 10.1080/00437956.1957.11659624

Wexler, K., and Culicover, P. (1983). *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.

White, E. J., Titone, D., Genesee, F., and Steinhauer, K. (2015). Phonological processing in late second language learners: the effects of proficiency and task. *Biling.: Lang. Cogn.* 20, 162–183. doi: 10.1017/S1366728915000620

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 06 November 2023
DOI 10.3389/flang.2023.1253816

OPEN ACCESS

EDITED BY
Jennifer Cabrelli,
University of Illinois Chicago, United States

REVIEWED BY
Charles B. Chang,
Boston University, United States
Kyle Parrish,
Goethe University Frankfurt, Germany

*CORRESPONDENCE
Brett C. Nelson
✉ brett.nelson@ucalgary.ca

RECEIVED 06 July 2023
ACCEPTED 13 October 2023
PUBLISHED 06 November 2023

CITATION
Nelson BC (2023) Phonological redeployment for [retracted tongue root] in third language perception of Kaqchikel stops.
Front. Lang. Sci. 2:1253816.
doi: 10.3389/flang.2023.1253816

Phonological redeployment for [retracted tongue root] in third language perception of Kaqchikel stops

Brett C. Nelson*

Division of Linguistics, School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

Phonological redeployment is the theoretical ability of language learners to utilize non-local phonological knowledge from known languages in the mapping and acquisition of novel contrasts in their target languages. The current paper probes the limits of phonological redeployment in a third language acquisition scenario. The phonological features [Advanced Tongue Root] and [Retracted Tongue Root] capture a range of phonological contrasts and harmony processes in both vowels and consonants of spoken languages across the world, including, but not limited to, vowel tensing and post-velar places of articulation (e.g. uvular). Kaqchikel (cak) exhibits both a tense-lax vocalic contrast in its vowels plus a velar-uvular Place contrast in its eight stop consonant phonemes. English (eng) exhibits a tense-lax vocalic distinction but no velar-uvular distinction among its six stop phonemes. Spanish (spa) exhibits neither of these contrasts in its vowels or among its six stop phonemes. How do multilingual learners of Kaqchikel already familiar with English and Spanish, but who differ in which is their first language (L1), compare in their categorical perception of Kaqchikel stop consonants? Despite English and Spanish having a three-way Place distinction among stops in common, in a phonemic categorization task, L1 English learners of Kaqchikel were better at correctly categorizing audio recordings of Kaqchikel uvular stops than L1 Spanish learners of Kaqchikel. To account for this surprising result, I propose that the L1 English group have easier access than the L1 Spanish group to the feature underlying English's tense-lax distinction. This access allows them to redeploy that phonological feature to accurately map out the novel four-way contrast of Kaqchikel's stop consonants, and the [±RTR] specified velar-uvular distinction in particular. Therefore, phonological redeployment must be considered in models of third language acquisition.

KEYWORDS

third language acquisition, phonology, redeployment, stop consonants, post-velar consonants, Kaqchikel, Spanish, English

1. Introduction

The phonological tongue root features [Advanced Tongue Root] (or [ATR]) and [Retracted Tongue Root] ([RTR]) capture a range of phonological contrasts and harmony processes in both vowels and consonants of spoken languages across the world (Beltzung et al., 2015). Among these are relatively less common vocalic contrasts such as /e–ɛ/ and /o–ɔ/, which are often involved in harmony processes, particularly among African languages, and the pharyngealization of oral consonants in Semitic and Salish languages (Davis, 1995; Shahin, 2002; Abo-Mokh and Davis, 2020). This relationship between tongue retraction and pharyngeals also implicates [RTR] in phonological contrasts among consonants articulated in and around the pharynx, including uvular, pharyngeal, and other post-velar consonants (Rose, 1996) along with harmony processes associated with those consonants (Sylak Glassman, 2014).

Kaqchikel (ISO 639-3: cak), a Mayan language used mostly by about 400,000 Kaqchikel people in the highlands of Guatemala and in diaspora across North America (Heaton and Xoyón, 2016), exhibits a contrast among its 10 vowels in which a series of 5 tense vowels is not specified for [RTR], while another series of (up to) 5 lax vowels is specified for [RTR]. This contrast is neutralized in unstressed syllables (the feature is lost), so that only tense vowels surface in those positions (Rill, 2013; Bennett, 2016). Among Kaqchikel's stops, there is also a velar–uvular contrast, which may similarly be derived by [RTR] (Shahin, 2002).

English (eng), a Germanic language used by billions of people across the world, similarly exhibits a contrast among its vowels in which two series are contrasted: one tense and one lax. While this contrast is typically attributed to a [tense] feature (Kim and Clements, 2015), Beltzung et al. (2015) notes that [tense] and [ATR] lead to nearly identical outcomes cross-linguistically. Moreover, some analyses have attributed this contrast directly to the feature [RTR] (Brown and Golston, 2006). However, unlike Kaqchikel, English only contrasts three places of articulation (PoA) among its stops: labial, coronal, and velar, each of which may be specified by just a single corresponding Place feature.

Spanish (spa), a Romance language used by hundreds of millions of people across the world, differs from both Kaqchikel and English by exhibiting only a single series of five vowels /i e a o u/, contrasting only on height and backness (Torres-Tamarit, 2019). No [tense] or tongue root feature is present in Spanish, and this can lead to difficulties learning vowel contrasts in other languages which do use those features, as examined in Escudero (2005), among many others. However, Spanish is similar to English in only contrasting three PoA among its stops, again differing from Kaqchikel's four-way contrast.

In this study of third language (L3) acquisition (L3A) of Kaqchikel by learners who already use English and Spanish, listeners were divided into groups based on their first language (L1), in order to investigate the effects of L1 on L3 phonological perception of stop consonant contrasts (Nelson, forthcoming).[1] The primary research question of the current study asks whether these learner groups differ in their categorical perception of Kaqchikel stops based on differential access to the [RTR] feature in their known languages.

Abbreviations: AIC, Akaike information criterion; [±ATR], [±advanced tongue root] phonological feature; AoL, age of learning; β, regression coefficient (standardized); C, consonant; cak, Kaqchikel (ISO 639-3 code); CI, confidence interval; *DF*, degrees of freedom; eng, English (ISO 639-3 code); F, female; L1, first language; L1A, first language acquisition; L2, second language; L2A, second language acquisition; L2LP, Second Language Linguistic Perception Model (Escudero, 2005); L3, third language; L3A, third language acquisition; M, male; n, sample size; NES, native (L1) English speakers; NKS, native (L1) Kaqchikel speakers; NSS, native (L1) Spanish speakers; p, probability value; PAM, Perceptual Assimilation Model (Best, 1995); PAM-L2, Perceptual Assimilation Model–Second Language (Best and Tyler, 2007); PPH, Phonological Permeability Hypothesis (Cabrelli Amaro, 2013b); R^2, coefficient of determination; [±RTR], [±retracted tongue root] phonological feature; *SD*, standard deviation; *SE*, standard error; Sig., Significance level; SLM, Speech Learning Model (Flege, 1995); spa, Spanish (ISO 639-3 code); V, vowel; χ^2, chi-squared statistic; z, standard score (for Wald test).

2. Background

2.1. Acquisition theory

2.1.1. Second vs. third language acquisition

Both subfields of L2A and L3A deal with the process of adding an additional language system to a person's repertoire of linguistic knowledge. Both ask questions about how previous knowledge impacts the implementation of new linguistic knowledge, and vice-versa: how new knowledge affects how previous linguistic knowledge is utilized. Both have various models and hypotheses within them that posit different causes for variable relative difficulty of language learning across different learners of the same language.

The two subfields differ as to which iteration of language acquisition the process is being studied. For L2A, the process being studied is an additional language learned later in life than the first language of that learner, usually after a certain point in the life of said learner. Therefore, a primary difference among learners is that first language. For L3A, on the other hand, the minimum prerequisite for the process being studied is the presence of multiple (minimally two, but, of course, there is the possibility for more) language systems in the linguistic knowledge of the learner. That is to say, for L3A there is additional complexity due to the increased number of potential sources and directions of influence between the multiple existing language systems and the new, target L3 system.

2.1.2. Second language acquisition

2.1.2.1. Basics of second language acquisition

The study of L2A has a relatively long history within modern linguistic inquiry, going back to at least the 1940s. A primary object of study in field of L2A is *interlanguage*, coined by Selinker (1972) to describe a speaker's L2 competence that was distinct from both that speaker's L1 competence as well as the competence of a native speaker of the target L2, though obviously being influenced by both. It is built upon three main processes: that of *L1 transfer*, *overgeneralization of L2 patterns*, and *fossilization*. According to the original theory of interlanguage, 95% of learners do not acquire native-like competence and thus continue to have an interlanguage after all learning has concluded (VanPatten and Benati, 2010), thus the expected result of L2A, or even L3A, cannot be native-like fluency.

Narrowing our focus, the interlanguage phonology of a given learner is influenced by a variety of factors, including similarities and differences between their known L1 and target L2 (see, e.g., Lado, 1957), markedness of features and structures they are

1 This study constitutes a part of a larger study investigating the acquisition of the Place and Laryngeal contrasts of Kaqchikel stops in various syllabic and word Positions by Spanish-English multilingual learners. The other parts of this study are briefly discussed in §3.3.

learning (Eckman, 1977), their age (Flege, 1995), and their specific dialectal experience in both L1 and L2 (Best and Tyler, 2007).

2.1.2.2. Phonological segments in L2A

Following decades of L2A investigation, Flege (1995) proposed the Speech Learning Model (SLM)[2] finding a correlation between a learner's age of learning (AoL) and their inability to "produce L2 vowels and consonants in a native-like fashion" (237). This segment-specific approach to the L2A of spoken languages' sound systems posits that foreign-accentedness arises from difficulty in learning, which in turn arises from acoustic similarity of L2 segments to known L1 segment categories. That is to say, L2 segments perceptually distinct from L1 segments are actually easier to identify, and therefore learn, than L2 segments perceived as being similar to L1 segments.

The Perceptual Assimilation Model-L2 (PAM-L2) of Best and Tyler (2007) contrasts with SLM, holding that perception of languages other than L1 is done within L1 categories, as opposed to the comparison of L2 segments to L1 segments in SLM. As such, according to PAM-L2, learners form their interlanguage by assimilating L2 segments into L1 categories, which may cause interference in the development of contrasts between L2 segments.

Best and Tyler (2007) predicted four possible cases of L1 assimilation for a set of contrasting segments in L2 (28–29):

1. *Two-Category* (TC): Only one L2 category is perceived as equivalent to a given L1 category.
2. *Category Goodness* (CG): Multiple L2 categories perceived as equivalent to the same L1 category, but one is a better fit than other(s).
3. *Single Category* (SC): Multiple L2 categories are perceived as equivalent to the same L1 category, but as equally good instances of it.
4. *Uncategorized*: No L1-L2 assimilation.

Best and Tyler (2007) ordered these in increasing order of difficulty, with TC Assimilations being easier to learn than CG Assimilations, which should be easier than SC Assimilations. Uncategorized contrasts, however, are subject to more variation, depending on the perceptual distance from each L2 category to known L1 categories. Long-term learning outcomes for the SC and unassimilated L2 segments would also depend upon lexical-functional differences among them.

Escudero (2005) proposed a model similar to PAM-L2 with L2 Linguistic Perception (L2LP), under which phonological L2A is initiated using L1 categories. However, mappings are made via auditory perception. L2LP predicts three scenario types, *new*, *subset*, and *similar*, and these scenarios range in the level of difficulty they present the learner from high to low, based on the nature of categorical remapping each requires. Thus, in both PAM-L2 and L2LP, the conflict that arises when L2 segments are categorized based on L1 categories using L1 cues is the primary contributor to interference in phonological L2A.

The SLM, PAM-L2, and L2LP were put to the test in the scenario of L1 English learners of Q'eqchi', a Mayan language related to Kaqchikel (the target language of the current study), in Wagner and Baker-Smemoe (2013), who compared L1 English to L1 Q'eqchi' participants in both perception and production of Q'eqchi' plain and ejective stops. They found that while none of the three models were perfect in their predictions of stop production based on perception results, they each "to some extent, predicted learning accuracy" (466).

According to Wagner and Baker-Smemoe (2013), SLM accurately predicted that L1 English learners would differ more from L1 Q'eqchi' speakers in their production of plain stops, relative to glottalized stops, but failed to predict their better ability in producing native-like cues for velars compared to other PoAs. PAM-L2 predicted the Laryngeal distinction at both velar and uvular PoAs to each be difficult to acquire SC categorizations, but the learners did not show difficulty with these pairs. PAM-L2 predicted ease of learning in the CG categorizations of the velar–uvular Place distinction, but Wagner and Baker-Smemoe (2013) found that this place distinction was more difficult than the laryngeal distinction. Similarly, L2LP made similar incorrect predictions as PAM-L2 regarding the relative difficulty of learning the Laryngeal distinction compared to the velar–uvular place distinction.

Wagner and Baker-Smemoe (2013) provide a possible explanation for the surprising results of the ease of learning distinctions among the velar and uvular stops in that the mapping of four categories onto a single L1 category forced learners to focus more intently on those contrasts, leading to better than expected learning outcomes for the velar and uvular stops of Q'eqchi' by L1 English learners. These results bear on the current study, as Kaqchikel, like Q'eqchi', contrasts uvular stops with velar stops. These outcomes should be replicated in the L3A scenarios of the current study.

2.1.2.3. Phonological features in L2A

Other models of L2A posit a filter effect that prevents some learners from developing adequate phonological mapping in their target language. This filter is caused by the supposed unlearnability of underlying features, the bundles of information that specify phonological contrasts of human language in theories of generative phonology (Chomsky and Halle, 1968). Building upon this, Clements (1985) and others developed an extension of feature theory called Feature Geometry, which holds that the underlying features exist in a hierarchical relationship to one another. Much of the work reviewed in the remainder of this background section depends on Feature Geometry, and I assume this theory for the current study as well.

Brown (1997, 1998, 2000) analyzes the relative difficulty of learning the English liquid contrast /l∼ɹ/ by L1 Japanese learners compared to L1 Mandarin Chinese[3] learners. Brown claims that Japanese and Korean feature geometries lack the [Coronal] feature necessary to specify the distinction between the English liquids, while Mandarin does use this feature, so L1 Mandarin learners of English are able to use this feature in acquiring /l∼ɹ/, leading to differences in outcomes between learners differing in L1 language background (Brown, 2000).

2 SLM was later revised in Flege and Bohn (2021). Though not immediately relevant here, I return to this in §5.1.5.

3 Alternatively called Standard Chinese, as a specific standardized variety of Mandarin. For consistency, I refer to this language as Mandarin, as Brown (2000) and later Yang et al. (2022) do.

LaCharité and Prévost (1999) hypothesized a weaker filter, in which only some features are unlearnable in L2A: those at intermediate, articulator nodes in the feature geometry. Terminal nodes projecting no dependent nodes, on the other hand, are learnable. Thus, it is unsurprising to observe that L1 Japanese learners of English seem unable to acquire the [Coronal] node in learning liquids, but L1 French learners of English more easily acquire English /θ/ compared to /h/ because /θ/ requires the addition of a terminal [distributed] node, while /h/ requires the addition of articulator [Pharyngeal] node. Moreover, LaCharité and Prévost (1999) predicted the acquisition of /ŋ/ by these same learners to be even easier than both /θ/ and /h/, as no new features need to be added to the feature geometry to specify /ŋ/.

Mah (2003) found the attribution of [Pharyngeal] to English /h/ to be unfounded as it "does not involve any constriction of the pharyngeal cavity" (24). Additionally, there is evidence that a [Pharyngeal] feature is in fact utilized in French, specifying its rhotic /ʁ/ (Mah, 2003). Instead, Mah (2011) followed Iverson and Salmons (1995) in specifying English /h/ with a Laryngeal node projecting the terminal node [spread glottis], which further problematizes the predictions and findings of LaCharité and Prévost (1999), as French stops use the Laryngeal node in representing voiceless and voiced obstruents. Thus, only the terminal node of the feature [spread glottis] would need to be added when francophones learn English /h/. Brown (2000) would predict this as impossible, but LaCharité and Prévost (1999) would predict this as possible but not as easy as other scenarios. Yet, Mah (2011) and Mah et al. (2016) again found that francophone learners of English are unable to perceive English /h/, despite being able to detect its acoustic cues in non-linguistic conditions. Therefore, Mah (2011) concluded that the learning problem arises due to francophones' inability to form a phonological representation for English /h/ in their interlanguage, missing the key feature [spread glottis], as Brown (2000) would predict.

2.1.2.4. Phonological redeployment in L2A

The filter hypotheses predict impossibility of learning based on what they assume to be irreconcilable disparities between underlying feature geometries. To account for the remarkable ability of some language learners to learn patterns and contrasts that these hypotheses and other theories of L2A predict as difficult, Archibald (2005) proposed that learners may dynamically redeploy previous linguistic knowledge to remedy their lack of the specific linguistic knowledge that L1 users of their target L2 have, including, but not limited to, phonological features.

González Poot (2011, 2014) offered redeployment as a potential facilitator in L2A of glottalized stops in Yukatek, another Mayan language distantly related to Kaqchikel. González Poot (2011) noted the differential use of the feature [constricted glottis] among Yukatek, English, and Spanish. In Yukatek, [constricted glottis] is distinctive, underlying the contrast between plain and glottalized stops. However, in Spanish, [constricted glottis] has no status whatsoever, while in English it may be used in word-final allophonic ejectives. González Poot (2011, 2014) did not investigate English learners of Yukatek, leaving open the possibility of redeployment of non-contrastive features, like [constricted glottis] in English, for future research. Nevertheless, González Poot (2011, 2014) found that the L1 Spanish L2 learners of Yukatek acquired its Laryngeal contrast, which González Poot (2014) credits to the ejectives' strong acoustic cues.

More recently, Yang et al. (2022) found that L1 Mandarin learners of Russian perceived Russian voiced stops as being highly similar to Mandarin voiceless unaspirated stops, which the SLM would predict would lead to difficulty in their L2A of Russian stops. Surely enough, in their productions of Russian-like nonce words, the learners did not produce a voice onset time (VOT) distinction between the phonemically voiced and voiceless stops, producing them both as short-lag stops, as if they all belonged to a single category of voiceless, unaspirated stops. They had no representation of the voicing distinction of Russian stops. Yang et al. (2022) interpreted these results as a refutation of redeployment theory. In their view of redeployment, the learners ought to have been able to redeploy [±voice] from Mandarin fricatives /ʂ/ and /ʐ/ in specifying Russian stops.[4]

In reviewing Yang et al. (2022) and Archibald (2023) clarified key points about redeployment. First, that redeployment is not phonetic, but rather phonological; it is embedded within learners' phonological systems. Second, that relatively robust cues to contrasts in the target language allow learners to better notice those contrasts, while contrasts with weaker cues may not get noticed. However, Archibald (2023) principal point was that noticing phonetic differences across languages is not the outcome of successful phonological learning. Instead it is an important preliminary step to phonological learning, which *may* be aided by redeployment of previous phonological knowledge and which *may* involve re-weighting of acoustic cues to account for the new contrasts they have begun to notice.

2.1.3. Third language acquisition

2.1.3.1. Basics of third language acquisition

Third language acquisition, when compared to L2A, is a younger field of study. While L2A has many decades of research behind it, L3A only has two or three decades of specific research, as it only arose out of the field of L2A in the 1990s. In addition to the L2A concept of interlanguage, an additional object of study in L3A is the concept of cross-linguistic influence: how the multiple different systems interact in the mind of the multilingual language user/learner (Hammarberg, 2001). Logically, any of a learner's language systems may interact and have cross-linguistic influence on any of their other language systems. However, theories differ as to the nature and amount of cross-linguistic influence possible in each direction of a given L3A scenario.

---

4 The status of [±voice] as a distinctive feature in Mandarin /ʂ/ and /ʐ/ is not straightforward, however. Duanmu (2007) and Lin (2007), both cited in Yang et al. (2022), offered differing accounts of the contrast. Lin (2007) transcribed Yang et al. (2022)'s /ʐ/ as /ɹ/, describing it as a voiced approximant, and later a liquid sonorant that "tends to become a voiced fricative" (47). Duanmu (2007) transcribes this segment as /ʐ/, but did not definitively describe its distinction from /ʂ/ as one of voicing, but rather of aspiration (24). Thus, I arrive at the same conclusion as Archibald (2023): there is no solid motivation for the distinctive status of [±voice] in Mandarin.

2.1.3.2. Phonological effects in L3A

As briefly discussed in §2.1.2, Mah (2003) investigated the filter effects proposed for phonological L2A. However, Mah (2003) did not find any differences in the perception of French vs. Spanish trills (/ ʀ/ and /r/, respectively) by L1 English learners, concluding that these learners were unable to construct appropriate phonological representation for either. Mah (2003) also noted that L2 exposure to non-L1 segments may only affect processing of the segment in L3 only when that exposure was in a childhood L2 and sufficient enough to learn the features necessary for representing the segment, implicating the AoL effect in L2A on subsequent L3A.

Cabrelli Amaro and Rothman (2010), noting similar age-related effects of L2A on L3A, offered the Phonological Permeability Hypothesis (PPH) of L3A. Later expanded upon in Cabrelli Amaro (2013b, 2017), the premise of PPH lies in the AoL effect, that there exists a critical/sensitive period of phonological learning in pre-pubescence, and that languages learned during this period are fundamentally different than languages learned after it. Thus, PPH predicts similarities among languages learned in childhood, and a separate set of similarities among languages learned in adulthood. Evidence for PPH comes in the form of relative susceptibility to regressive transfer from L3 during L3A: L2 is more permeable to this cross-linguistic influence than L1 is (Cabrelli Amaro, 2017). This implicates the conceptual similarity between L2 and L3 being greater than that between L1 and L3, as well as the higher degree of phonological fossilization present in systems learned in childhood as compared to systems learned in adulthood. In short, languages learned during childhood are more entrenched in a person's greater phonological system than subsequently learned languages.

Wrembel et al. (2019), noting previous studies showed that L2A enhances auditory awareness in L3A, offered an extension of PAM to the domain of phonological L3A. In their study of the acquisition of Polish by teenagers who were multilingual in other European languages, Wrembel et al. (2019) found that learners did operate under the assumptions of PAM, and that pairs of L3 and L2 sounds (including both consonants and vowels) were assimilated more often than pairs of L3 and L1 sounds. Therefore the models of phonological L2A do seem to have some predictive power in L3A in that comparisons of L3 categories are made to known categories, but learner do prefer assimilating L3 categories into L2 categories rather than those of L1.

2.1.3.3. Selective transfer in phonological L3A

Most recently, Archibald (2022) analyzed the L3A of English by L1 Arabic, L2 French learners in Algeria and Tunisia (as presented by Benrabah, 1991 and Ghazali and Bouchhioua, 2003), arguing that phonological transfer in L3A comes from both/all known languages on a property-by-property basis (as opposed to wholesale) along the lines of the Linguistic Proximity Model (Westergaard et al., 2017). Specifically, the learners of L3 English were found to transfer the vocalic system from French, but their consonants, including pharyngealized stops, from Arabic. Furthermore, the Tunisian novice-level English learners in Ghazali and Bouchhioua (2003) seemed to transfer sentence-level prominences from French, but word-level stress rules from Arabic.

Archibald (2022) used the Contrastive Hierarchy of Dresher (2009) and Feature Geometry to show that learners select different sources for their phonological L3A transfer on a property-by-property basis. Learners do this based on the evidence available to them in their learning environments, the knowledge from the integrated I-grammar of their known languages, and the general constraints provided by Universal Grammar. In the end, learners tend to make the decision to transfer the phonological subsystem which most optimally accounts for the L3 contrasts they can observe. Note, however, that they still must be able to observe or notice the contrasts, as per Archibald (2023), before they begin to integrate them into their I-grammar's phonology.

As theories of phonological learning, especially those allowing for redeployment, depend on underlying specification of segments, the following subsection makes clear the featural specifications that I assume for each of the three languages of the current study.

2.2. Phonological background

2.2.1. Kaqchikel phonology

2.2.1.1. Kaqchikel vocalic phonology

Kaqchikel's vocalic phonology is typical of Mayan languages in its basis as a five-vowel inventory. However, it is atypical of Mayan languages in that it does not contrast these five base vowels for length, instead exhibiting a tense–lax contrast. The maximal ten-vowel inventory of Standard Kaqchikel is shown in Figure 1A. The tense–lax contrast of Kaqchikel vowels only surfaces in stressed syllables. Outside of stressed syllables, only tense vowels surface (Rill, 2013). In practice, due to the location of stress in Kaqchikel being fixed to the final syllable of the word, lax vowels may only surface in the final syllable of words (Brown et al., 2006).

Most varieties of Kaqchikel, however, do not mark the tense–lax contrast for all five vowel archiphonemes, and exhibit mergers of the tense–lax distinction in some or even all vowels. However, this is subject to much community-based variation. The standardized orthography of Kaqchikel marks the tense–lax distinction for all five vowels, and the textbooks used at *Oxlajuj Aj*, the language school where the learners in the current study had enrolled, use the standardized orthography (Brown et al., 2006; Maxwell and Little, 2006). Furthermore, though the teachers at Oxlajuj Aj come from various Kaqchikel communities, they all accommodate to the standard in their teaching. Thus, the target language of all students is taken to be Standard Kaqchikel and its 10 vowel inventory, with the tense–lax distinction realized for all five vowel archiphonemes.

The featural specification for Standard Kaqchikel is as follows: There are four distinctive features that specify Kaqchikel's ten vowels: [±high], [±back], [±round], and [RTR]. The [+high] vowels are /i ɪ ʊ u/, while all other vowels /e ɛ a ə ɔ o/ are [−high]. The [+back] vowels are /u ʊ o ɔ/, while all other vowels /i ɪ e ɛ ə a/ are [−back]. The [+round] vowels include all [+back] vowels plus central vowels /a ə/[5], leaving the front, unrounded

5 Evidence for this perhaps unintuitive specification comes from a vowel harmony process in which the vowel of a verbal suffix /-Vʔ/ matches the [±back] specification of the vowel of the verb root /CVC/ but the suffix's vowel is always [+round] and [+tense]: /i ɪ e ɛ a ə/ harmonize to /a/, while /o ɔ/ harmonize to /o/ and /u ʊ/ harmonize to /u/ (Brown et al., 2006, p. 172). An alternative analysis would have the central vowels specified [+back]

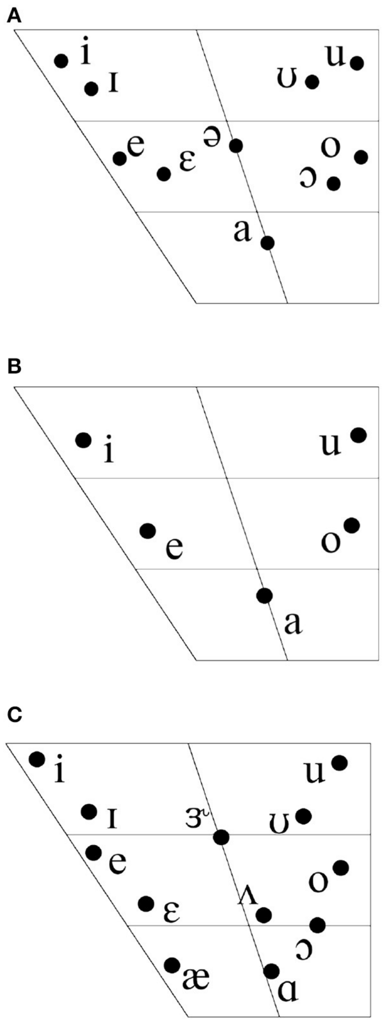


FIGURE 1
Monophthong vowel inventories: **(A)** Standard Kaqchikel vowel monophthong phonemes. **(B)** Guatemalan Spanish vowel monophthong phonemes. **(C)** Typical US American English vowel monophthong phonemes.

vowels /i ɪ e ɛ/ as [−round]. Finally, the [+RTR] vowels are the lax vowels /ɪ ɛ ə ɔ ʊ/, while tense vowels /i e a o u/ are [−RTR].

2.2.1.2. Kaqchikel consonantal phonology

The consonantal inventory of Kaqchikel appears at the top of Table 1. Due to space constraints and the scope of the current study, I discuss only the phonology of this inventory's stops here. Broader consideration of the consonantal inventory is given in Nelson (forthcoming). All stops in Kaqchikel (and also Spanish and English) are [+consonantal], [−sonorant], [−continuant], and [−strident], differing from each other only in Place and Laryngeal specification.

Kaqchikel stops exhibit a four-way place contrast. These four PoAs are, in order from the front of the mouth backwards: bilabial, alveolar, velar, and uvular. The distinction between the first three is usually specified under Feature Theory with a single Place feature for each: Labial for bilabials, Coronal for alveolars, and Dorsal for velars. However, the uvular distinction is less commonly made. Therefore, phonological systems that do contrast uvulars and/or other sounds articulated beyond the velum require more featural complexity to derive their Place contrasts.

Both of Kaqchikel's Labials /p ɓ̥/ are specified for the Place feature [Labial], with no further Labial features/nodes necessary. Similarly, the Coronals /t t'/ are specified for the Place feature [Coronal]. However, Kaqchikel distinguishes between anterior and posterior coronal consonants. Therefore, these two anterior stops are specified [−posterior].

The four remaining Kaqchikel stops are all specified for the Place feature [Dorsal]. However, the velars /k k'/ must be distinguished from the uvulars /q ʛ̥/. Previous analyses of languages with velars and post-velars offer various features and specifications in order to derive these contrasts. Sylak Glassman (2014) provided a detailed history of the featural representation of velars, uvulars, and other post-velars, including [flat] (Jakobson, 1962), [±high] (Chomsky and Halle, 1968), tongue body features (Ladefoged, 1971), [guttural] (Hayward and Hayward, 1989), [pharyngeal] (McCarthy, 1994), and [RTR] (Rose, 1996). This final feature [RTR] is of primary interest to the current study, as it is already utilized for the tense–lax distinction of Kaqchikel vowels. Variably placed within Feature Geometry as a dependent of Pharyngeal, Dorsal, and Tongue Root (TR), I follow Davis (1995) and Shahin (2002) in assuming that post-velars are distinguished from velars via the feature [±RTR], projecting from a TR node, in turn projecting from the Dorsal node held in common between velars and uvulars (76: (45a) vs. (45c)). Acoustic evidence for these specifications of [±RTR, including lowering and backing of front vowels preceding uvulars, is forthcoming in Nelson (forthcoming).

Thus, I assume that Kaqchikel velars contain a Dorsal node, projecting TR, which in turn projects [−RTR] (Figure 2A), while Kaqchikel uvulars contain the same Dorsal and TR nodes, but with [+RTR] as the terminal projection (Figure 2B). In consideration of space, I assume all [±RTR] features are dependent of the articulator node TR, and thus omit the TR node from feature geometry diagrams in the current paper.

Kaqchikel stops[6] show an additional contrast that other sound classes in the language do not: a Laryngeal contrast in which one series is glottalized via a closure/restriction of the glottis in addition to the closure made in the mouth at the stop's PoA. This typically results in an ejective, in which the glottal closure releases and forces the high pressure air upward and outward through the mouth. However, Kaqchikel glottalized labial and uvular stops are often realized as voiceless implosives (Bennett, 2016, p. 485), when, in addition to the oral closure, the vocal folds that form the glottis are closed and move downward without vibrating (Ladefoged and

and [−round] with this process matching the [±round] specification of the root vowel and instead requiring [+back] and [+tense] in the suffix. Given the phonetic variability of /ə/ that allows for fronted [e], raised [ɨ], and rounded [ʌ] (Patal Majzul et al., 2000), I analyze the central vowels as [+round] rather than [+back]. This choice has no direct implications on the current study.

6 Affricates, which fall outside the scope of the current analysis due to their articulatory complexity, also show this contrast.

TABLE 1 Consonant phoneme inventories of Kaqchikel, Spanish, and English.

Standard Kaqchikel	Bilabial	Alveolar	Palatal	Velar	Uvular	Glottal
Plain stop	p	t		k	q	
Glottalized stop	ɓ	t'		k'	ʛ̥	ʔ
Plain affricate		ts	tʃ			
Glottalized affricate		ts'	tʃ'			
Fricative		s	ʃ	x		
Nasal	m	n				
Lateral approximant		l				
Tap		ɾ				
Glide			j	w		

Guatemalan Spanish	Bilabial	Labio-dental	Dental	Alveolar	Post-alveolar	Velar
Voiceless stop	p		t			k
Voiced stop	b		d			g
Voiceless affricate					tʃ	
Voiceless fricative		f		s	ʃ	x
Voiced fricative					ʝ	
Nasal	m			n	ɲ	
Tap				ɾ		
Trill				r		
Lateral approximant				l		

US American English	Bilabial	Labio-dental	Inter-dental	Alveolar	Post-alveolar	Velar	Glottal
Fortis stop	pʰ			tʰ		kʰ	
Lenis stop	b̥			d̥		g̊	
Fortis affricate					tʃʰ		
Lenis affricate					d̥ʒ̊		
Fortis fricative		f	θ	s	ʃ		h
Lenis fricative		v	ð	z	ʒ		
Nasal	m			n		ŋ	
Approximant				ɹ	j	w	
Lateral approximant				l			

Stop consonants are in bold. Kaqchikel stop consonants are shaded based on their Place of Articulation.

Johnson, 2015, p. 165). Regardless of their phonetic realizations, the glottalized stops contrast with plain, non-glottalized stops at each PoA. The plain stops are realized as voiceless, *unaspirated* stops in syllable- and word-initial positions (i.e., onset), but as voiceless, *aspirated* stops in syllable- and word- final positions (i.e., in coda).[7]

The contrast between plain stops and glottalized stops can be uncontroversially derived via specification of [constricted glottis]. The members of the glottalized series /ɓ t' k' ʛ̥/ are specified as having [constricted glottis], while members of the plain series /p t k q/ are specified with an absence of the feature (represented by the null set symbol Ø). The underspecification of a Laryngeal feature for Kaqchikel plain stops allows them to take on the feature [spread glottis] in final positions, thereby surfacing as voiceless, aspirated stops (Nelson, 2023).

7 See Nelson (2023) for a detailed discussion and analysis of aspiration and related word-final processes in Kaqchikel.

2.2.2. Spanish phonology

2.2.2.1. Spanish vocalic phonology

The five vowel inventory of Spanish is shown as Figure 1B. Following Barrios et al. (2016), these five vowels, /i e a o u/, can be specified by just three features, [±back], [±high], and [±low]. The vowels /u o a/ are specified [+back], while the two front vowels are [−back]. The high vowels /i u/ are [+high], leaving the remaining three non-high vowels as [−high]. The only vowel specified [+low] is /a/, while all other vowels are [−low].

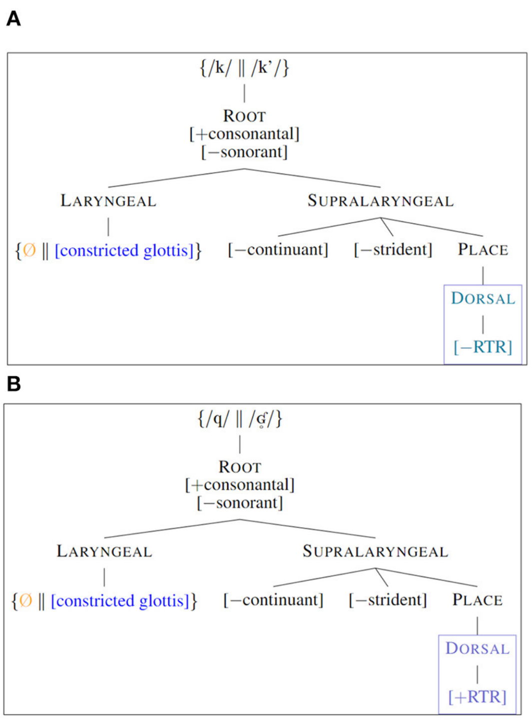


FIGURE 2
Feature geometry for Kaqchikel Dorsal stops: **(A)** velars /k/ and /k'/, **(B)** uvulars /q/ and /ʠ̥/.

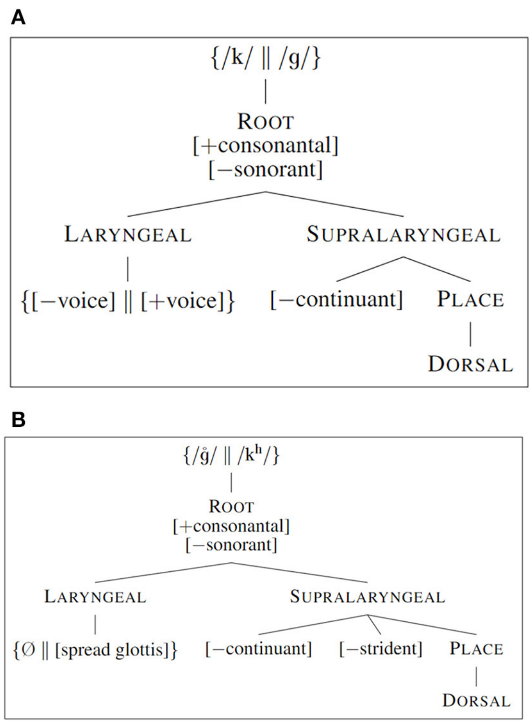


FIGURE 3
Feature geometry for Spanish and English Dorsal stops: **(A)** Spanish velars /k/ and /g/, **(B)** English velars /g̊/ and /kʰ/.

2.2.2.2. Spanish consonantal phonology

The center of Table 1 shows the consonantal inventory of Guatemalan Spanish, which includes some phonemes not typically found in other Spanishes, such as /ʃ/. Our focus remains on the stops however, of which there are six phonemes in Guatemalan Spanish, /p t k b d g/, divided among three PoAs and two Laryngeal series.

I assume Spanish stops to be specified for Place based on a feature geometry with three Place features (Clements and Hume, 1995; Halle et al., 2000; Padgett, 2002), a node for each of Labial, Coronal, and Dorsal place. The bilabials /p b/ have Labial Place. The dentals /t d/ have Coronal Place, and, with both stops being dental while post-alveolar coronals are distinguished in Spanish, the projection of terminal node for [−posterior] is necessary. Finally, the velars /k g/ have Dorsal Place, with no further features projected from the Dorsal node. The feature geometry diagrams for the two Spanish velars are shown as Figure 3A.

I follow Torres-Tamarit (2019) in assuming that, like other Romance languages, the two-way Laryngeal contrast in Spanish is derived by the feature [±voice]. Evidence for this feature lies in the acoustic cues associated with each series. The series /p t k/ exhibits positive, but short VOT in onset position, while the series /b d g/ has negative VOT. Thus, I assume that the first series is a voiceless series specified as [−voice], while the second is a voiced series specified as [+voice].

2.2.3. English phonology

2.2.3.1. English vocalic phonology

Figure 1C shows the comparatively dense vowel inventory of English. Even within a single country, like the United States where all L1 English learners of Kaqchikel who participated in the current study were born and raised, there is considerable variation in the language's vowel inventory. Thus, my description of the inventory is general in that it consists of *around* 12 vowels.

English distinguishes these many vowels on dimensions of height, backness, rounding, and tenseness. To account for these many distinctions, more vocalic features are necessary than for the previous two languages. These features are: [±back], [±high], [±low], [±round], and [±tense]. I take all vowels to be specified for all features. The [+back] vowels are /u ʊ o ɔ ʌ ɑ/. The [+high] vowels are /i ɪ u ʊ/. The [+low] vowels are /æ ɑ/. The [+round] vowels are /ɝ ɔ o ʊ u/. Finally, the [+tense] vowels are /i e o u/ (Kim and Clements, 2015), though, as mentioned previously, Brown and Golston (2006) proposed that one of [ATR] or [RTR] capture this distinction (advanced or retracted in their model).

2.2.3.2. English consonantal phonology

The consonantal inventory of United States American English is shown at the bottom of Table 1. In English, there are six stop phonemes, just as in Spanish, with three PoAs on two Laryngeal series.

I take English to use the same specifications for Place as Spanish: the bilabials /pʰ b̥/ are specified for [Labial], the coronal stops /t^h d̥/ (which in English are alveolar) are specified for Coronal with a terminal node of [−posterior] projecting from it, and the velars /k^h g̊/ are specified for [Dorsal]. The feature geometry for English Dorsal stops is shown as Figure 3B.

Based on Laryngeal Realism (Honeybone, 2001), I take the operative Laryngeal feature in English to be [spread glottis] (Iverson and Salmons, 1995). The fortis (or aspirated) stops /p^h t^h k^h/ are marked in this respect, and thus bear the Laryngeal feature [spread glottis], while the lenis (unaspirated) stops /b̥ d̥ g̊/ are unmarked and do not bear the feature. The lenis series being unmarked for a Laryngeal feature in turn makes them susceptible to intervocalic (passive) voicing in some environments, while in other environments, the [spread glottis] feature is suppressed so that fortis stops may surface as unaspirated (e.g., in sC- onset cluster).

Additionally, in multiple English varieties, word-final stops increasingly surface as ejectives, with at least 26% of such stops being realized as ejective in a corpus analysis by Price et al. (2020). However, MRI analysis in Price et al. (2022) showed that many English ejectives are produced without an elevated larynx, implicating an airstream mechanism distinct from the laryngeal egressive airstream typically ascribed to ejective production in languages like Kaqchikel. Could experience with these emerging allophonic glottalized stops provide learners of Kaqchikel with knowledge transferable to their acquisition of the Kaqchikel laryngeal contrast?

2.3. Learning problem

With these differences in stop phonology among these three languages in hand, we can now formulate the learning problem encountered by learners of Kaqchikel who have already learned and use both Spanish and English. Spanish contrasts voiced stops from voiceless stops, based on the Laryngeal feature [±voice], while English contrasts aspirated stops from unaspirated stops with the Laryngeal feature [spread glottis]. These contrasts, underlying features, and associated cues are neither equivalent to each other, nor equivalent to the Laryngeal contrast of Kaqchikel, a glottalization contrast between glottalized stops and plain stops based instead on [constricted glottis]. How can these learners use their knowledge of Laryngeal contrasts in Spanish and English to learn the new glottalization contrast of Kaqchikel? Adding to this learning problem of Laryngeal specification, Spanish and English both contrast their respective stop consonants at three PoAs: labials, coronals, and velars. Kaqchikel, on the other hand, adds uvulars to these three. Thus, learning the glottalization contrast cannot simply be an extension of the Laryngeal contrasts already known, as new cues must be learned for the uvulars and the contrast between them as well.

2.4. Research questions

The current paper investigates how learners manage to solve the learning problem presented by the new categories and contrasts among Kaqchikel stops by asking the following research questions.

1. Do Spanish-English multilingual learners differ in their perception of Kaqchikel stop consonants based on their L1 (i.e., their L1 Group)?
2. If so, which language is privileged in this regard?
3. Is any such privilege present across the whole stop subsystem, or restricted to individual parts, namely the Laryngeal contrast on one hand and Place contrasts on the other hand(s)?
4. Is there a difference in stop perception based on its Position within its word?

I hypothesize, based on the phonological differences between Spanish and English, that there will be differences between these two groups. These differences should arise in the groups' perception of the Laryngeal contrast of Kaqchikel, but not the place contrasts. English phonology offers a Laryngeal contrast based on [spread glottis], which is a better match than Spanish's [voice] for Kaqchikel's system based on [constricted glottis]. Therefore, based on the L2 Status Factor Model (Bardel and Falk, 2007) and the findings of Wrembel et al. (2019) that PAM in L3 prefers mappings of L3 on L2, the L2 English group (NSS) should outperform the L2 Spanish group (NES) in perception of Kaqchikel's stop contrasts.

On the other hand, if both groups select optimal mappings, as is predicted under redeployment models, no inter-group differences should emerge among the learners, as they should make similar mappings based on their shared knowledge of both Spanish and English. In this case, it may be more beneficial to consider individual cases of L3A of Kaqchikel rather than group-wise comparisons based on learners' L1s.

As there is no difference in Place phonology between English and Spanish stops, the decision to transfer the phonological structure from one language over that of another should not impact their perception. Therefore, I do not predict group differences based on PoA. Instead, the findings of Wagner and Baker-Smemoe (2013) lead me to predict relatively good uvular categorization by both learner groups.

However, I do hypothesize that there is an effect of stop's position that could interact with the potential Group effects, or overall effects of Laryngeal and Place contrasts on categorical perception of Kaqchikel stops. English stops regularly appear in a variety of syllabic positions, including at the beginnings of words and syllables and at the ends of words and syllables. Spanish stops do not regularly appear in domain-final positions. However, Spanish voiceless stops in initial position do map more closely to Kaqchikel plain stops in initial position, in both having unmarked phonological specification and in the acoustic cues associated with them. Therefore, based on property-by-property transfer (Westergaard et al., 2017; Archibald, 2022), knowledge of Spanish domain-initial voiceless stops could be transferred to account for Kaqchikel initial plain stops, while English domain-final stops could be transferred to account for Kaqchikel final stops.

3. Materials and methods

3.1. Participants

In order to investigate the stated research questions, I recruited 18 multilingual Kaqchikel–Spanish–English users to perform various tasks in those three languages. The first group were

TABLE 2 Descriptive statistics of each group's ages (in years) at time of study plus their ages of learning (AoL, also in years) and listening and speaking self-assessment scores (out of 3) for each language. Group means (with standard deviations in parentheses).

Group	Age	Kaqchikel			Spanish			English		
		AoL	*Listening*	*Speaking*	*AoL*	*Listening*	*Speaking*	*AoL*	*Listening*	*Speaking*
NKS	38.4	(0.0)	2.8	2.6	5.2	2.8	2.6	24.0	1.4	1.4
n = 5	(15.5)	(0.0)	(0.4)	(0.5)	(4.8)	(0.4)	(0.4)	(13.3)	(0.5)	(0.5)
NSS	26.0	22.7	1.3	1.3	(0.0)	3.0	3.0	4.2	3.0	2.3
n = 6	(4.0)	(3.1)	(0.5)	(0.5)	(0.0)	(0.0)	(0.0)	(3.4)	(0.0)	(0.5)
NES	26.7	24.3	1.3	1.3	8.0	2.3	2.3	(0.0)	3.0	3.0
n = 6	(6.0)	(4.4)	(0.5)	(0.5)	(5.5)	(0.5)	(0.5)	(0.0)	(0.0)	(0.0)

L1 users of Kaqchikel (heretofore labeled as "Native Kaqchikel Speakers" or NKS), while the two other groups were L3 learners of Kaqchikel. The learner groups were divided according to the L1 of each participant, with 6 learners being L1 Spanish (receiving the identifier "Native Spanish Speakers", NSS) and 7 learners being L1 English ("Native English Speakers", NES), in a mirror-image design (Ortin and Fernandez-Florez, 2019).

One NES learner's (NES7) results were excluded due to their insufficient Spanish experience relative to other NES participants. This left 5 NKS, 6 NSS, and 6 NES. Each participant also gave a self-assessment of their Kaqchikel, Spanish, and English skill levels in listening and speaking using a three-point rating scale (1 ="beginner", 2 ="intermediate", and 3 ="fluent"). The descriptive statistics for the groups' ages and their age of learning and self-assessments for each language appear in Table 2.

3.2. Categorization task

3.2.1. Materials

3.2.1.1. Stimuli

Stimuli for the categorization task were audio recordings of 160 Kaqchikel, Spanish, and English words as spoken by a female L1 Kaqchikel L2 Spanish, L3 English speaker. This speaker was born and raised in Guatemala, but had resided elsewhere at various points in her life, including the United States and Spain. As a Kaqchikel language researcher and teacher, she was able to produce the stimuli in a Standard Kaqchikel variety.

She recorded the stimuli into a Zoom H4N Handy Recorder equipped with a Sony ECM-44B condenser microphone attached to the her lapel, by reading aloud from wordlists with each target word placed within a carrier sentence that matched the language of the word. For Kaqchikel, this carrier sentence was *Xinb'ij [word] la q'ij la'* "I said [word] that day)". For Spanish it was *Dijo [word] la semana pasada* "I said [word] last week". For English words the carrier sentence was *I said [word] last week*. She also read from an additional Kaqchikel wordlist, recordings of which formed part of the category labels for the categorization task. She read each wordlist two times, with half of stimuli selected for presentation in the task coming from the first readings and half coming from the second readings. Recordings were trimmed down to just the target word using speech analysis and modification software, Praat.

In order to test for any differences in categorization based on the stops' positions within their respective words, the stimuli featured each of the three languages' six or eight stops in both initial/onset and final/coda positions. Additionally, each stop phone for all three languages appeared before each of the four vowels /i e o u/ in the initial condition and after each of the four vowels /i e o u/ in the final condition. Each stimulus word was presented twice giving a total of 320 trials for each participant. Of the 320 trials, only 128 presented Kaqchikel words; the remaining 192 presented Spanish or English words. For the purpose of the current study's analysis, the 192 Spanish and English trials were distractors that kept the listener in a trilingual mode of perception throughout testing. They would be cross-linguistically categorizing Spanish and English stops into Kaqchikel categories, but cis-linguistically categorizing Kaqchikel stops into Kaqchikel categories. The use of stimuli from three languages is based on the bilingual categorization task of Wagner and Baker-Smemoe (2013), in which English and Q'eqchi' stops were categorized into English categories. The list of Kaqchikel stimuli words presented to listeners is shown in Tables 3, 4.

The categories that participants would be asked to sort each word into were the eight phonemic oral stops of Kaqchikel. These categories were represented visually by an icon that corresponded to a word which began with the stop phoneme it was serving as a label for. I give these eight category label words in Table 5.

In addition to categorizing words based on their initial and final stop consonants, participants also rated the goodness of each categorization. This rating was made on a scale with 5 discrete points. Ratings could be indicated by clicking a point on the scale or by pressing the corresponding number on the keyboard.

3.2.1.2. Equipment

The experiment was run on the same Dell Inspiron 13 7370 laptop computer for all participants. Participants listened to the stimuli at maximum volume through a pair of Sony MDRZX110 over-the-ear headphones wired to the laptop. The category and rating interface appeared on the laptop's display, allowing participants to indicate their responses using the laptop's track-pad or a wireless mouse connected to the laptop via Universal Serial Bus (USB) adapter. All equipment was wiped with rubbing alcohol before each participant used it.

TABLE 3 Categorization task stimulus words: Kaqchikel, initial stops only.

Stop position	*Stop laryngeal*	*Stop place*	*V*	*Orthographic word*	*Phonemic word*	*Word meaning*
Initial	Plain	Labial	i	*pich'*	/pitʃ'/	Tender corn
			e	*pe*	/pe/	Come
			o	*poy*	/poj/	Scarecrow
			u	*pur*	/pur/	Snail
		Coronal	i	*tix*	/tiʃ/	Elephant
			e	*tem*	/tem/	Column
			o	*tol*	/tol/	Gourd container
			u	*tum*	/tum/	Drum
		Velar	i	*kis*	/kis/	Fart
			e	*kem*	/kem/	Weaving
			o	*kow*	/kow/	Hard
			u	*kux*	/kuʃ/	Weasel
		Uvular	i	*qi'*	/qiʔ/	Ourselves
			e	*qey*	/qej/	Our teeth
			o	*qo'ch*	/qoʔtʃ/	Crow; raven
			u	*qupe*	/qupe/	Right?
	Glottalized	Labial	i	*b'is*	/ɓ̥is/	Sadness
			e	*b'ey*	/ɓ̥ej/	Road
			o	*b'o'j*	/ɓ̥oʔx/	Cotton
			u	*b'usaj*	/ɓ̥usax/	Sheet of paper; page
		Coronal	i	*t'im*	/t'im/	Plastic
			e	*t'esël*	/t'esɛl/	Very fat
			o	*t'ok*	/t'ok/	Fist
			u	*t'uq*	/t'uq/	Setting hen
		Velar	i	*k'im*	/k'im/	Straw
			e	*k'el*	/k'el/	Parakeet
			o	*k'oj*	/k'ox/	Tamale-wrapping-leaf plant
			u	*k'ul*	/k'ul/	Blanket
		Uvular	i	*q'ij*	/ɠ̥ix/	Sun; day
			e	*q'e'l*	/ɠ̥eʔl/	Pitcher
			o	*q'or*	/ɠ̥or/	Atole
			u	*q'ux*	/ɠ̥uʃ/	Moss

3.2.2. Methods

The trilingual categorization and rating task was designed and blocked out using the PsychoPy software, which generated and ran Python code that presented stimuli to participants and recorded their responses in spreadsheet format. All participants completed the task during the summer of 2019, prior to the COVID-19 pandemic and any potential adjustments to methods that it would have forced.

Prior to any practice or testing trials, each participant was given unlimited time to familiarize themself with the experimental interface and category labels via a PowerPoint slide matching the interface they would see in test trials. This slide included clickable icons, each of which played an audio recording of the word associated with that icon. Each icon was a monochrome depiction of the meaning of its associated word. This training was necessary as the audio for each category label would not be playable during categorization trials. Therefore, participants would have to associate the initial stop of each category label word with the icons on the screen during this time.

The experiment was divided into two blocks, with 160 trials testing initial stops randomized within the first block, and 160 trials testing final stops randomized within the second block. Prior to

TABLE 4 Categorization task stimulus words: Kaqchikel, final stops only.

Stop position	Stop laryngeal	Stop place	V	Orthographic word	Phonemic word	Word meaning
Final	Plain	Labial	i	*sip*	/sip/	Tick
			e	*xq'ep*	/ʃɢ̥ep/	It was snapped
			o	*ch'op*	/tʃ'op/	Pineapple
			u	*ch'up*	/tʃ'up/	Passionfruit
		Coronal	i	*xit*	/ʃit/	Jade
			e	*xet*	/ʃet/	Long hair bound atop the head
			o	*xot*	/ʃot/	Griddle
			u	*sut*	/sut/	Hair whorl
		Velar	i	*jik*	/xik/	Straight
			e	*ach'ek*	/atʃ'ek/	Dream
			o	*t'ok*	/t'ok/	Fist
			u	*kuk*	/kuk/	Squirrel
		Uvular	i	*xb'iq*	/ʃɓ̥iq/	It was degrained
			e	*weq*	/weq/	Floor; story; level
			o	*t'oq*	/t'oq/	Thick
			u	*t'uq*	/t'uq/	Setting hen
	Glottalized	Labial	i	*sib'*	/siɓ̥/	Smoke
			e	*jaleb'*	/xaleɓ̥/	Disguise
			o	*yob'*	/joɓ̥/	Dimple
			u	*xub'*	/ʃuɓ̥/	Whistle
		Coronal	i	*xit'*	/ʃit'/	It was filled well
			e	*let'et'*	/let'et'/	Bicycle
			o	*yot'*	/jot'/	Dimple
			u	*rut'*	/rut'/	Receipt
		Velar	i	*sik'*	/sik'/	Cigar
			e	*k'ek'*	/k'ek'/	Miserly
			o	*k'ok'*	/k'ok'/	Spicy-smelling
			u	*ruk'*	/ruk'/	Her louse
		Uvular	i	*liq'*	/liɢ̥/	Slimy
			e	*xmeq'*	/ʃmeɢ̥/	It was warmed
			o	*moq'*	/moɢ̥/	Handful
			u	*tuq'*	/tuqɢ̥/	Purple

each testing block, a screen displayed instructions for the task in both English and Spanish. Once participants read this screen and indicated they were ready, practice trials began. After completing four practice trials, which included a word from each of the three languages being studied, another screen reminding participants of their instructions was shown. Again, once participants read this screen and indicated they were ready, testing trials began.

For each trial an audio file of a word in either Kaqchikel, Spanish, or English played a single time. The categorization interface displayed eight black icons corresponding to the Kaqchikel stop category labels learned from the PowerPoint slide prior to the task. Once the participant clicked any of the icons, the screen would switch to the rating interface, where the participant would indicate how good of a match that word's stop was to the category they selected. There was no reminder of which categorization the participant made. Once the rating was confirmed, the next trial began. This repeated for all 160 test trials in the block. Between blocks there was a break. Once the participant indicated they were ready to continue, they saw a screen showing the instructions for the "Final" block, and the process was repeated focusing on the final stop of stimuli words. After they completed all trials, a final screen thanked the participant

TABLE 5 Categorization task category label word associations.

Category Laryngeal	Category place	Category phoneme	Orthographic word	Phonemic word	Word meaning
Plain	Labial	/p/	*patx*	/patʃ/	Duck
	Coronal	/t/	*tukr*	/tukr/	Owl
	Velar	/k/	*kumätz*	/kumǝts/	Snake
	Uvular	/q/	*qo'l*	/qoʔl/	Turkey
Glottalized	Labial	/ɓ̥/	*b'alam*	/ɓ̥alam/	Jaguar
	Coronal	/t'/	*t'ot'*	/t'ot'/	Conch shell
	Velar	/k'/	*k'aj*	/k'ax/	Droplet
	Uvular	/ʛ̥/	*q'aq'*	/ʛ̥aʛ̥/	Fire

for their time. BN then debriefed the participant compensated for their time.

3.2.3. Measures

For each trial, every participant's response was recorded using the PsychoPy software. This included the Kaqchikel stop category selected as the best match for the played word's stop consonant (i.e., the categorization), the amount of time elapsed between the playing of the audio and the participant's categorization (Response Time, RT), the participant's rating of their categorization, and the amount of time elapsed between the initiation of the rating interface and the selection of the rating.

For the current analysis, rates of correct categorizations can also be calculated. These rates (or accuracy scores) could then be used to assess the accuracy of a participant's or group of participants' categorical perception of a given division of Kaqchikel's stop consonants: by Place feature, by Laryngeal feature, by word Position, or any combination of any of these three factors. Alternatively, for a logistic regression analysis, as is to be performed here, each response is coded as either *correct* (with a binary value of 1) or *incorrect* (0).

The category with the most selections for a given division of the sample (i.e., a listener, a language group, or the sample as a whole) for a given stimulus specification (i.e., a Kaqchikel stop phoneme with or without regard to its position in its word) was determined to be the modal categorization for that division for that stimulus specification. In the event that multiple categorizations are tied as a modal categorization, the categorization with the highest mean rating was determined to be the modal categorization. Modal categorizations can be used to compare different participants' or participant groups' performance on this task to each other.

When the categorization is done cross-linguistically, ratings show what types of assimilation or categorization scenario are occurring. Higher ratings indicate high category goodness, while lower ratings indicate low category goodness. Thus, even in the case where a learner selects the same category phoneme for multiple stimulus phones, relative goodness among their categorizations can be determined. Cross-language rating tasks of this type have been used in previous multilingual studies to show how multilinguals perceive sounds from both their known and target languages at different levels of similarity to sounds in one of those languages [e.g., a previously learned language in Wagner and Baker-Smemoe (2013) or the target language in Wrembel et al. (2019)].

The focus of the current analysis lies only on the accuracy of categorizations of Kaqchikel stops into Kaqchikel categories. I leave complete analysis of Categorizations and Response Times to future work [see Nelson (forthcoming)].

3.3. Other tasks

In addition to the categorization task, participants were asked to complete other tasks of perception and production. The first tasks all participants completed were production tasks. These tasks consisted of the reading of wordlists in all three languages of the study, Kaqchikel, Spanish, and English. The procedure for recording these wordlists matched the procedure for recording of the perception stimuli.

All participants also completed a language background questionnaire, responses of which informed the description of the participants in §3.1. Each participant completed this written questionnaire during a break in their AX Discrimination task. For all participants, the Categorization task described previously in this section was the final task completed as it was most revealing of the objective of the study as a whole: documenting perception and production of stop consonants in Kaqchikel, Spanish, and English.

4. Results

4.1. Raw accuracy results

Listeners were accurate in their categorization of Kaqchikel stimuli stops into Kaqchikel stop phoneme categories on 54.7% [$\frac{1190}{2176}$, standard deviation (*SD*) = 29.8%] of all trials. For the NKS group, the overall accuracy rate was 68.1% ($\frac{436}{640}$, *SD* = 20.6%). The learner groups had lower accuracy rates: NSS were correct on 40.2% ($\frac{309}{768}$, *SD* = 18.0%) of trials, while NES were correct on 57.9% ($\frac{445}{768}$, *SD* = 13.6%) of their trials. Note that performance at the level of chance is 12.5% accuracy, not 50%, as there are 8 potential categories, despite there only being two outcomes (correct or incorrect).[8] Table 6 shows the accuracy of categorization for each group (as well as the standard deviations within each group) according to the levels of each of the factors Place, Laryngeal, and Position.

When trials are divided by the Place of the stop being categorized, the Place with the highest accuracy across all listeners was coronal, followed by labial, velar, and then uvular. As for how each group performed relative to each Place, NKS were most accurate with uvulars, followed closely by coronals, velars, and then labials, while NSS were most accurate on labials, followed by coronals, velars, and then uvulars. NES were most accurate in categorizing coronals, followed very closely by the other three Places: velars and uvulars, and then labials.

Dividing trials by the Laryngeal feature of the stimulus stop, across all participants, plain stops were categorized correctly on 51.4% of trials ($\frac{559}{1088}$, $SD = 22.9\%$). This was slightly lower than glottalized stops, which had an accuracy of 58.0% ($SD = 20.2\%$). This pattern held for each of the three language Groups as well.

When Laryngeal and Place features are considered together, the phoneme whose stimuli were most accurately categorized correctly across all listeners was the glottalized uvular /ʠ̥/ (65.4%, $\frac{178}{272}$, $SD = 23.8\%$). This was followed in order of decreasing accuracy by the plain coronal /t/, the plain labial /p/, the glottalized velar /k'/, the glottalized labial /ɓ̥/, the glottalized coronal /t'/, the plain velar /k/, and finally the plain uvular /q/ (34.9%, $SD = 33.0\%$).

Based on the Position of the stimulus stop, initial stops, with an accuracy rate of 58.0% ($\frac{631}{1088}$, $SD = 20.0\%$), were categorized correctly more often than final stops (51.4%, $SD = 21.4\%$) across all listeners. This was true within each L1 Group as well.

Pulling all factors together and considering a stop's Place, Laryngeal, and Position in determining its accuracy of categorization (as in Table 6, initial stops were categorized with greater accuracy than final stops for every stop phoneme (i.e., Place + Laryngeal combination), except the glottalized uvular /ʠ̥/, which had greater accuracy in final Position than when in initial Position. NKS deviated from this overall pattern in categorizing final glottalized coronal /t'/ with greater accuracy than initial /t'/. NSS deviated from the overall pattern in categorizing three of the four glottalized stop phonemes with greater accuracy in final Position than in initial Position, adding /t'/ and /k'/ to /ʠ̥/ in that regard. Finally, NES also deviate from the overall pattern by having greater accuracy in categorizing final stops than initial stops for /ɓ̥/ and /q/, in addition to /ʠ̥/.

4.2. Multiple logistic regression analysis

4.2.1. Logistic mixed effects model

I analyzed the participant's ability to correctly categorize Kaqchikel stops into Kaqchikel categories under a logistic mixed effects model[9]. I fitted a logistic mixed model [estimated using the Bound Optimization B Quadratic Approximation (BOBYQA) optimizer on the Akaike Information Criterion (AIC)] in order to predict Correct categorizations based on Group, Place, Laryngeal, and Position factors. Included in this model were three interaction factors: one between listener Group and stimulus Place feature, one between Group and stimulus Laryngeal feature, and one between Place and Laryngeal features. I also included a single random effect caused by individual difference among Listeners.

The formula for the optimized model in the `lme4` package's `glmer()` function was:

```
Correct ~ (1|Listener) + Group + Place +
Laryngeal + Position +
Group*Place + Group*Laryngeal +
Place*Laryngeal.
```

As indicated by its conditional R^2 of 0.33, the model's total explanatory power is substantial. The fixed effects part of the model carries a marginal R^2 of 0.18. The model's results are listed in Table 7. Note that the contrasts within factors were made using sum contrast coding, with estimates representing deviations from the mean. For convenience, I list the effect of each level of every factor. In the rightmost column, I also give the observed accuracy rate for the portion of the sample corresponding to each effect. These can be compared to the model's predicted probabilities, which I give in the following paragraphs.

The logistic regression analysis of this model reveals several significant effects among each of the factors as well as two of the three interaction terms. This indicates that the factors and those two interactions are meaningful predictors of correct categorization of Kaqchikel stops. Figure 4 shows the predicted probability of a Kaqchikel stop being correctly categorized based on the L1 of the listener, its Place features, its Laryngeal features, and its Position at the beginning or end of its word.

4.2.2. Main effects

4.2.2.1. Group

Of the three listener Groups in the current study, the model reveals a significant effect of two of them. Unsurprisingly, NKS listeners are significantly more likely than average to provide a correct categorization of a Kaqchikel stop ($\beta = 0.77$, 95% CI: $[0.15, 1.40]$; $z = 2.42, p = 0.015$). On the other hand, listeners of the NSS Group are significantly less likely to provide a correct categorization ($\beta = -0.85, [-1.45, -0.25]$; $z = -2.79, p = 0.005$). The effect of a listener belonging to NES was found to be non-significant ($\beta = 0.07, [-0.52, 0.66]$; $z = 0.25, p = 0.809$).

Figure 5A shows the predicted probability of a correct categorization[10] for each Group. NKS has a predicted probability of 74% correct, NSS has 36%, and NES has 59%.

8 Accuracy of classification could be analyzed at a featural level, with one accuracy score for choosing a category that shares its Place with the correct phoneme and one score for choosing a category that shares its Laryngeal feature with the correct phoneme. I do not make that level of analysis here.

9 I had initially analyzed these data under mixed Analysis of Variance design. At the urging of a reviewer, I analyzed the data using this model design.

10 $Prob = \frac{e^{\beta}}{1+e^{\beta}}$, where $Prob$ is the predicted probability, e is Euler's number, and β is a coefficient of the model.

TABLE 6 Categorization accuracy by groups based on stop Laryngeal feature, position, and place of articulation.

	Plain							
	Initial (C-)				Final (-C)			
	Labial /p/		Coronal /t/		Labial /p/		Coronal /t/	
Group	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*
NKS	70.0%	40	85.0%	40	40.0%	40	72.5%	40
$n = 5$	$SD = 28.8\%$		$SD = 20.5\%$		$SD = 38.9\%$		$SD = 35.8\%$	
NSS	62.5%	48	52.1%	48	50.0%	48	43.8%	48
$n = 6$	$SD = 22.4\%$		$SD = 34.8\%$		$SD = 32.6\%$		$SD = 31.4\%$	
NES	64.6%	48	70.1%	48	66.7%	48	56.3%	48
$n = 6$	$SD = 33.9\%$		$SD = 33.2\%$		$SD = 47.2\%$		$SD = 42.4\%$	
	Velar /k/		Uvular /q/		Velar /k/		Uvular /q/	
	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*
NKS	67.5%	40	62.5%	40	57.5%	40	55.0%	40
$n = 5$	$SD = 30.1\%$		$SD = 28.0\%$		$SD = 37.1\%$		$SD = 36.0\%$	
NSS	39.6%	48	10.4%	48	29.2%	48	2.1%	48
$n = 6$	$SD = 24.2\%$		$SD = 20.0\%$		$SD = 25.8\%$		$SD = 5.1\%$	
NES	64.6%	48	39.6%	48	39.6%	48	47.9%	48
$n = 6$	$SD = 30.0\%$		$SD = 25.5\%$		$SD = 41.4\%$		$SD = 42.1\%$	
	Glottalized							
	Initial (C-)				Final (-C)			
	Labial /ɓ̥/		Coronal /t’/		Labial /ɓ̥/		Coronal /t’/	
	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*
NKS	75.0%	40	62.5%	40	65.0%	40	70.0%	40
$n = 5$	$SD = 15.3\%$		$SD = 38.5\%$		$SD = 32.3\%$		$SD = 44.7\%$	
NSS	58.3%	48	41.7%	48	41.7%	48	50.0%	48
$n = 6$	$SD = 41.6\%$		$SD = 38.5\%$		$SD = 30.3\%$		$SD = 34.5\%$	
NES	60.4%	48	62.5%	48	37.5%	48	45.8%	48
$n = 6$	$SD = 24.3\%$		$SD = 37.9\%$		$SD = 35.4\%$		$SD = 37.6\%$	
	Velar /k’/		Uvular /ʠ̥/		Velar /k’/		Uvular /ʠ̥/	
	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*	Accuracy	*n*
NKS	67.5%	40	82.5%	40	62.5%	40	95.0%	40
$n = 5$	$SD = 31.4\%$		$SD = 16.8\%$		$SD = 39.5\%$		$SD = 6.8\%$	
NSS	37.5%	48	35.4%	48	45.8%	48	43.8%	48
$n = 6$	$SD = 22.4\%$		$SD = 12.3\%$		$SD = 24.6\%$		$SD = 32.4\%$	
NES	72.9%	48	64.6%	48	54.2%	48	79.2%	48
$n = 6$	$SD = 31.0\%$		$SD = 18.4\%$		$SD = 31.3\%$		$SD = 28.1\%$	

4.2.2.2. Place

Two of the four Places are significant predictors of categorization accuracy. Coronal stops are significantly more likely to be categorized correctly by Kaqchikel-Spanish-English multilingual listeners ($\beta = 0.21, 95\%$ CI: $[0.04, 0.37]; z = 2.46, p = 0.014$), while uvular stops are significantly less likely to be accurately categorized ($\beta = -0.20, [-0.38, -0.02]; z = -2.14, p = 0.032$). The effect of labial Place is non-significant but positive ($\beta = 0.10, [-0.06, 0.26]; z = 1.18, p = 0.237$), while that of velar Place is non-significant but negative ($\beta = -0.11, [-0.27, 0.05]; z = -1.31, p = 0.189$).

The predicted probabilities for a correct categorization of a stop from each of the four Kaqchikel Places are shown in Figure 5B.

TABLE 7 Mixed effects logistic regression analysis of the factors' contribution to the probability of accurately categorizing a Kaqchikel stop. The rightmost columns give the number of observed that contributed to the model plus the accuracy rate of those observed trials.

AIC= 2580.6, Marginal R^2 = .18, Conditional R^2 = .33							Observations	
Fixed effects	*Estimate [95% CI]*		*SE*	*z*	*p*	*Sig.*	*n*	*Accuracy %*
INTERCEPT	0.29	[-0.14, 0.72]	0.22	1.33			2, 176	54.7
Group								
NKS	0.77	[0.15, 1.40]	0.32	2.42	0.015	*	640	68.1
NSS	-0.85	[-1.45, -0.25]	0.30	-2.79	0.005	**	768	40.2
NES	0.07	[-0.52, 0.66]	0.30	0.25	0.809		768	57.9
Place								
Labial	0.10	[-0.06, 0.26]	0.08	1.18	0.237		544	57.4
Coronal	0.21	[0.04, 0.37]	0.08	2.46	0.014	*	544	58.6
Velar	-0.11	[-0.27, 0.05]	0.08	-1.31	0.189		544	52.6
Uvular	-0.20	[-0.38, -0.02]	0.09	-2.14	0.032	*	544	50.2
Laryngeal								
Plain	-0.21	[-0.31, -0.11]	0.05	-4.16	<0.001	***	1, 088	51.4
Glottalized	0.21	[0.11, 0.31]	0.05	4.16	<0.001	***	1, 088	58.0
Position								
Initial	0.17	[0.07, 0.26]	0.05	3.47	<0.001	***	1, 088	58.0
Final	-0.17	[-0.26, -0.07]	0.05	-3.47	<0.001	***	1, 088	51.4
Group × Place								
NKS × Labial	-0.45	[-0.69, -0.21]	0.12	-3.62	<0.001	***	160	62.5
NKS × Coronal	-0.01	[-0.26, 0.24]	0.13	-0.09	0.927		160	72.5
NKS × Velar	-0.17	[-0.42, 0.07]	0.12	-1.37	0.170		160	63.7
NKS × Uvular	0.63	[0.36, 0.91]	0.14	4.51	<0.001	***	160	73.8
NSS × Labial	0.59	[0.36, 0.81]	0.12	5.04	<0.001	***	192	53.1
NSS × Coronal	0.19	[-0.04, 0.41]	0.12	1.59	0.113		192	46.9
NSS × Velar	0.08	[-0.15, 0.31]	0.12	0.68	0.494		192	38.0
NSS × Uvular	-0.85	[-1.12, -0.59]	0.14	-6.26	<0.001	***	192	22.9
NES × Labial	-0.14	[-0.36, 0.08]	0.11	-1.22	0.224		192	57.3
NES × Coronal	-0.17	[-0.40, 0.05]	0.11	-1.52	0.129		192	58.9
NES × Velar	0.09	[-0.13, 0.31]	0.11	0.80	0.421		192	57.8
NES × Uvular	0.22	[-0.02, 0.46]	0.12	1.81	0.071		192	57.8
Group × Laryngeal								
NKS × Plain	-0.07	[-0.21, 0.08]	0.07	-0.88	0.380		320	63.8
NKS × Glottalized	0.07	[-0.08, 0.21]	0.07	0.88	0.380		320	72.5
NSS × Plain	-0.06	[-0.20, 0.07]	0.07	-0.89	0.376		384	36.2
NSS × Glottalized	0.06	[-0.07, 0.20]	0.07	0.89	0.376		384	44.3
NES × Plain	0.13	[-0.00, 0.26]	0.07	1.89	0.058		384	56.3
NES × Glottalized	-0.13	[-0.26, 0.00]	0.07	-1.89	0.058		384	59.7
Place × Laryngeal								
Labial × Plain	0.29	[0.13, 0.45]	0.08	3.49	<0.001	***	272	59.2
Labial × Glottalized	-0.29	[-0.45, -0.14]	0.08	-3.49	<0.001	***	272	55.5

(Continued)

TABLE 7 (Continued)

AIC= 2580.6, Marginal R^2 = .18, Conditional R^2 = .33							Observations	
Fixed effects	*Estimate [95% CI]*		*SE*	*z*	*p*	*Sig.*	*n*	*Accuracy %*
Coronal × Plain	0.39	[0.22, 0.55]	0.08	4.64	<0.001	***	272	62.5
Coronal × Glottalized	-0.39	[-0.55, -0.22]	0.08	-4.64	<0.001	***	272	54.8
Velar × Plain	0.02	[-0.14, 0.19]	0.08	0.28	0.782		272	48.9
Velar × Glottalized	-0.02	[-0.19, 0.14]	0.08	-0.28	0.782		272	56.2
Uvular × Plain	-0.70	[-0.88, -0.51]	0.08	-7.44	<0.001	***	272	34.9
Uvular × Glottalized	0.70	[0.51, 0.88]	0.08	7.44	<0.001	***	272	65.4
Random Effects							*SD*	*n*
LISTENER							0.20	17

Significance levels: * : $p < 0.050$; ** : $p < 0.010$; *** : $p < 0.001$.
Color of ∗ refers to color of effect in predicted probability plots.
Sum contrast coding. Estimate coefficients reported in log-odds.
Model significantly better than null model (AIC= 2702.0, $\chi^2 = 157.37$, $DF = 18$).

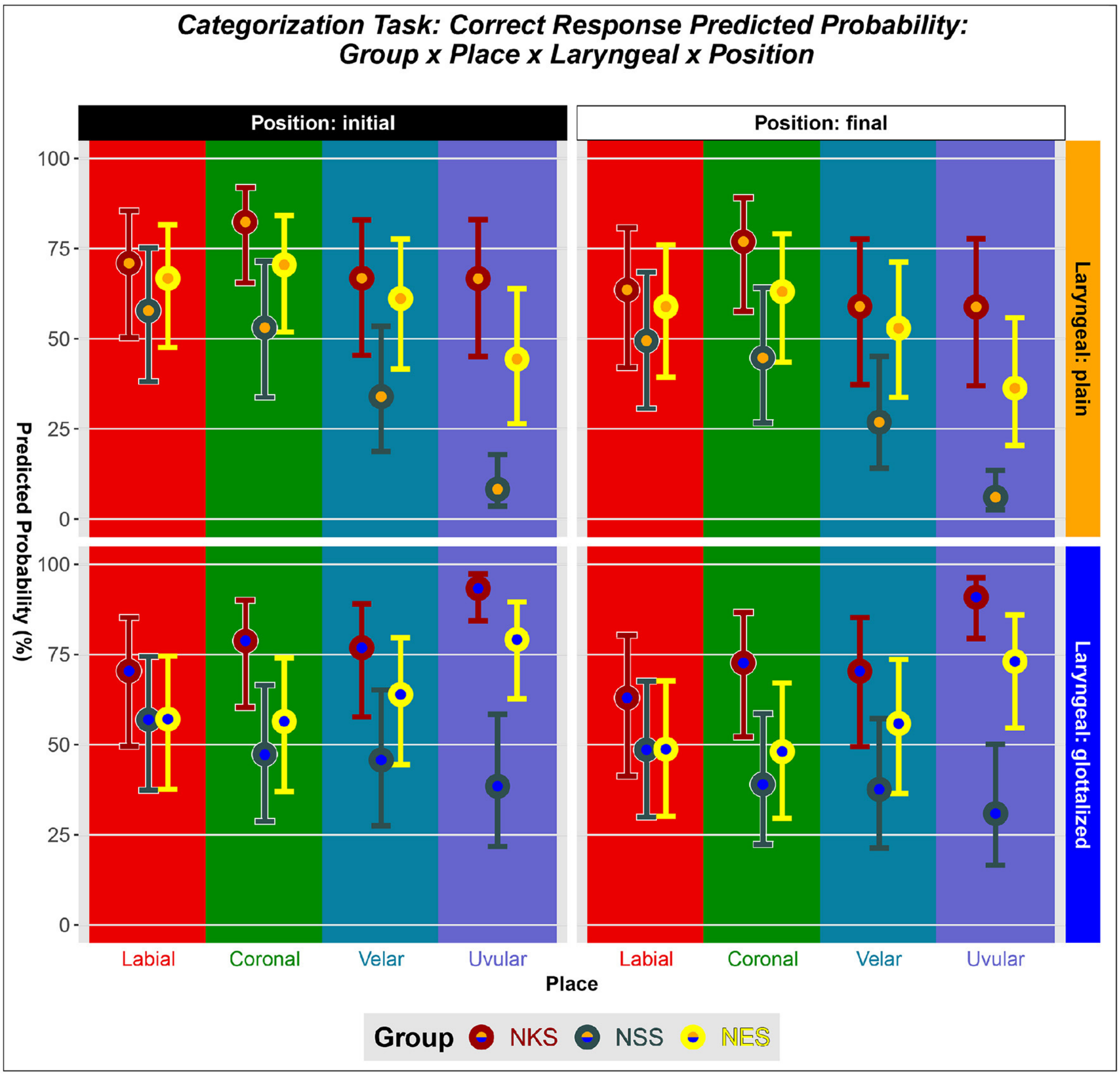


FIGURE 4
Predicted probability plot showing the model's predicted probability for each combination of the levels of Group, Place, Laryngeal, and Position.

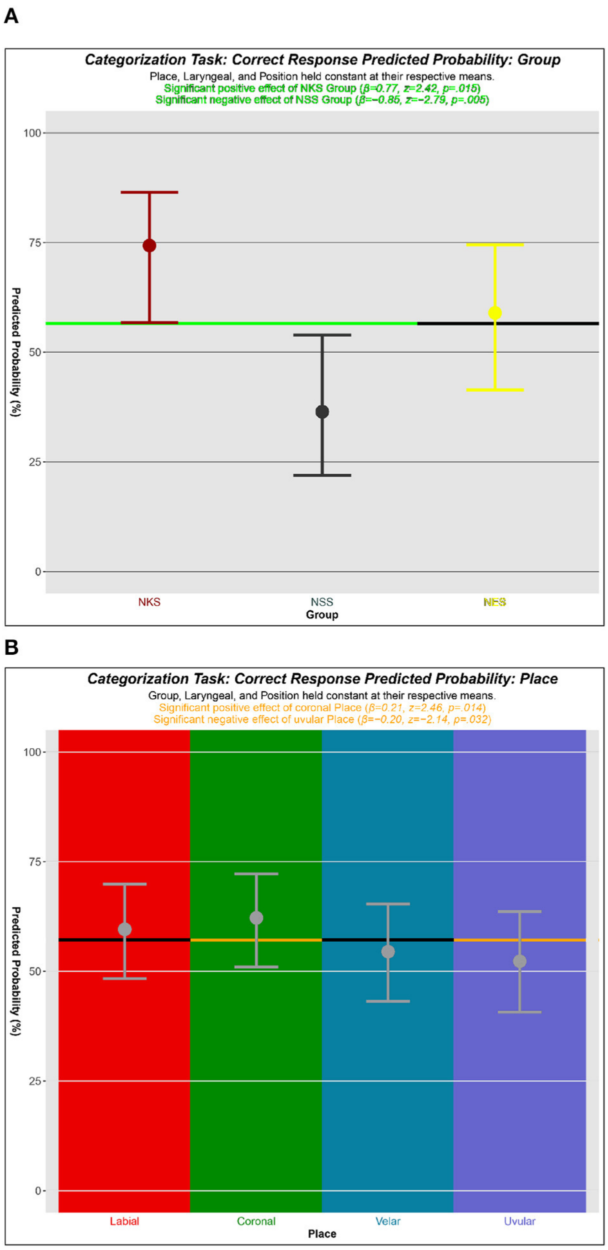


FIGURE 5
Probability plots for the Group and Place factors based on the mixed effects logistic regression analysis. The horizontal line in each represents the mean to which each level of Group or Place was compared. Blacked out sections of these lines indicate that there was not a significant effect of that section's Group or Place: **(A)** Group, **(B)** Place.

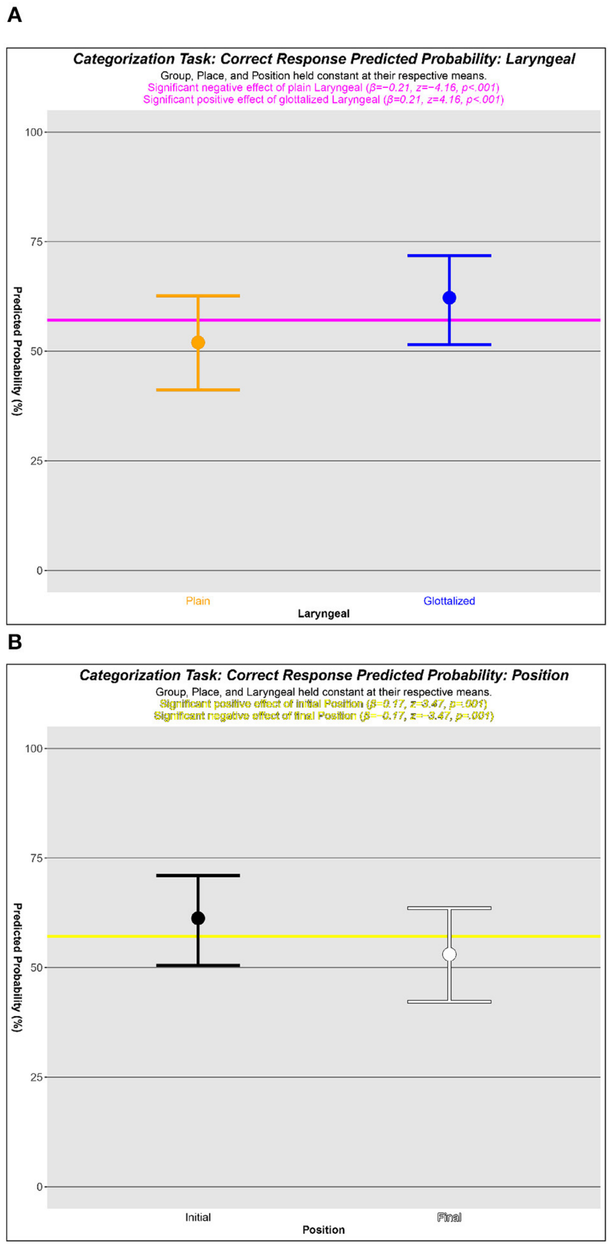


FIGURE 6
Probability plots for the Laryngeal and Position factors based on the mixed effects logistic regression analysis. The horizontal line in each represents the mean predicted probability to which each level of Laryngeal or Position was compared: **(A)** Laryngeal, **(B)** Position.

Listeners are predicted to correctly categorize labial stops 60% of the time, while the predicted probability for coronals is 62%, velars 54%, and uvulars 52%.

4.2.2.3. Laryngeal

As there are two series of stops based on a Laryngeal feature in Kaqchikel, a positive effect of a stop being in one series would be accompanied by an equivalent negative effect of a stop belong to the other series. In this case, the plain stops of Kaqchikel are significantly less likely to be correctly categorized ($\beta = -0.21$, 95% CI: $[-0.31, -0.11]$; $z = -4.16, p < 0.001$), while glottalized stops are significantly more likely to be correctly categorized ($\beta = 0.21$, 95% CI: $[0.11, 0.31]$; $z = 4.16, p < 0.001$).

Figure 6A shows the predicted probabilities of a listener providing a correct response for a stop in each of the two Laryngeal series of Kaqchikel. Listeners are predicted to be correct on 52% of plain stop trials, while for glottalized stop trials the predicted probability is 62%.

4.2.2.4. Position

Similar to the Laryngeal factor, the Position factor was measured at two levels, word-initial and word-final. As such, a positive effect of one Position is accompanied by an equivalent negative effect of the other. In the current model, initial Position has a significant positive effect on accurate stop categorization ($\beta = 0.17$, 95% CI: $[0.07, 0.26]$; $z = 3.47, p < 0.001$), so final Position has a significant negative effect ($\beta = 0.17$, 95% CI: $[-0.26, -0.07]$; $z = -3.47, p < 0.001$).

Figure 6B shows the predicted probabilities for correct categorization of a stop based on its word Position. The model predicts that listeners provide a correct categorization on 61% of initial stop trials, but only 53% of final stop trials.

4.2.3. Interactions

4.2.3.1. Group by place

For the two-way interaction term between listener Group and stimulus Place, the model shows significant effects at two of the four Places. At labial Place, there is a negative effect of NKS Group ($\beta = -0.45$, 95% CI: $[-0.69, -0.21]$; $z = -3.62, p < 0.001$), accompanied by a positive effect of NSS Group ($\beta = 0.59, [0.36, 0.81]$; $z = 5.04, p < 0.001$). At uvular Place, on the other hand, there is a positive effect of NKS ($\beta = 0.63, [0.36, 0.91]$; $z = 4.51, p < 0.001$), with a negative effect of NSS ($\beta = -0.85, [-1.12, -0.59]$; $z = -6.26, p < 0.001$). NES again has no significant effect, never differing significantly from the expected mean ($p > 0.071$ at all three Places).

The predicted probability of a correct response for each Group×Place combination is shown in Figure 7. NKS listeners have a predicted probability of 67% correct for labials, 78% for coronals, 69% for velars, and 82% for uvulars. For NSS, these are 53% for labials, 46% for coronals, 36% for velars, and 17% for uvulars. For NES, the predicted probabilities are 58% for labials, 60% for coronals, 59% for velars, and 60% for uvulars.

4.2.3.2. Group by Laryngeal

The model shows that none of the two-way interactions between listener Group and stimulus Laryngeal are significant ($p > 0.058$ for each). A listener from each Group is predicted to exhibit the main effects of their Group and of the stop's Laryngeal feature. Thus, NKS listeners have significantly higher than average accuracy, while NSS listeners have significantly lower than average accuracy, while they are significantly less likely than average to correctly categorize a Plain stop, but are significantly more likely than average to correctly categorize a Glottalized stop. For space considerations, I do not show the probability plot for this non-significant interaction term.

The predicted probability for a NKS listener to correctly categorize a plain stop is 69%, but 79% for a glottalized stop. NSS listeners are predicted to correctly categorize plain stops on 30% of trials, but for Glottalized stop trials this is 43%. NES listeners are predicted to correctly categorize 57% of plain stop trials and 61% of glottalized stop trials.

4.2.3.3. Place by Laryngeal

Finally, for the two-way interaction term between stimulus Place and Laryngeal features, the model shows significant effects of Laryngeal at three of the four levels of Place: labial, coronal, and uvular. Plain labials are more likely to be categorized correctly than average ($\beta = 0.29$, 95% CI: $[0.13, 0.45]$; $z = 3.49, p < 0.001$) while glottalized labials are less likely ($\beta = -0.29$, 95% CI: $[-0.45, -0.14]$; $z = -3.49, p < 0.001$). This is also true for coronals, with plain /t/ more likely to be categorized correctly ($\beta = 0.39, [0.22, 0.55]$; $z = 4.64, p < 0.001$) and glottalized /t'/ less likely ($\beta = -0.39, [-0.55, -0.22]$; $z = -4.64, p < 0.001$). However, *plain* uvulars are *less* likely to be categorized correctly ($\beta = -0.70, [-0.88, -0.51]$; $z = -7.44, p < 0.001$), but *glottalized* uvulars *more* likely ($\beta = 0.70, [0.51, 0.88]$; $z = 7.44, p < 0.001$). Velar accuracy does not significantly differ from its predicted mean for either of its two Laryngeal specifications ($z = \pm 0.28, p = 0.782$).

Figure 8 shows the predicted probability of a correct response for each Place×Laryngeal combination. The predicted probability of a correct response for plain labials is 61%, but for glottalized labials it is 58%. For plain coronals this predicted probability is 66%, compared to glottalized coronals' 58%. For plain velars, the predicted probability of a correct categorization is 50%, while for glottalized velars it is 59%. Finally, the plain uvular predicted probability of a correct categorizations is 31%, but for glottalized uvulars it is 73%.

5. Discussion

5.1. Significance

5.1.1. Research answers

Answering the research questions in §2.4, the results indicate that:

1. L1 Spanish, L2 English multilinguals differ from their L1 English, L2 Spanish counterparts in their ability to correctly categorize some of Kaqchikel's stop consonants.
2. L1 English appears privileged, as when differences were significant, L1 English, L2 Spanish multilinguals are more accurate than L1 Spanish, L2 English multilinguals.
3. The difference is limited to the Place contrast, and specifically the categorization of uvulars.
4. Positional differences are minimal, but not insignificant. Closer investigation of the effects of positional allophony on both L1 and learners' categorization accuracy is warranted.

5.1.2. Group differences

In the mixed effects logistic regression model there is a significant interaction between the L1 Group and the Place of the stimulus stop on accuracy. Specifically, NSS accuracy is significantly lower for uvulars compared to both other L1 Groups. NSS do particularly well at categorizing labials, as there is a significant positive interaction between NSS and labial Place, showing they aren't failing the task outright. They are simply poor at categorizing uvulars.

Meanwhile, the NES group never differ significantly from predicted mean accuracy. They are not involved in any significant Group by Place interaction effects. Their predicted accuracy is always well above chance, and always lies between NKS above them

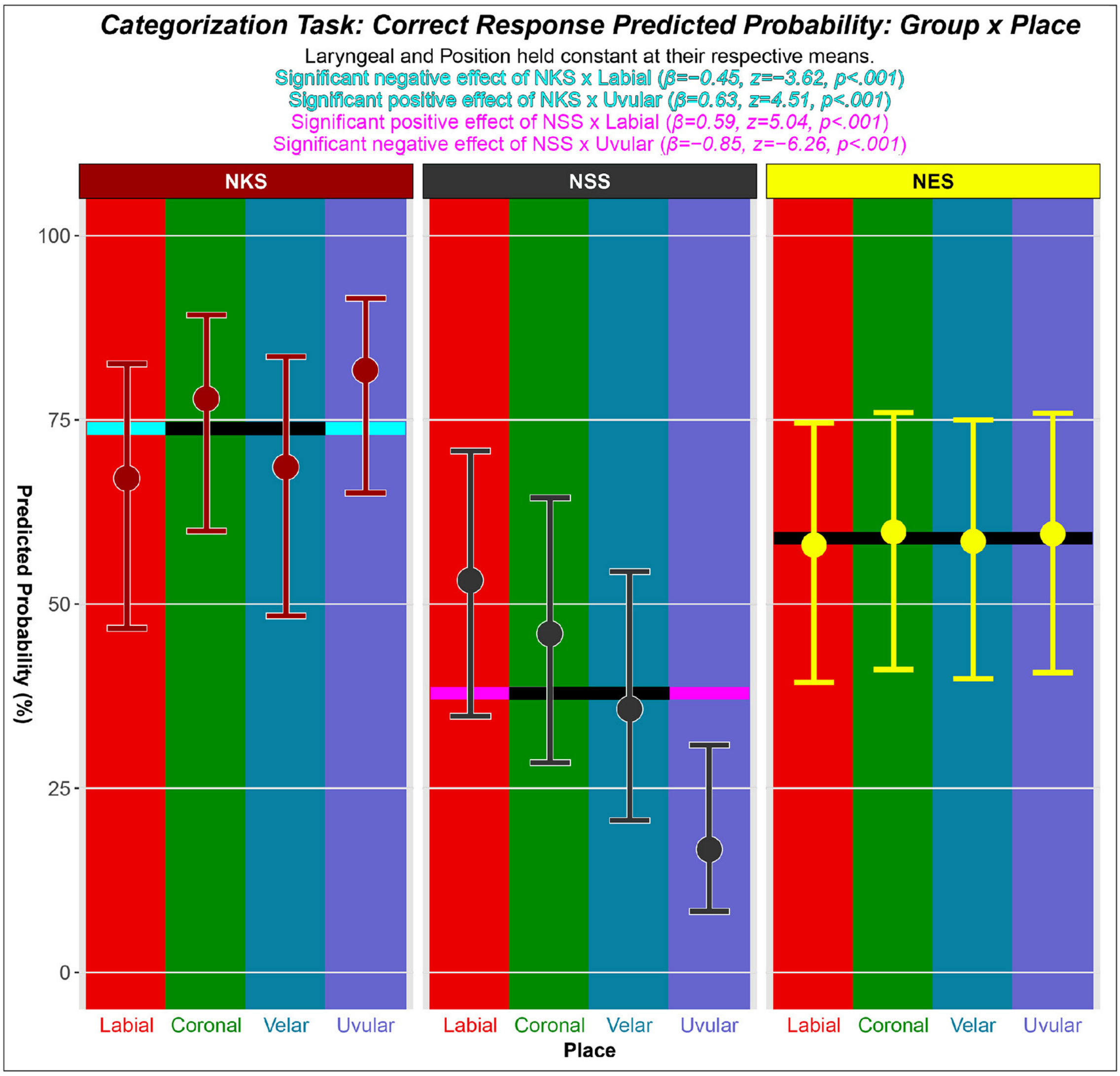


FIGURE 7
Probability plot for the Group by Place interaction based on the mixed effects logistic regression analysis. The horizontal line within each Group's plot represents that Group's mean predicted probability, to which each level of Place was compared. Blacked-out sections of these lines indicate that there was not a significant interaction for the respective levels of Group and Place.

and NSS below them. However, the difference between NES and NSS predicted probabilities is significant only at uvular Place.

5.1.3. Uvular categorization

Examining the NSS group's uvular categorizations, the proximity to floor performance (in terms of both observed accuracy and predicted probability), especially with plain uvulars, is a major contributor to this effect. In short, NSS listeners are much less able to identify and categorize /q/ than are the other two groups. I did not predict greater confusion, and therefore difficulty, for either learner group compared to the other for any Place of articulation. I had made this prediction because the relevant Place specifications are virtually identical in Spanish and English. Both systems exhibit a three-way contrast between bilabials, coronals, and velars. Neither system seemed to provide an advantage (or disadvantage) over the other if it was selected for transfer into the L3 system. Therefore, the current results showing such a difference between learner groups would seem quite surprising. However, the patterns observed here need not be so surprising!

Phonological redeployment, as discussed in §2.1, is the L2A process proposed by Archibald (2005) by which learners of a language utilize components of phonological knowledge to account for differences between their known language and their target language. I did predict that NES would be able to redeploy their Laryngeal feature [spread glottis] or potential non-contrastive uses of [constricted glottis] to account for Kaqchikel's Laryngeal distinction. While that result is not as apparent, there is still a potential case of redeployment at play in the present results, specifically in the differences in uvular categorization. While there is no major difference in the Place features of stop consonants

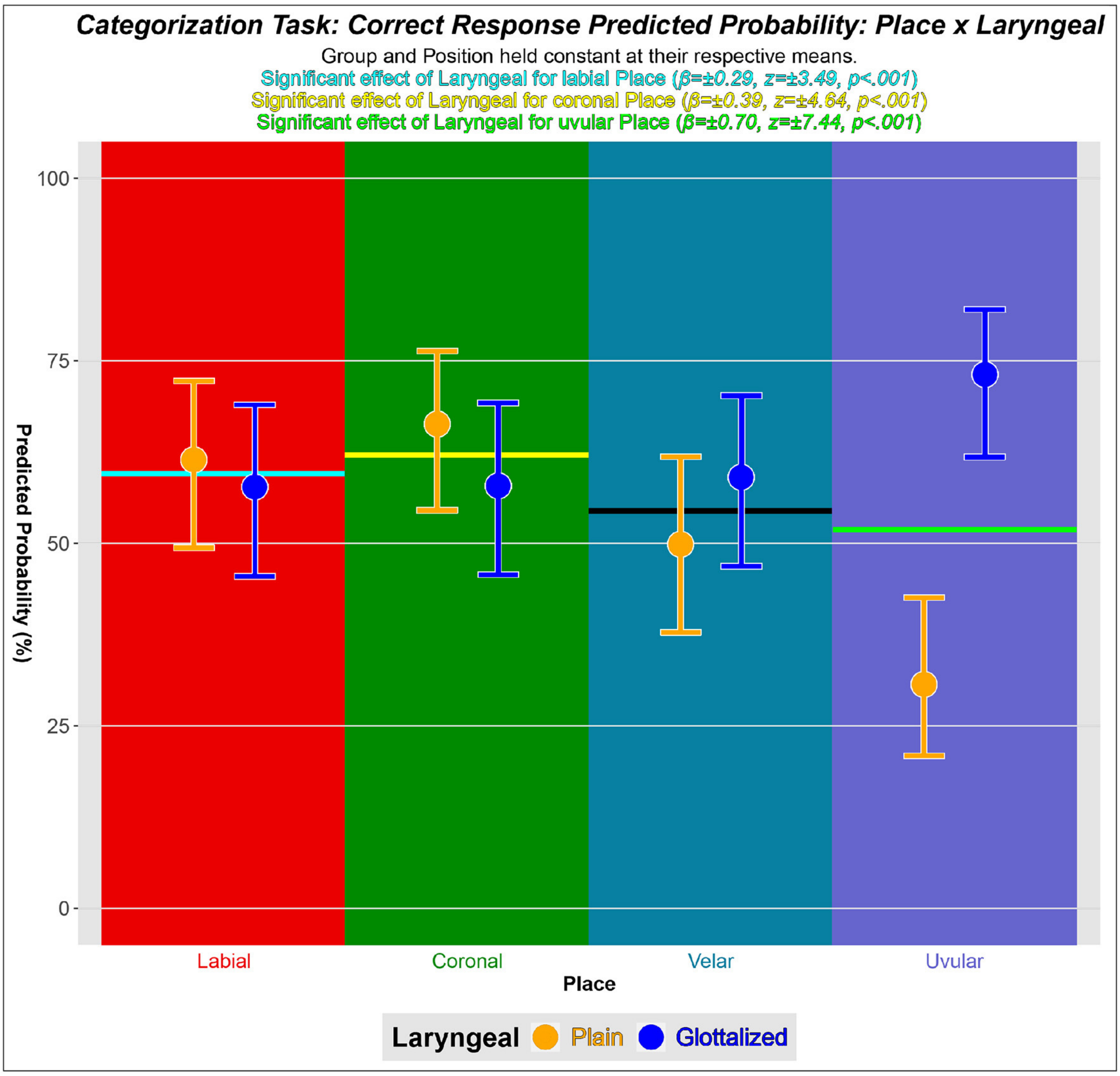


FIGURE 8
Probability plot for the Place by Laryngeal interaction based on the mixed effects logistic regression analysis. The horizontal line within each Place's plot represents that Place's mean predicted probability, to which each level of Laryngeal was compared. Blacked out sections indicate that there was no significant interaction for the respective levels of Place and Laryngeal.

in Spanish and English, English does use more distinctive phonological features to specify its vowels than Spanish uses for its vocalic inventory.

Spanish requires fewer features to specify its vowels than does English. Those features are [±high], [±back], and [±low]. English, on the other hand, requires more features, adding [±round] and [±tense]. [±tense] behaves remarkably similar to [±ATR] cross-linguistically, especially in regards to vocalic specification (Beltzung et al., 2015). Thus, if [±tense] and [±ATR] make equivalent specifications, and [±ATR] has an inverse feature in [±RTR], then a relationship between [±tense] and [±RTR] exists. This relationship allows for redeployment of the English vocalic feature in specifying Kaqchikel dorsal stops, which are specified by [±RTR]: velars with [−RTR] and uvulars with [+RTR] (Davis, 1995; Shahin, 2002). Analyses of English that attribute its tense–lax distinction directly to the critical feature [±RTR], such as Brown and Golston (2006), avoid the need to establish any analogical relationships between [±tense], [±ATR], and [±RTR]. Representation matters!

Based on the differences in their self-assessment of English speaking skills (Table 2), I assume that these NES learners have more solidified knowledge of English phonology than the NSS learners. I therefore propose that multilingual NES are more likely to redeploy English phonological features than multilingual NSS are. Specifically, this involves taking the vocalic feature used to specify the distinction between tense and lax vowels, whether it be [±tense], or one of the related tongue root features [±ATR] and [±RTR], which they may already be using in specifying the tense–lax contrast of Kaqchikel, and applying it to the novel consonantal contrast between velars and uvulars.

5.1.4. Laryngeal categorization

Shifting the lens of analysis to the laryngeal contrast, the main effect of Laryngeal suggests there is a difference in how listeners, regardless of their L1, categorize plain versus glottalized stops in Kaqchikel. Against my prediction, there is no evidence that either learner Group used knowledge of allophonic glottalized stops in English more than the other Group in their categorization of Kaqchikel glottalized stops. Based on the theory of phonological redeployment in Archibald (2022), this is unsurprising: that would be learners redeploying phonetic rather than phonological knowledge.

However, the significant interaction effects between Place and Laryngeal, particularly as visualized in Figure 8, shows that a primary contributor to higher accuracy of categorizing glottalized stops (both predicted and observed) is the extremely high accuracy of categorizing glottalized uvular stops. A likely explanation for this is the relatively robust acoustic cues associated with the release burst of the uvular implosive /ʛ̥/. As explained by González Poot (2014) and Archibald (2023), such robust cues help the learners notice, and then better account for, contrasts for which they lack explicit phonological knowledge.

5.1.5. Position-dependent cues

Once more, I return to the modest, but significant effect of Position on Kaqchikel stop categorization. The model predicts slightly higher accuracy for categorization of stops in initial position relative to stops in final position, as visualized in Figures 4, 6B. Again, I refer to Archibald (2005), González Poot (2014), and Archibald (2023), who also predicted that onset stops (including initials) would have more robust cues than coda stops (including finals). That prediction is borne out here.

Flege and Bohn (2021) presented a revised version of Flege (1995) SLM, with many changes to its scope, methods of analysis, and predictions. Although phonetic categories and learning remains its focus, the revised SLM de-emphasized effects directly related to the listener's age. Instead they keep open the possibility for continuous learning across the learner's life-span, including re-weighting of perceptual cues, just as Archibald (2023) would later suggest as a primary language learning task along the journey to adequate phonological representations.

5.1.6. Toward an integrated model of L3A

Finally, Rothman (2011, 2015) Typological Primacy Model offers a potential alternative explanation for these results, with a caveat. Under that model, the learner transfers one of their known language systems to serve as the initial state of L3A. After this initial state, additional L3 experience could lead the learner to adjust their L3 interlanguage via cross-linguistic influence (González Alonso and Rothman, 2017; Rothman et al., 2019), potentially via redeployment.

If a learner of Kaqchikel, and perhaps these NSS learners, had selected Spanish, with its relatively few vocalic features, redeployment of [tense]/[RTR] would not be an option. Only after they adjust their interlanguage to allow for the influence of English and its [tense]/[RTR] feature would they then be able to redeploy that feature in optimally mapping both the tense–lax and velar–uvular distinctions of Kaqchikel. However, as these learners are not at the initial stage of their L3A, the source of transfer cannot be determined here and is therefore left to future work that analyzes learners closer to, or at, the initial stage of Kaqchikel L3A.

The main limitations to the results I have presented here are related to this inability to determine the source of initial transfer. These learners represent a very particular subset of learners of Kaqchikel. In order to build the best picture of L3A, both generally and of Kaqchikel, the effects observed here should be investigated in other subsets of learners [a methodological concern going back to at least Cabrelli Amaro (2013a)]. Additionally, as noted by a reviewer, the phonological status of the tense–lax distinction among these NSS is not probed here. This is a notorious learning problem for L1 Spanish learners of L2 English [examined, e.g., in Escudero (2005)]. If a NSS has not acquired this English contrast, then it logically follows that its underlying feature is unavailable for redeployment in L3 Kaqchikel.

5.2. Conclusion

In this study, I examined a case of L3A, in which multilingual learners already familiar with English and Spanish are learning Kaqchikel, a Mayan language of Guatemala. The phonological structures of primary interest were the stop consonants, particularly the glottalization contrast and the distinction of a uvular/post-velar PoA.

Learners, which were grouped based on which of the two known languages was their L1, as well as a comparison group of multilingual L1 Kaqchikel users, completed a audio perception task in which they categorized Kaqchikel's stop consonants into the eight phonemic categories of Kaqchikel. All listeners performed well above chance. However, several effects emerged in their performance on this task. One of these effects was an interaction between L1 Group and Place of articulation: NSS listeners were less accurate in categorizing uvulars than both NES and NKS, indicating they had not yet formed the category of uvular PoA and were unable to contrast them with other places, and particularly velars.

I propose that NES listeners were more accurate than NSS due to having relatively easier access to knowledge of English phonology, which contains more material that can be redeployed when learning an additional language's phonology, and in this case Kaqchikel's. The key feature that the NES listeners have better access to is the vocalic feature [±tense], or alternatively [±RTR] or [±ATR]. Whatever this feature is, its status as a distinctive feature among English vowels allows it to be redeployed to capture the novel category of Kaqchikel uvular stops. NES learners of Kaqchikel redeployed an English feature specific to vowels in order to capture a Place contrast among Kaqchikel stops. NSS learners, on the other hand, do not appear to be able to do this yet.

However, hope is not lost for L1 Spanish L2 English learners of L3 Kaqchikel, or of any order of language learning between these languages or languages with similar distinctions across them. NES appear able to redeploy English's vocalic feature thanks in part to English being their L1, with better access to its underlying phonology. Nevertheless, these NSS learners should still have access to phonological knowledge of English. It may simply be just a matter of time and practice with L2 English and L3 Kaqchikel before they too make the connections between English lax vowels and Kaqchikel lax vowels and uvular stops already made by their NES counterparts. Thus phonological redeployment, in the L3A context, is a tool available for all to use in language learning, and not one restricted to those who happen to have a L1 with more phonological material available to be redeployed in a given L3A scenario.

5.3. Resource identification initiative

The following resources were used in the creation and presentation of experimental materials or the analysis and visualization of experimental results:

- Adobe Photoshop (RRID:SCR_014199)
- Overleaf (RRID:SCR_003232)
- Praat (RRID:SCR_016564)
- PsychoPy (RRID:SCR_006571)
- Python Programming Language (RRID:SCR_008394)
- R Project for Statistical Computing (RRID:SCR_001905)
- R package: Companion to Applied Regression (RRID:SCR_022137)
- R package: dplyr (RRID:SCR_016708)
- R package: ggeffects (RRID:SCR_022496)
- R package: ggplot2 (RRID:SCR_014601)
- R package: Harrell Miscellaneous (RRID:SCR_022497)
- R package: lme4 (RRID:SCR_015654)
- R package: Regression Modeling Strategies (RRID:SCR_023242)
- R package: tidyverse (RRID:SCR_019186)

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: Open Science Framework (osf.io): https://osf.io/r7kz8 (NKS); https://osf.io/aqsnu (NSS); https://osf.io/vw56h (NES).

Ethics statement

The studies involving humans were approved by Conjoint Faculties Research Ethics Board, University of Calgary. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

BN: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing—original draft, Writing—review & editing.

Funding

Funding for travel, lodging, and compensation of participants was provided personally by BN. Additional reimbursements were made by the institutional/university funding of BN's supervisor, Dr. Darin Flynn.

Acknowledgments

Many sincere thanks go out to associated members of the School of Languages, Linguistics, Literatures and Cultures at the University of Calgary. I offer special thanks to my doctoral thesis supervisor, Dr. Darin Flynn, and members of my thesis committee, Dr. Mary Grantham O'Brien and Dr. Angeliki Athanasopoulou, for aiding in the conception and analysis of the current study. I appreciate Dr. Allison Shapp's helpful second opinion on phonetic transcription. I am grateful to two reviewers for their extremely helpful feedback on how best to describe, analyze, and explain these observations.

Territorial acknowledgments

Different parts of this research took place on the traditional territories of many different Indigenous peoples: some taken by force, some taken by treaty, and some not conceded. These traditional territories include those of the Kaqchikel Maya of the central highlands of southern *Iximulew* (Guatemala); the Chitimacha, Choctaw, and Houma people of *Bulbancha* (Southern Louisiana); and the peoples of Treaty 7 territory in Southern Alberta, where most of the analysis of the current research took place. Treaty 7 peoples include the Blackfoot Confederacy (composed of the Siksika, Piikani, and Kainai First Nations), the Tsuut'ina First Nation, and the Stoney Nakoda, which includes the Chiniki, Bearspaw, and Goodstoney First Nations. Métis Nation of Alberta, Region III also has ties to the land upon which the City of Calgary now sits. The University of Calgary, within the City of Calgary, stands near the land where the Bow and Elbow Rivers meet. The Blackfoot name for this place is *Mohkinstsis*, the Stoney Nakoda call it *Wîchîspa*, and the Tsuut'ina refer to it as *Guts'ists'i*.

Conflict of interest

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

References

Abo-Mokh, N., and Davis, S. (2020). "What triggers 'imāla: Focus on a Palestinian variety with phonological analysis," in *Perspectives on Arabic Linguistics XXXII*, ed. E. van Gelderen (John Benjamins), 35–53. doi: 10.1075/sal.9.02abo

Archibald, J. (2005). Second language phonology as redeployment of L1 phonological knowledge. *Canad. J. Ling.* 50, 285–314. doi: 10.1353/cjl.2007.0000

Archibald, J. (2022). Segmental and prosodic evidence for property-by-property transfer in L3 English in northern Africa. *Languages* 7, 1–20. doi: 10.3390/languages7010028

Archibald, J. (2023). Phonological redeployment and the mapping problem: cross-linguistic e-similarity is the beginning of the story, not the end. *Second Lang. Res.* 39, 287–297. doi: 10.1177/02676583211066413

Bardel, C., and Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Lang. Res.* 23, 459–484. doi: 10.1177/0267658307080557

Barrios, S., Jiang, N., and Idsardi, W. J. (2016). Similarity in L2 Phonology: evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Second Lang. Res.* 32, 367–395. doi: 10.1177/0267658316630784

Beltzung, J.-M., Patin, C., and Clements, G. N. (2015). "The feature [atr]," in *Features in Phonology and Phonetics: Posthumous Writings by Nick Clements and Coauthors*, eds. A. Riallan, R. Ridouane, and H. van der Hulst (Berlin: Walter de Gruyter), 217–246. doi: 10.1515/9783110399981-011

Bennett, R. (2016). Mayan phonology. *Lang. Ling. Compass* 10, 469–514. doi: 10.1111/lnc3.12148

Benrabah, M. (1991). "Learning English segments with two languages," in *Proceedings of the International Congress of Phonetic Sciences*, ed. P. Romeas (Aix-en-Provence: Université de Aix-en-Provence), 334–337.

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, ed. W. Strange (York: York Press), 171–204.

Best, C. T., and Tyler, M. D. (2007). "Nonnative and second-language speech perception: Commonalities and complementaries," in *Second-language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege*, eds. O.-S. Bohn, and M. J. Munro (Amsterdam: John Benjamins), 13–34. doi: 10.1075/lllt.17.07bes

Brown, C. A. (1997). *Acquisition of segmental structure: consequences for speech perception and Second Language Acquisition*. PhD thesis, McGill University.

Brown, C. A. (1998). The role of L1 grammar in the L2 acquisition of segmental structure. *Second Lang. Res.* 14, 136–193. doi: 10.1191/026765898669508401

Brown, C. A. (2000). "The interrelation between speech perception and phonological acquisition: from infant to adult," in *Second Language Acquisition and Linguistic Theory*, ed. J. Archibald (Oxford: Blackwell), 4–63.

Brown, J. C., and Golston, C. (2006). Embedded structure and the evolution of phonology. *Inter. Stud.* 7, 17–41. doi: 10.1075/is.7.1.03bro

Brown, R. M., Maxwell, J. M., and Little, W. E. (2006). *La ütz awäch? Introduction to Kaqchikel Maya Language*. Austin, TX: University of Austin Press.

Cabrelli Amaro, J. (2013a). Methodological issues in L3 phonology. *Stud. Hispan. Lusophone Ling.* 6, 101–117. doi: 10.1515/shll-2013-1142

Cabrelli Amaro, J. (2013b). *The phonological permeability hypothesis: measuring regressive L3 influence to test L1 and L2 phonological representations*. PhD thesis, University of Florida.

Cabrelli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *Int. J. Bilingual.* 21, 698–717. doi: 10.1177/1367006916637287

Cabrelli Amaro, J., and Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *IRAL* 48, 275–296. doi: 10.1515/iral.2010.012

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonol. Yearb.* 2, 225–252. doi: 10.1017/S0952675700000440

Clements, G. N., and Hume, E. (1995). "The internal organization of speech sounds," in *Handbook of Phonological Theory*, ed. J. Goldsmith (Oxford: Blackwell), 245–306.

Davis, S. (1995). Emphasis spread in arabic and grounded phonology. *Ling. Inquiry* 26, 465–498.

Dresher, B. E. (2009). *The Contrastive Hierarchy in Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511642005

Duanmu, S. (2007). *The Phonology of Standard Chinese*. 2nd edition. Oxford: Oxford University Press.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Lang. Learn.* 27, 315–330. doi: 10.1111/j.1467-1770.1977.tb00124.x

Escudero, P. (2005). *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. PhD thesis, Utrecht University.

Flege, J. E. (1995). "Second language speech learning: Theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, W. Strange (York: York Press), 233–277.

Flege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed. R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002

Ghazali, S., and Bouchhioua, N. (2003). "The learning of English prosodic structures by speakers of Tunisian Arabic: Word stress and weak forms," in *15th International Conference on Phonetic Sciences* 961–964.

González Alonso, J., and Rothman, J. (2017). Coming of age in L3 initial stages transfer models: deriving developmental predictions and looking towards the future. *Int. J. Biling.* 21, 683–697. doi: 10.1177/1367006916649265

González Poot, A. A. (2011). *Conflict resolution in the Spanish SLA of Yucatec ejectives: L1, L2 and universal constraints*. PhD thesis, University of Calgary.

González Poot, A. A. (2014). "Conflict resolution in the Spanish L2 acquisition of Yucatec ejectives: L1, L2 and universal constraints," in *Proceedings of the 2014 annual conference of the Canadian Linguistic Association / Actes du congrés de l'Association canadienne de linguistique 2014* 1–15.

Halle, M., Vaux, B., and Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Ling. Inquiry* 31, 387–444. doi: 10.1162/002438900554398

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition," in *Cross-Linguistic Influence in Third Language Acquisition*, eds. J. Cenoz, B. Hufeisen, and U. Jessner (Bristol, UK: Multilingual Matters), 21–41. doi: 10.21832/9781853595509-003

Hayward, K. M., and Hayward, R. J. (1989). "Guttural": Arguments for a new distinctive feature. *Transact. Philol. Soc.* 87, 179–193. doi: 10.1111/j.1467-968X.1989.tb00626.x

Heaton, R., and Xoyón, I. (2016). Assessing language acquisition in the Kaqchikel program at Nimaläj Kaqchikel Amaq'. *Lang. Document. Conserv.* 10, 497–521.

Honeybone, P. (2001). *Germanic obstruet lenition: some mutual implications of theoretical and historical phonology*. PhD thesis, University of Newcastle upon Tyne.

Iverson, G. K., and Salmons, J. C. (1995). Aspiration and laryngeal representation in germanic. *Phonology* 12, 369–396. doi: 10.1017/S0952675700002566

Jakobson, R. (1962). "Mufaxxama: The "emphatic" phonemes in Arabic," in *Phonological Studies, Selected Writings*, ed. R. Jakobson (Clearwater, FL: Mouton and Co.), 510–522.

Kim, H., and Clements, G. N. (2015). "The feature [tense]," in *Features in Phonology and Phonetics: Posthumous Writings by Nick Clements and Coauthors*, eds. A. Riallan, R. Ridouane, and H. van der Hulst (Berlin: Walter de Gruyter), 159–178. doi: 10.1515/9783110399981-008

LaCharité, D., and Pré vost, P. (1999). "The role of L1 and of teaching in the acquisition of English sounds by francophones," in *BUCLD 23 Proceedings*, eds. A. Greenhill, H. Littlefield, and C. Tano (Somerville, MA: Cascadilla Press), 373–385.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago, IL: University of Chicago Press.

Ladefoged, P., and Johnson, K. (2015). *A Course in Phonetics*. 7th edition. Boston, MA: Cengage Learning.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.

Lin, Y.-H. (2007). *The Sounds of Chinese*. Cambridge: Cambridge University Press.

Mah, J. (2003). *The acquisition of phonological features in a second language*. Master's thesis, University of Calgary.

Mah, J. (2011). *Segmental representations in interlanguage grammars: the case of francophones and English /h/*. PhD thesis, McGill University.

Mah, J., Goad, H., and Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: The case of French speakers and English [h]. *Front. Psychol.* 7, 1469. doi: 10.3389/fpsyg.2016.01469

Maxwell, J. M., and Little, W. E. (2006). Tijonïk Kaqchikel Oxlajuj Aj: Curso de Idioma y Cultura Maya Kaqchikel [Kaqchikel Maya Language and Culture Course]. Editorial Junajpu.

McCarthy, J. J. (1994). "The phonetics and phonology of Semitic pharyngeals," in *Phonological Structure and Phonetic Form, Papers in Laboratory Phonology*

III, ed. P. A. Keating (Cambridge: Cambridge University Press), 191–233. doi: 10.1017/CBO9780511659461.012

Nelson, B. C. (2023). "Insertion of [spread glottis] at the right edge of words in Kaqchikel," in *Epenthesis and Beyond Workshop 2021* 35. doi: 10.4324/9781032682907-4

Nelson, B. C. (forthcoming). *Learning the sounds of silence: adult acquisition of Kaqchikel (Mayan) plain and glottalized stop consonants*. PhD thesis, University of Calgary.

Ortin, R., and Fernandez-Florez, C. (2019). Transfer of variable grammars in third langauge acquisition. *Int. J. Multiling*. 16, 442–458. doi: 10.1080/14790718.2018.1550088

Padgett, J. (2002). Feature classes in phonology. *Language* 78, 81–110. doi: 10.1353/lan.2002.0046

Patal Majzul, F., García Matzar, L. P. O., and Espantzay Serech, I. C. (2000). *Rujunamaxik ri Kaqchikel chi' : Variación dialectal en Kaqchikel [Dialectal variation in Kaqchikel*]. Cholsamaj.

Price, L., Pouplier, M., and Hoole, P. (2020). Ejective production mechanisms in English. *J. Acoust. Soc. Am*. 148, 2655. doi: 10.1121/1.5147392

Price, L., Pouplier, M., and Hoole, P. (2022). "Kehlkopfanhebung in englischen wortfinalen Ejektiven: eine Echtzeit-MRT-Studie [Laryngeal raising in English word-final ejectives: a real-time MRI study]," in *18. Phonetik und Phonologie im deutschsprachigen Raum [18th conference on Phonetics and Phonology in the German-speaking area*].

Rill, J. (2013). "Vowel tensing in Kaqchikel Maya," in *Penn Linguistics Conference* 37.

Rose, S. (1996). Variable laryngeals and vowel lowering. *Phonology* 13, 73–117. doi: 10.1017/S0952675700000191

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: the typological primacy model. *Second Lang. Res*. 27, 107–127. doi: 10.1177/0267658310386439

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered. *Bilingualism*. 18, 179–190. doi: 10.1017/S136672891300059X

Rothman, J., González Alonso, J., and Puig-Mayenco, E. (2019). "Transfer in multilingual morphosyntax," in *Third Language Acquisition and Linguistic Transfer* (Cambridge: Cambridge University Press), 116–187. doi: 10.1017/9781316014660.004

Selinker, L. (1972). Interlanguage. *Int. Rev. Appl. Ling. Lang. Teach*. 10, 209–231. doi: 10.1515/iral.1972.10.1-4.209

Shahin, K. N. (2002). *Postvelar Harmony*. Amsterdam: John Benjamins. doi: 10.1075/cilt.225

Sylak Glassman, J. C. (2014). *Deriving natural classes: the phonology and typology of post-velar consonants*. PhD thesis, University of California, Berkeley.

Torres-Tamarit, F. (2019). "Phonemic contrast and neutralization," in *The Routeledge Handbook of Spanish Phonology*, eds. S. Colina, and F. Martínez-Gil (London: Routledge), 3–33. doi: 10.4324/9781315228112-2

Van Patten, B., and Benati, A. G. (2010). *Key Terms in Second Language Acquisition*. London: Bloomsbury Publishing.

Wagner, K. O. C., and Baker-Smemoe, W. (2013). An investigation of the production of ejectives by native (L1) and second (L2) language speakers of Q'eqchi' Mayan. *J. Phonet*. 41, 453–467. doi: 10.1016/j.wocn.2013.08.002

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: the linguistic proximity model. *Int. J. Biling*. 21, 666–682. doi: 10.1177/1367006916648859

Wrembel, M., Marecka, M., and Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *Int. J. Multiling*. 16, 513–533. doi: 10.1080/14790718.2019.1583233

Yang, Y., Chen, X., and Xiao, Q. (2022). Cross-similarity in L2 speech learning: evidence from the acquisition of Russian stop contrasts by Mandarin speakers. *Second Lang. Res*. 38, 3–29. doi: 10.1177/0267658319900919

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 20 December 2023
DOI 10.3389/flang.2023.1303511

Check for updates

OPEN ACCESS

EDITED BY
Baris Kabak,
Julius Maximilian University of Würzburg, Germany

REVIEWED BY
Mathias Scharinger,
University of Marburg, Germany
Silke Hamann,
University of Amsterdam, Netherlands

*CORRESPONDENCE
Kakeru Yazawa
✉ yazawa.kakeru.gb@u.tsukuba.ac.jp
James Whang
✉ jamesw@snu.ac.kr

RECEIVED 28 September 2023
ACCEPTED 22 November 2023
PUBLISHED 20 December 2023

CITATION
Yazawa K, Whang J, Kondo M and Escudero P (2023) Feature-driven new sound category formation: computational implementation with the L2LP model and beyond.
Front. Lang. Sci. 2:1303511.
doi: 10.3389/flang.2023.1303511

Feature-driven new sound category formation: computational implementation with the L2LP model and beyond

Kakeru Yazawa[1]*, James Whang[2]*, Mariko Kondo[3,4] and Paola Escudero[5]

[1]Institutes of Humanities and Social Sciences, University of Tsukuba, Tsukuba, Japan, [2]Department of Linguistics, Seoul National University, Seoul, Republic of Korea, [3]School of International Liberal Studies, Waseda University, Tokyo, Japan, [4]Graduate School of International Culture and Communication Studies, Waseda University, Tokyo, Japan, [5]The MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, NSW, Australia

One of the primary questions of second language (L2) acquisition research is how a new sound category is formed to allow for an L2 contrast that does not exist in the learner's first language (L1). Most models rely crucially on perceived (dis)similarities between L1 and L2 sounds, but a precise definition of what constitutes "similarity" has long proven elusive. The current study proposes that perceived cross-linguistic similarities are based on feature-level representations, not segmental categories. We investigate how L1 Japanese listeners learn to establish a new category for L2 American English /æ/ through a perception experiment and computational, phonological modeling. Our experimental results reveal that intermediate-level Japanese learners of English perceive /æ/ as an unusually fronted deviant of Japanese /a/. We implemented two versions of the Second Language Linguistic Perception (L2LP) model with Stochastic Optimality Theory—one mapping acoustic cues to segmental categories and another to features—and compared their simulated learning results to the experimental results. The segmental model was theoretically inadequate as it was unable explain how L1 Japanese listeners notice the deviance of /æ/ from /a/ in the first place, and was also practically implausible because the predicted overall perception patterns were too native English-like compared to real learners' perception. The featural model, however, showed that the deviance of /æ/ could be perceived due to an ill-formed combination of height and backness features, namely */low, front/. The featural model, therefore, reflected the experimental results more closely, where a new category was formed for /æ/ but not for other L2 vowels /ɛ/, /ʌ/, and /ɑ/, which although acoustically deviate from L1 /e/, /a/, and /o/, are nonetheless featurally well-formed in L1 Japanese, namely /mid, front/, /low, central/, and /mid, back/. The benefits of a feature-based approach for L2LP and other L2 models, as well as future directions for extending the approach, are discussed.

KEYWORDS

Second Language Linguistic Perception (L2LP) model, Stochastic Optimality Theory, Gradual Learning Algorithm, category formation, features, computational modeling, Japanese, American English

1 Introduction

Second language (L2) learners often encounter a "new" sound that does not exist in their first language (L1). Establishing a phonological representation for such new sounds is essential to L2 learning, because otherwise the lexical distinctions denoted by the phonological contrast cannot be made for successful communication. Various models have been proposed to explain how a new sound category may develop in the learner's mind, with most models focusing on the cross-linguistic perceptual relationships between L1 and L2 sounds, although the exact underlying mechanism remains to be elucidated. In this study, we propose that the process of L2 category formation can be better explained by assuming feature-level representations as the fundamental unit of perception, rather than segmental categories.[1] To this end, we compare two versions of formal modeling, i.e., segment- and feature-based, of how L1 Japanese listeners form a new category for L2 American English (AmE) /æ/[2] by implementing the theoretical predictions of the Second Language Linguistic Perception (L2LP) model (Escudero, 2005; van Leussen and Escudero, 2015; Escudero and Yazawa, in press) with a computational-phonological approach of Stochastic Optimality Theory (StOT; Boersma, 1998).

In the field of L2 speech perception research, two models have been particularly dominant over the last few decades (Chen and Chang, 2022): the Speech Learning Model (SLM; Flege, 1995; Flege and Bohn, 2021) and the Perceptual Assimilation Model (PAM; Best, 1995; Best and Tyler, 2007). According to SLM, learners can form a new category for an L2 sound if they discern its phonetic difference(s) from the closest L1 category and, if not, a single composite category will be used to process both L1 and L2 sounds. The likelihood of category formation therefore depends primarily on the perceived cross-linguistic phonetic dissimilarity, but other factors such as the quantity and quality of L2 input obtained in meaningful conversations are said to be also relevant. PAM agrees with SLM in that perceived cross-linguistic dissimilarity guides category formation, but with the caveat that assimilation occurs not only at the phonetic level but also at the phonological or lexical-functional level. For example, if there are many minimal pairs involving an L2 contrast that assimilates phonetically to a single L1 category, the increased communicative pressure can lead to the formation of a new phonological category to allow for distinct phonological representations of these lexical items (e.g., AmE [ʌ] and [ɑ] both being assimilated to Japanese [a], but AmE *nut* [nʌt] and *not* [nɑt] leading to Japanese /na$_1$t/ and /na$_2$t/, where /a$_1$/ and /a$_2$/ are distinct phonological categories that occupy phonetically overlapping but distinct parts of a single L1 category). While the predictions of both models have been supported by numerous studies, there is one fundamental issue that remains to be resolved: It is unclear on what basis categorical similarity should be defined. In the words of Best and Tyler (2007, p.26), "one issue [...] has not yet received adequate treatment in any model of nonnative or L2 speech perception: How listeners identify nonnative phones as equivalent to L1 phones, and the level(s) at which this occurs." Over a decade later, Flege and Bohn (2021, p.31) restated the unresolved problem: "It remains to be determined how best to measure cross-language phonetic dissimilarity. The importance of doing so is widely accepted but a standard measurement procedure has not yet emerged."

To illustrate this elusive goal with a concrete example, consider our case of L1 Japanese listeners learning /æ/ and adjacent vowels in L2 AmE. In cross-linguistic categorization experiments, Strange et al. (1998) found that the AmE vowel was perceived as a very poor exemplar of Japanese /a(a)/,[3] receiving the lowest mean goodness-of-fit rating (two out of seven) among all AmE vowel categories, while spectrally adjacent /ɛ/, /ʌ/, and /ɑ/ received higher ratings as Japanese /e/ (four out of seven), /a/ (four out of seven), and /a(a)/[4] (six out of seven), respectively. Duration-based categorization of AmE vowels as Japanese long and short vowels was observed when the stimuli were embedded in a carrier sentence, but not when they were presented in isolation. Shinohara et al. (2019) further found that the category goodness of synthetic vowel stimuli as Japanese /a/ deteriorated as the second formant (F2) frequency was increased. These studies suggest that AmE /æ/ is perceptually dissimilar from L1 Japanese /a(a)/, presumably in terms of F2 but possibly in conjunction with other cues such as the first formant (F1) frequency and duration, and is thus subject to new category formation according to SLM and PAM. L2 perception studies on Japanese listeners also showed that the AmE /æ/-/ʌ/ contrast was more discriminable than the /ɑ/-/ʌ/ contrast (Hisagi et al., 2021; Shafer et al., 2021; Shinohara et al., 2022) and that AmE /æ/ was identified with higher accuracy than /ʌ/ or /ɑ/ (Lambacher et al., 2005). These results imply that Japanese listeners perceptually distinguish AmE /æ/ from AmE /ʌ/ and /ɑ/, which themselves are assimilated to Japanese /a(a)/. However, it remains unclear why only AmE /æ/ would be perceptually distinct in the first place. Spectral distance between the L1 and L2 categories in Figure 1, which shows the production of Japanese and AmE vowels by four native speakers of each language (Nishi et al., 2008), does not seem

1 Our view of features departs from the generally assumed universal set of binary phonological features. Specifically, we assume in this paper that features are language-specific and emergent rather than universal and innate (Boersma et al., 2003, 2022), that they are privative rather than binary (Chládková et al., 2015b), and that they are phonetically based (i.e., tied to acoustic cues) but still phonological (i.e., phonemically distinctive) in nature (Boersma and Chládková, 2011), all of which are also assumed in the Second Language Linguistic Perception model discussed below (cf. Boersma, 2009; Escudero, 2009). These specifications of features will be discussed in more detail in later sections.

2 AmE is considered as the target variety of English because it is widely used in the formal English language education in Japan and is most familiar to the learners (Sugimoto and Uchida, 2020).

3 Japanese has five vowel qualities /i/, /e/, /a/, /o/, and /u/, which form five short (1-mora) and long (2-mora) pairs. Long vowels are transcribed with double letters (e.g., /aa/) in this study because they can underlyingly be a sequence of two identical vowels. The transcription "/a(a)/" here indicates "either /a/ or /aa/" because the AmE vowel was perceived as Japanese /a/ when presented in isolation but as Japanese /aa/ when embedded in a carrier sentence.

4 Similar to AmE /æ/, AmE /ɑ/ was perceived as Japanese /a/ when presented in isolation but as Japanese /aa/ when embedded in a carrier sentence.

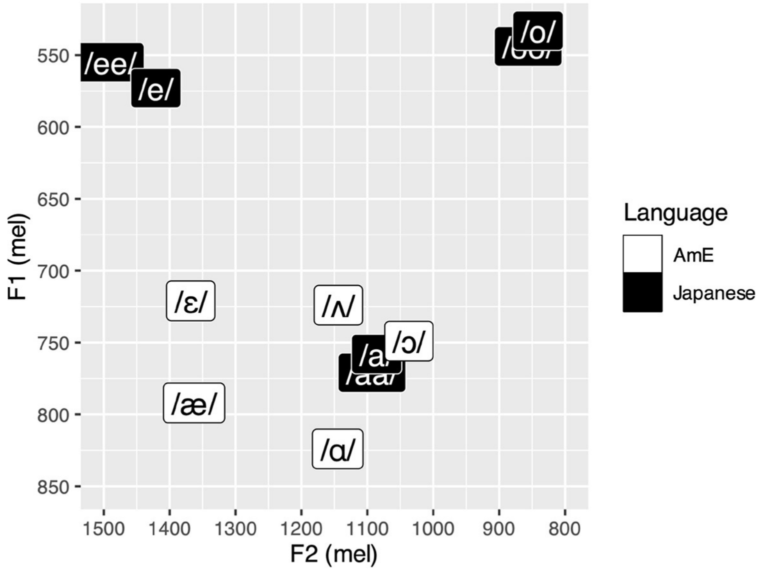


FIGURE 1
Mel-converted average F1 and F2 frequencies of relevant AmE and Japanese vowels. Adapted with permission from Nishi et al. (2008), licensed under Copyright 2008, Acoustical Society of America.

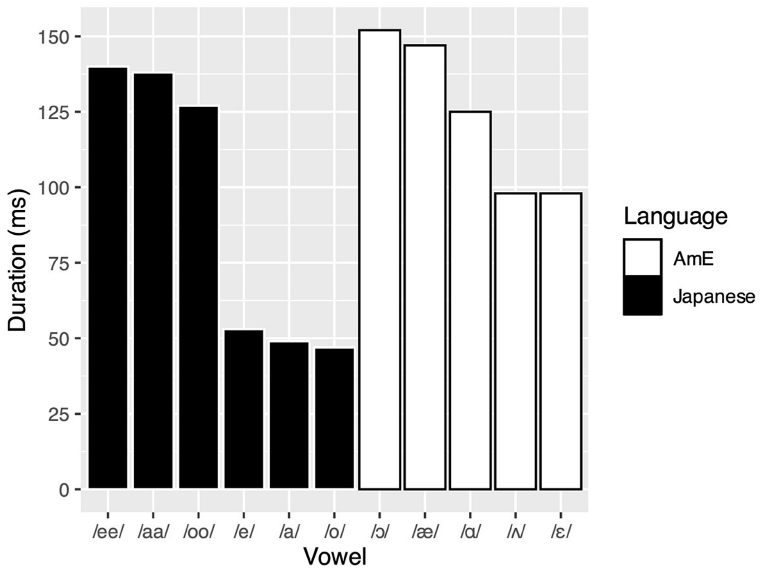


FIGURE 2
Average duration of relevant AmE and Japanese vowels. Adapted with permission from Nishi et al. (2008), licensed under Copyright 2008, Acoustical Society of America.

to predict perceived category goodness very well. For example, the figure shows that AmE /ʌ/ is spectrally closer than AmE /ɑ/ to Japanese /a(a)/, but it was the latter AmE vowel that was judged to be a better fit in Strange et al. (1998). AmE /ɛ/ is also quite far from Japanese /e/ in spectral distance, but its perceived category goodness as Japanese /e/ was nonetheless as high as that of AmE /ʌ/ as Japanese /a/. One may attribute this pattern to duration given the duration-based categorization in Strange et al. (1998), but the actual duration of the target vowels (Figure 2) does not seem to provide useful clues, either. For example, since AmE /ɛ/ and /ʌ/ have almost identical duration values, the former is acoustically more distant from the L1 categories after all, despite both AmE vowels receiving equal goodness ratings. This brings us back to the question: How is L1-L2 perceptual dissimilarity determined?

The L2LP model approaches L2 category formation from a different perspective. While the model shares with SLM and PAM the view that perceptual learning is both auditory- and meaning-driven, it is unique in assuming the interplay of multiple levels of linguistic representations. Although many L2LP studies have focused on the perceptual mapping of acoustic cues onto segmental representations, some have also incorporated feature-level representations, which may be useful for modeling the perception of new L2 sounds. For example, Escudero and Boersma (2004) proposed that L1 Spanish learners' overuse of the duration cue in perceiving L2 Southern British English (SBE) /iː/-/ɪ/ contrast can be adequately modeled by assuming that the vowels are perceptually represented as "/i, long/" and "/i, short/," respectively. That is, the learners developed a new length feature that does not exist in their L1 (because vowel length is non-phonemic in Spanish) and integrated the feature with an existing L1 segmental representation, yielding a perceptual pattern that is not seen in either Spanish or English. This type of learning scenario is called the UNFAMILIAR NEW scenario in L2LP, where L2 representations outnumber L1 representations and thus learners must establish a new category to bridge the cross-linguistic gap (hence NEW) but an important cue for the L2 contrast is not utilized in L1 phonology (hence UNFAMILIAR). The current learning scenario of our interest is also considered NEW because AmE has more vowels than Japanese, but the necessary cues for optimal perception of the target L2 vowels—F1, F2, and duration—are all FAMILIAR (because Japanese vowels contrast in height, backness, and length). We hypothesize that a feature-based modeling as in Escudero and Boersma (2004) may also be useful for modeling the FAMILIAR NEW scenario, although no previous study has formally tested this possibility yet. Another unique characteristic of L2LP is that the model's theoretical components can be computationally implemented, or simulated, to provide more concrete and testable predictions. While various computational frameworks can be used for this purpose, previous studies have generally used StOT (Escudero and Boersma, 2004; Boersma and Escudero, 2008; Yazawa et al., 2020) because it outperforms other machine learning algorithms (Escudero et al., 2007) and is compatible with the phonological theory of Optimality Theory (OT; Prince and Smolensky, 1993). The current study follows this line of work and evaluates how segment- and feature-based StOT modeling compare in explaining the process of new L2 category formation.

The incentive for feature-based modeling is not only theoretically grounded but also empirically motivated, as emerging evidence suggests the involvement of features in L1 and L2 perception. With respect to native perception, Scharinger et al. (2011) used magnetoencephalography (MEG) to map the entire Turkish vowel space onto cortical locations and found that dipole locations could be structured in terms of features (height, backness, and roundedness) rather than raw acoustic cues (F1, F2, and F3). Mesgarani et al. (2014) further used high-density direct cortical surface recordings to reveal the representation of the entire AmE sound inventory, finding response selectivity at single electrodes corresponding to features (voice, place, manner, height, and backness) rather than individual vowels and consonants. Given these results, it seems reasonable to assume that L2 sounds are also perceived through (L1) features. While most L2 perception studies have focused on segmental categories, some have explored the potential role of L1 features, with a prominent focus on phonological length (and lack thereof). Perhaps the best known study is McAllister et al. (2002), who compared the perception of L2 Swedish vowel length by L1 listeners of Estonian, AmE, and Spanish, where only Estonian has contrastive vowel length. The study found that the Estonian group outperformed the other two groups in perceptual accuracy, suggesting that the L1 length feature is positively transferred to L2 perception or, to put it another way, the lack of the length feature is negatively transferred. Pajak and Levy (2014) extended this finding by showing that native listeners of a language with vowel length contrasts showed enhanced discrimination of nonnative consonant length contrasts (i.e., geminates). This finding suggests that the L1 length feature may be shared across vowels and consonants, which appears to be accessible in L2 perception. Research on native Australian English listners (Tsukada, 2012; Tsukada et al., 2018; Yazawa et al., 2023) has also found that they can discriminate and identify Japanese vowel and consonant length contrasts fairly well without any prior knowledge or training, contrary to native AmE listeners struggling to learn the contrasts (Hirata, 2004, 2017). Taken together, previous research suggests that the presence or absence of a certain feature in the L1 (or its specific variety) predicts the ease or difficulty of L2 perception. However, to our knowledge, no prior study has provided a formal account of how existing L1 features mediate L2 category formation, which is what we aim to achieve in this study.

The remainder of this paper is organized as follows. First, in Section 2, we present a forced-choice perception experiment that investigates the use of spectral and temporal cues in the perception of the L1 Japanese and L2 AmE vowels of interest. This is intended to complement the previous studies, which did not investigate potential effects of F1 and duration cues. Section 3 then presents a formal computational modeling of new L2 category formation within the L2LP framework. Two versions of StOT-based simulations are compared, namely segmental and featural, to evaluate which better explains and replicates the experimental results. We then discuss the experimental and computational results together in Section 4, addressing the implications of feature-based modeling for L2 speech perception models (i.e., L2LP, SLM, and PAM) as well as the directions for future research. Finally, Section 5 draws the conclusion.

2 Experiment

The perception experiment reported in this section was designed to investigate how L2 AmE /æ/ and adjacent vowels are perceived in relation to L1 Japanese vowels based on three acoustic cues (F1, F2, and duration), to help model the category formation and cross-linguistic assimilation processes. Following our previous study (Yazawa et al., 2020), the experiment manipulates the ambient language context to elicit L1- and L2-specific perception modes without changing the relevant acoustic properties of the stimuli, as detailed below.

2.1 Participants

Thirty-six native Japanese listeners (22 male, 14 female) participated in the experiment. They were undergraduate or graduate students at Waseda University, Tokyo, Japan, between the ages of 18 and 35 (mean = 21.25, standard deviation = 2.97). All participants had received six years of compulsory English language education in Japanese secondary schools (from ages 13 to 18), which focused primarily on reading and grammar. They had also received some additional English instruction during college, the quality and quantity of which varied according to the courses they were enrolled in. None of the participants had spent more than a total of three months outside of Japan. TOEIC was the most common standardized test of English proficiency taken by the participants (n = 18), with a mean score of 688 (i.e., intermediate level). All participants reported normal hearing.

2.2 Stimuli

Two sets of stimuli—"Japanese" (JP) and "English" (EN)—were prepared. Both had the same phonetic form [bVs], with the spectral and temporal properties of the vowel varying in an identical manner. The JP stimuli were created from a natural token of the Japanese loanword *baasu* /baasu/ "birth," as produced by a male native Japanese speaker from Tokyo, Japan. The token was phonetically realized as [baːs] because Japanese /u/ can devoice or delete word-finally (Shaw and Kawahara, 2017; Whang and Yazawa, 2023). The EN stimuli, on the other hand, were created from a natural token of the English word *bus* /bʌs/, as produced by a male native AmE speaker from Minnesota, United States. For both tokens, the F1, F2, and duration of the vowel were manipulated with STRAIGHT (Kawahara, 2006) to vary in four psychoacoustically equidistant steps: F1 at 700, 750, 800, and 850 mel; F2 at 1,100, 1,200, 1,300, and 1,400 mel; and duration at 100, 114, 131, and 150 ms (i.e., natural logarithm). These steps were intended to fully cover the spectral and temporal variability of AmE /ɛ/, /æ/, /ʌ/, and /ɑ/, while also partially covering that of Japanese /ee/, /e/, /aa/, and /a/ (Figures 1, 2). The third formant (F3) was set to 1,700 mel. The fundamental frequency and intensity contours were also changed to have a mean of 120 Hz and a peak of 70 dB, respectively. The manipulations resulted in a total of 64 (4 × 4 × 4) stimuli for each of the two language sets.

2.3 Procedure

The experiment included two sessions—again "Japanese" (JP) and "English" (EN)—using the JP and EN stimuli, respectively. In order to elicit language-specific perception modes across sessions, all instructions, both oral and written, were given only in the language of the session. The two sessions were consecutive, and the session order was counterbalanced across participants to control for order effects; 18 participants (11 male, seven female) attended the EN session first, while the other 18 (11 male, seven female) attended the JP session first. In the JP session, participants were first presented with each of the 64 JP stimuli in random order and then chose one of the following four words that best matched what they had heard: *beesu* /beesu/ "base," *besu* /besu/ "Bess," *baasu* /baasu/ "birth," and *basu* /basu/ "bus." The choices are all existing loanwords in Japanese and were written in *katakana* orthography. Participants were instructed that they were not required to use all of the four choices. The block of 64 trials was repeated four times, with a short break in between, giving a total of 256 (64 × 4) trials for the session. The EN session followed a similar a procedure, where participants categorized the randomized 64 EN stimuli as the following four real English words (though there was no requirement to use all choices): *Bess* (/bɛs/), *bass* (/bæs/[5]), *bus* (/bʌs/), and *boss* (/bɑs/). The stimulus block was again repeated four times, for a total of 256 trials for the session.

Participants were tested individually in an anechoic chamber, seated in front of a MacBook Pro laptop running the experiment in the Praat ExperimentMFC format (Boersma and Weenink, 2023) and wearing Sennheiser HD 380 Pro headphones through which the stimuli were played at a comfortable volume. The entire experiment took ∼30 to 40 min to complete, for which monetary compensation was provided.

2.4 Analysis

In order to quantify the participants' use of the acoustic cues, a logistic regression analysis was performed on the obtained response data per session and per response category, using the *glm()* function in R (R Core Team, 2023). The model structure is as follows:

$$\ln(\frac{P}{1-P}) = \alpha + \beta_{F1} \times \text{step}_{F1} + \beta_{F2} \times \text{step}_{F2} + \beta_{dur} \times \text{step}_{dur} \quad (1)$$

where P is the probability that a given response category (e.g., JP /aa/) is chosen, and $1 - P$ is the probability that the other three categories (e.g., JP /ee/, /e/, or /aa/) are chosen. The odds $\frac{P}{1-P}$ is log-transformed to fit a sigmoidal curve to the data, which is more appropriate than the straight line of a linear regression model for analyzing speech perception data. The intercept α is the bias coefficient, which reflects how likely the particular response category is to be chosen in general. The stimulus-tuned coefficients βs represent the extent to which the F1, F2, and duration steps, coded from "1" (smallest) to "4" (largest), cause a change in the likelihood of the response category being chosen.

2.5 Results

Table 1 shows the results of the logistic regression analyses on all participants' pooled data. The coefficients β_{F1}, β_{F2}, and β_{dur} can be plotted to graphically represent the estimated locations of response categories in the stimulus space (Morrison, 2007), which are shown in Figure 3.

Let us briefly examine the overall response patterns in the figure. Regarding the JP responses, the relative positions of /e/ and /a/ on the β_{F1}-β_{F2} plane are as expected, since mid front /e/ should show lower β_{F1} and higher β_{F2} than low central /a/. Phonologically long /ee/ and /aa/ are proximal to their short counterparts in β_{F1} and β_{F2}, but larger in β_{dur}. This is consistent with the traditional description of Japanese long vowels as a sequence of two identical vowels at the phonological level. As for the EN responses, the

5 Participants were reminded that the pronunciation of *bass* was not /beɪs/ "low frequency sound" but /bæs/ "a type of fish" in a short practice before the EN session, where natural tokens of the four English words were used as tokens. The JP session also followed a practice with natural tokens of the four Japanese words as tokens. The Japanese and English tokens were produced by the same speakers as those in Section 2.2.

TABLE 1 Results of logistic regression analyses on all participants' data in the experiment.

Session	Vowel	α	β_{F1}	β_{F2}	β_{dur}
JP	/ee/	−6.943	−0.261	0.988	0.667
JP	/e/	−2.167	−0.309	0.749	−0.570
JP	/aa/	−2.228	0.226	−0.390	0.916
JP	/a/	1.552	0.117	−0.184	−0.840
EN	/ɛ/	−5.293	−0.386	1.404	−0.108
EN	/æ/	−3.535	0.391	0.298	0.161
EN	/ʌ/	−1.313	0.158	0.150	−0.195
EN	/ɑ/	1.735	−0.335	−0.908	0.216

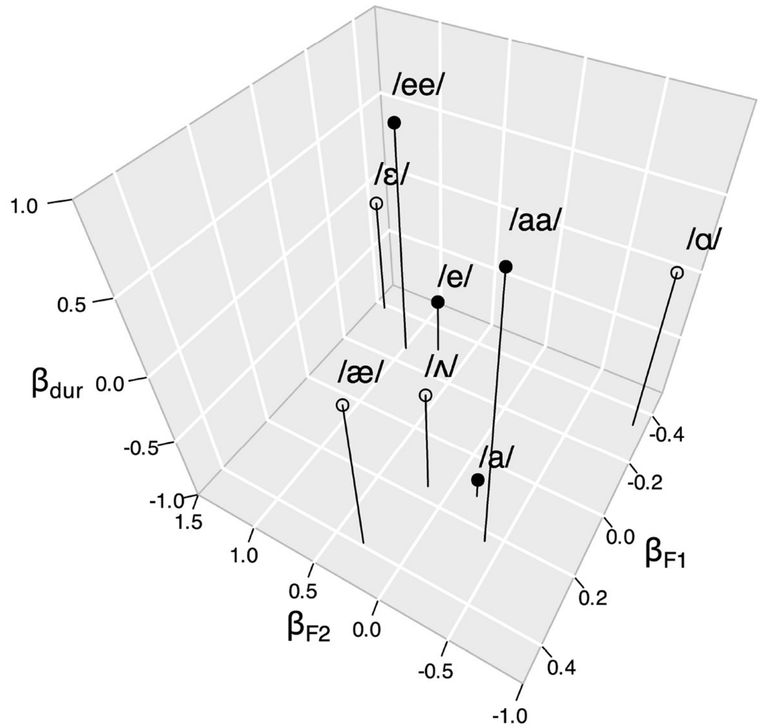


FIGURE 3
Plot of logistic regression coefficients in Table 1 (black = JP, white = EN).

relative positions of /ɛ/ and /ʌ/ are similar to those of JP /e/ and /a/, while /æ/ seems to be somewhat distant, on the β_{F1}-β_{F2} plane. Far away from all other categories is /ɑ/, with very low β_{F1} and β_{F2}. As for β_{dur}, the four EN categories seem to occupy an intermediate position between JP long and short categories.

To further investigate the response patterns, linear mixed-effects (LME) models were applied to the by-participant results of the logistic regression analyses, using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages in R. Each model is structured as follows:

$$\text{lmer}(\beta_{F1|F2|dur} \sim \text{response.category} + (1|\text{participant}) + (1|\text{session.order})) \quad (2)$$

The model tests whether the response categories differ on a stimulus-tuned coefficient (β_{F1}, β_{F2}, or β_{dur}) at a statistically significant level, controlling for the potential variability across participants and session order. Note that both JP and EN categories are included in the model, as the coefficients can in principle be compared across sessions, since the JP and EN stimuli share the same acoustic properties.

The LME model for β_{F1} with JP /a/ as the reference level showed significantly smaller estimates for EN /ɛ/ ($\beta = -1.110$, *s.d.* = 0.265, $t = -4.180$, $p < 0.001$) and /ɑ/ ($\beta = -0.561$, *s.d.* = 0.265, $t = -2.116$, $p = 0.035$), suggesting that the two EN categories are higher in perceived vowel height than the reference. The model for β_{F2} also yielded significantly larger estimates for EN /ɛ/ ($\beta = 5.391$, *s.d.* = 0.372, $t = 14.492$, $p < 0.001$) and /æ/ ($\beta = 1.248$, *s.d.* = 0.372, $t = 3.355$, $p < 0.001$), as well as a significantly smaller estimate for

EN /ɑ/ ($\beta = -0.892$, $s.d. = 0.372$, $t = -2.397$, $p = 0.017$), than the reference JP /a/. This suggests that EN /ɛ/ and /æ/ are perceptually represented as more fronted, and EN /ɑ/ as more back, than JP /a/. No significant difference was found between EN /ʌ/ and JP /a/ in either β_{F1} or β_{F2}. As for β_{dur}, all EN categories had significantly larger estimates than the reference JP /a/ ($p < 0.05$ for EN /ʌ/ and $ps < 0.001$ for /ɛ/, /æ/, and /ɑ/). An additional LME model with EN /ʌ/ as reference found significantly larger β_{dur} estimates for JP /ee/ ($\beta = 1.646$, $s.d. = 0.378$, $t = 4.359$, $p < 0.001$) and /aa/ ($\beta = 1.423$, $s.d. = 0.378$, $t = 3.768$, $p < 0.001$), but no significant difference was found for the other three EN categories. The results suggest that the four EN categories are represented with an intermediate perceptual duration between the long and short JP categories, with no significant difference between the EN categories themselves.

2.6 Interpretation

The above results can be interpreted as follows. First, AmE /ɛ/ and /ʌ/ are qualitatively assimilated to Japanese /e/ and /a/, given the similar β_{F1} and β_{F2} estimates between EN /ɛ/ and JP /e/ and between EN /ʌ/ and JP /a/, respectively. If a separate category had been formed for AmE /ɛ/, which is lower in phonetic height than Japanese /e/, then β_{F1} for EN /ɛ/ should have been larger than that for JP /e/, but this was not the case. Also, given the non-significant differences in β_{F1} and β_{F2} between EN /ʌ/ and JP /a/, it is unlikely that AmE /ʌ/ was reliably discriminated from Japanese /a/. In contrast, AmE /æ/ was most likely perceived as a separate category. Given its significantly larger β_{F2} than JP /a/, the AmE vowel may be represented as "a fronted version of /a/." While these results are consistent with previous findings, it has additionally been shown that AmE /æ/ is distinguished from Japanese /a/ by the F2 cue and not by the F1 cue.

The result for EN /ɑ/, however, was somewhat unexpected. Although AmE /ɑ/ is reported to be qualitatively assimilated to Japanese /a/ (Strange et al., 1998), the β_{F1} and β_{F2} estimates for EN /ɑ/ responses were significantly lower than for JP /a/. There are a few possible explanations for this finding. First, the learners may have associated AmE /ɑ/ with Japanese /o/ at the orthographic level, since the AmE sound is often written with "o" (e.g., *boss*, *lot*, *not*), as is the Japanese sound when written in the Roman alphabet (e.g., *bosu* /bosu/ "boss"). This possibility is particularly plausible because the participants had learned English mostly in written rather than oral form. Second, the participants may have been referring to AmE /ɔ/ rather than /ɑ/ when they chose *boss* as their response. The experimental design assumed that the vowel in *boss* is /ɑ/ because of the widespread and ongoing low back merger in many dialects of AmE (Labov et al., 2006), but some AmE speakers may still maintain the contrast and produce the word with /ɔ/, which would be perceptually assimilated to the Japanese /o/ quality (Strange et al., 1998). These two possibilities are complementary rather than mutually exclusive, and they both indicate that the very low β_{F1} and β_{F2} for the participants' *boss* responses can be attributed to Japanese /o/.

Finally, it is worth noting that the duration cue was not utilized very actively in the EN session. Judging from their intermediate β_{dur} between JP long and short categories, the AmE vowel categories appear to be unspecified in terms of phonological length. This result is consistent with Strange et al. (1998)'s finding that Japanese listeners did not show duration-based categorization when AmE vowels were presented in isolation as in the current experiment.

3 Simulation

Following the above experimental results, we now present in this section a formal computational modeling of how L1 Japanese listeners may develop a new sound category for L2 AmE /æ/ (or not for other categories) within the L2LP framework. We compare two versions of simulations using StOT, one segment- and the other feature-based, as they make divergent predictions about how L1 and L2 linguistic experience shapes listeners' perception. These predictions are compared with the experimental result to evaluate which version is more plausible. We begin by outlining the general procedure of the simulations, followed by the segmental and then by featural simulations.

3.1 General procedure

With StOT, speech perception can be modeled with a set of Optimality Theoretic, negatively formulated *cue constraints* (Escudero, 2005, 2009; Boersma and Escudero, 2008; Boersma, 2009) that modulate the mapping of acoustic cues (e.g., [F1 = 700 mel]) onto phonological representations (i.e., segmental categories or distinctive features in our case). StOT differs from regular OT in that constraints are arranged on a continuous rather than a discrete ranking scale, and constraint rankings are allowed to shift rather than being fixed. Each constraint is assigned a *ranking value* representing the stringency of the constraint (e.g., 100.0). At each time of evaluation, the ranking value is temporarily perturbed by a random value called *evaluation noise*, drawn from a normal distribution with a mean of 0 and a specified standard deviation (e.g., 2.0). The resulting value, called *selection point*, is used to evaluate the candidates. For example, if a constraint C_1 has a ranking value of 100.0 and the evaluation noise is 2.0, then the selection point for that constraint can be 100.4, 101.5, 99.3, etc. at each evaluation. Since the selection points change each time, the constraint rankings are not absolute as in regular OT (e.g., $C_1 > C_2$) but are probabilistic (e.g., C_1 with a ranking value of 100.0 will usually outrank C_2 with a ranking value of 98.0, but the latter constraint may outrank the former in some cases). This allows StOT to deal with probabilistic variation in speech perception.

The ranking values of the constraints are not determined manually, but are learned computationally from the input data through the Gradual Learning Algorithm (GLA), an error-driven algorithm for learning optimal constraint rankings in StOT (Boersma and Hayes, 2001). GLA is error-driven in that it adjusts the ranking values of relevant constraints when there is a mismatch between the output and the correct form, which the listeners are assumed to have access to via lexical knowledge and semantic context.[6] Specifically, the ranking values of the constraints that lead to the incorrect winner are increased (i.e., strengthened), while those of the constraints that would lead to the correct form are decreased (i.e., weakened). The degree to which the ranking values can change is set by a small number called *plasticity* (e.g., 1.0), which simulates the learner's current neural or cognitive plasticity. The plasticity is set to gradually decrease over time, so that learning is fast but imprecise at an early stage (infancy and childhood) and slow but precise at a later stage (adulthood). The overall scheme allows GLA to model the effects of the lexicon and age on perceptual learning.

The segmental and featural versions of the simulations use the above two computational tools, with the same parameter settings whenever possible. All constraints have an initial ranking value of 100.0, and the evaluation noise is fixed at 2.0. The plasticity is initially set to 1.0, decreasing by a factor of 0.7 per virtual year. The number of yearly input tokens was 10,000. These settings are mostly taken from previous studies, Boersma and Escudero (2008) in particular. To compare the results of the simulations with those of the experiment, we restrict the relevant auditory information provided to our virtual listeners to a range of F1 from 700 to 850 mel and a range of F2 from 1,100 to 1,400 mel, i.e., the same as the spectral stimulus space in the experiment. Duration is not included in the simulations because the target L2 AmE vowels appear to be unspecified in terms of length. Similar to the F1 and F2 steps in the experiment, the F1 and F2 ranges are divided into "bins" of equal width on the mel scale. While four bins per range would allow for a direct comparison between the experiment and the simulation, each range was assigned 16 bins for more precise modeling; as discussed in more detail in Sections 3.2 and 3.3, using 16 bins also effectively illustrates how a range of acoustic values map to certain abstract categories or features. Thus, there are 16 F1 bins with a width of 10 mel (i.e., [F1 = 700 mel], [F1 = 710 mel], ... [F1 = 850 mel]) and 16 F2 bins with a width of 20 mel ([F2 = 1,100 mel], [F2 = 11,20 mel], ... [F2 = 1,400 mel]), all of which receive a cue constraint.

The input data for training the virtual listeners are randomly generated using the parameters in Table 2. The mean formant values are taken from Nishi et al. (2008), as shown in Figure 1. The standard deviations are approximate estimates based on the formant plots in the study, as specific values are not available. Japanese /o/ is included here because it is necessary to model the perception of the AmE *boss* vowel. For simplicity, the three Japanese vowels /e/, /a/, and /o/ are assumed to occur at the same frequency (33.3%), as are the four AmE vowels /ɛ/, /æ/, /ʌ/, and /ɑ/ (25.0%). Although we are not entirely sure about the status of the AmE *boss* vowel, our virtual learners will hear both [ɑ] and [ɔ] tokens equally often (i.e., 12.5%), although there is only one target category /ɑ/ to acquire because the low back contrast is optional. In other words, the learners hear both merged and unmerged speakers but will eventually become merged listeners themselves, which we believe is a feasible scenario.

In the following two sections, we present how segmental and featural versions of virtual StOT listeners, trained with the same L1 Japanese and L2 AmE input, may develop a new category for /æ/ (and not for other AmE vowels), like the real listeners in our experiment. Each section begins with a brief illustration of cue constraints, namely cue-to-segment or cue-to-feature constraints. In line with the *Full Transfer* hypothesis of L2 acquisition (Schwartz and Sprouse, 1996), L2LP assumes that the initial state of L2 perception is a *Full Copy* of the end-state L1 grammar. Thus, we first train the perception grammar with Japanese input tokens for a total of 12 virtual years, which is copied to serve as the basis for L2 speech perception. Based on L2LP's further assumption that L2 learners have *Full Access* (Schwartz and Sprouse, 1996) to L1-like learning mechanisms, the copied perception grammar is then trained with AmE input in the same way, but with a decreased plasticity ($1.0 \times 0.7^{12} = 0.014$ at age 12, which further decreases by a factor of 0.7 per year).

3.2 Segmental simulation

3.2.1 Cue-to-segment constraints

Most previous studies aimed at formally modeling the process of L2 speech perception within L2LP (e.g., Escudero and Boersma, 2004; Boersma and Escudero, 2008; Yazawa et al., 2020) used segment-based cue constraints, such as "a value of *x* on the auditory continuum *f* should not be mapped to the phonological category *y*," which we also use here. For instance, suppose that a Japanese listener hears a vowel token with [F1 = 850 mel] and [F2 = 1,400 mel] (i.e., an [æ]-like token, cf. Figure 1) and perceives it as either /e/ or /a/. Table 3 shows how this can be modeled by a total of 4 cue constraints, each of which prohibits the perception of a segmental category (e.g., */e/) based on an acoustic cue (e.g., [F1 = 850 mel]). At the top of the leftmost column is the perceptual input, i.e., the vowel token, followed in the same column by candidates for the perceptual output, i.e., what the listener perceives given the input. In this example, the constraint "[F2 = 1,400 mel] */a/" happens to outrank the constraint "[F1 = 850 mel] */e/," making the candidate /e/ as the winner.

Note, however, that the same vowel token will not always be perceived as /e/ due to the probabilistic nature of StOT. It is possible that in some cases the constraint "[F1 = 850 mel] */e/" will outrank "[F2 = 1,400 mel] */a/," making /a/ as the alternative winner. The probability of such an evaluation is increased by GLA if and when the listener notices that the intended from should be /a/ rather than /e/ through their lexical knowledge and the semantic context (e.g., *aki* "autumn" should have been perceived instead of *eki* "station" given the conversational context). Table 4 illustrates how such learning takes place. Here, the ranking values of the constraints that led to the perception of the incorrect winner ("✓") are increased ("←"), while the ranking values of the constraints that would lead to the correct form ("✗") are decreased ("→"), by the current plasticity value. This makes it more likely that the same token will be perceived as /a/ rather than /e/ in future evaluations.

6 The distinction between lexical and semantic levels of representations goes beyond the scope of our simulations; see Boersma (2011) for a discussion.

TABLE 2 Input training parameters for the simulations.

Language	Vowel	F1 (mel)		F2 (mel)		Frequency (%)
		Mean	*S.d.*	Mean	*S.d.*	
Japanese	/e/	573	100	1,421	150	33.3
Japanese	/a/	758	100	1,086	150	33.3
Japanese	/o/	533	100	841	150	33.3
AmE	/ɛ/	721	50	1,368	100	25.0
AmE	/æ/	792	50	1,363	100	25.0
AmE	/ʌ/	724	50	1,144	100	25.0
AmE	/ɑ/ ([ɑ])	824	50	1,145	100	12.5
AmE	/ɑ/ ([ɔ])	749	50	1,037	100	12.5

TABLE 3 Example of segmental perception grammar.

[F1 = 850, F2 = 1,400]	[F2 = 1,400] */a/	[F1 = 850] */e/	[F1 = 850] */a/	[F2 = 1,400] */e/
☞ /e/		*		*
/a/	*!		*	

TABLE 4 Constraint updating in segmental grammar.

[F1 = 850, F2 = 1,400]	[F2 = 1,400] */a/	[F1 = 850] */e/	[F1 = 850] */a/	[F2 = 1,400] */e/
✗ /e/		←*		←*
✓ /a/	*!→		*→	

3.2.2 L1 perception

Our virtual segmental learner starts with a "blank" perception grammar, which has a total of 96 cue-to-segment constraints (16 F1 bins + 16 F2 bins, multiplied by three segmental categories /e/, /a/, and /o/), all ranked at the same initial value of 100.0.[7] The learner then begins to receive L1 input, namely random tokens of Japanese /e/, /a/, and /o/, which occur with equal frequency. The formant values of each vowel token is randomly determined based on the means and standard deviations in Table 2, which are then rounded to the nearest bins to be evaluated by the corresponding constraints. Whenever there is a mismatch between the perceived and intended forms, GLA updates the ranking values of the relevant cue constraints by adding or subtracting the current plasticity value.

Figure 4 shows the result of L1 learning. The grammar was tested 100 times on each combination of F1 and F2 bins. The vertical axis in the figure shows the probability of segmental categories being perceived given the F1-F2 bin combination, as calculated by logistic regression analyses as in (1) but without the duration coefficients. It can be seen that the virtual listener perceives /e/ when F1 is low and F2 is high, and /a/ when F1 is high and F2 is low, similar to the perception patterns of the real listeners in the experiment (cf. Figure 3). Note that /o/ can also be perceived when F1 and F2 are both very low.

3.2.3 L2 perception

The segmental learner is then exposed to L2 AmE data for the first time in life. Following L2LP's *Full Copying* hypothesis, the 96 cue constraints and their ranking values are copied over. Given the experimental results, the L1 vowel labels /e/, /a/, and /o/ in the copied constraints are relabeled as L2 /ɛ/, /ʌ/, and /ɑ(ɔ)/, respectively. This alone would be sufficient to explain real learners' perception of seemingly L1-assimilated vowels: /ɛ/ (= /e/) is perceived when F1 is low and and F2 is high, /ʌ/ (= /a/) when F1 is high and F2 is low, and /ɑ(ɔ)/ (= /o/) when F1 and F2 are both very low (cf. Figures 3, 4).

There is a problem, however, in that the perception of AmE /æ/ cannot be adequately modeled by mere copying. Since the grammar can only perceive three existing segmental categories, a new category for /æ/ must be manually added to the grammar. The act of adding a new category itself is not theoretically unsupported, since learners may notice a lexical distinction denoted by the vowel contrast (e.g., *bass* vs. *bus*)—perhaps due to repeated communicative errors—which motivates them to form a new phonological category. However, we encounter a puzzle here: how would the lexical distinction between *bass* and *bus* help the listeners notice the phonological contrast between /æ/ and /ʌ/ if these words sound the "same" to them? Wouldn't these words simply be represented as homophones? For example, we can see in Figure 4 that most tokens of both the *bass* vowel /æ/, which typically has high F1 and F2, and the *bus* vowel /ʌ/, which typically has low F1 and F2, are perceived as Japanese /a/. Thus, leaving open the possibility of L2 listeners manually adding a new category still begs the question of what the precise mechanism that allows the learner to do so is.

Even if we ignore this theoretical problem and add 32 new constraints (16 F1 bins + 16 F2 bins) for /æ/ (e.g., "[F1 = 1400 mel] */æ/") to the L2 grammar, we encounter another difficulty: The simulated learning outcome does not resemble actual

7 The grammar is not truly "blank" because it already knows three segmental categories onto which the cues are mapped. Boersma et al. (2003) modeled how abstract categories can emerge from phonetic and lexical input, but we chose not to include such modeling in our simulation because our focus is not on how an L1 grammar is established but on how the L1 established grammar is copied and then restructured by L2 learning.

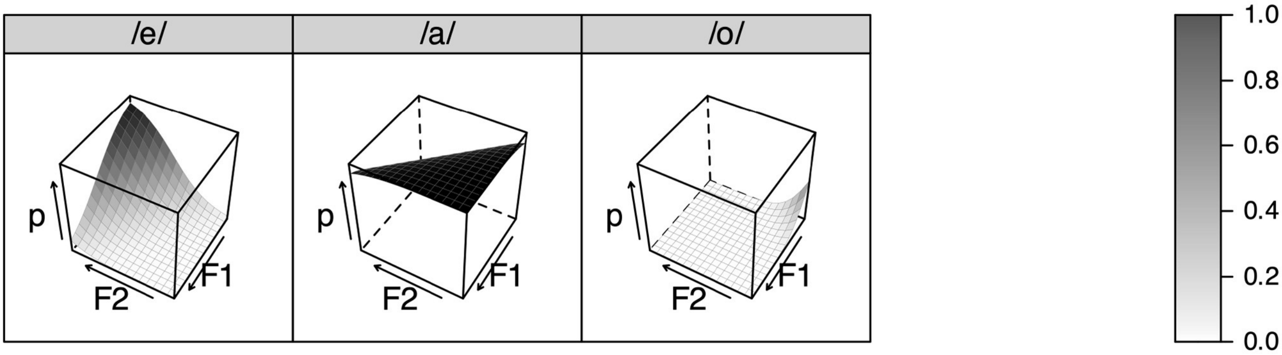


FIGURE 4
Simulated segmental perception after learning Japanese as L1 for 12 years, with no representation for /æ/.

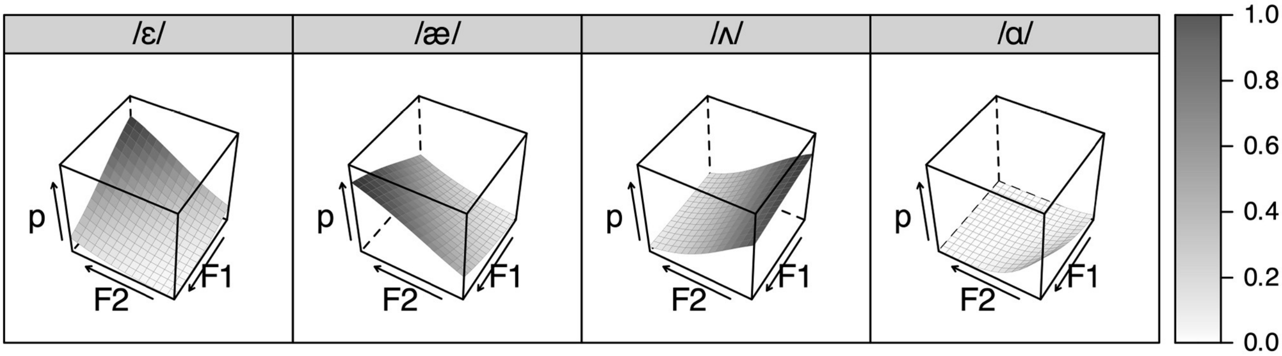


FIGURE 5
Simulated segmental perception after subsequently learning AmE as L2 for 6 years.

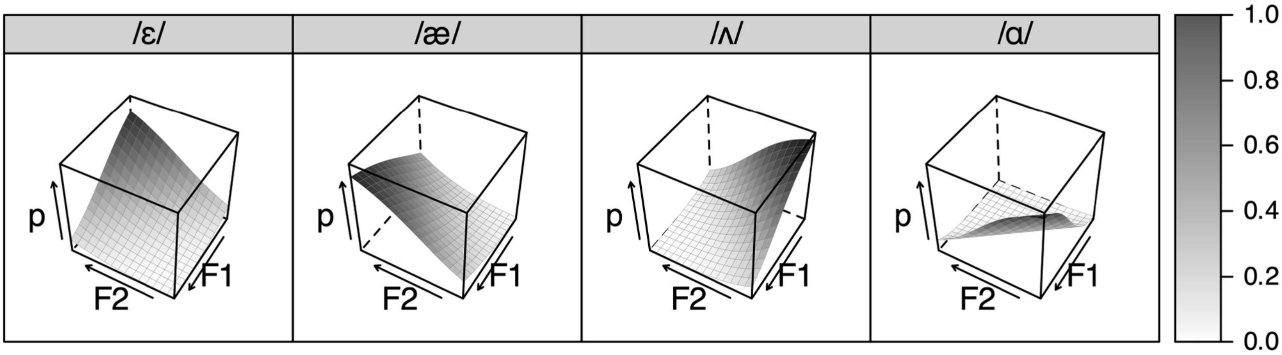


FIGURE 6
Simulated segmental perception after learning only AmE as L1 for 18 years.

perceptual behavior. In fact, the model overperforms. This can be seen in Figure 5, which shows the result after learning L2 AmE for six years. Despite the decreased plasticity, the grammar has learned to correctly perceive not only /æ/ but also other vowels /ɛ/, /ʌ/, and /ɑ/, according to the acoustic distributions of the input. This is clearly different from the real learners' perception observed in the experiment, where the latter three vowels /ɛ/, /ʌ/, and /ɑ/ were perceived as Japanese /e/, /a/, and /o/, respectively. The simulated learner therefore becomes too nativelike, showing almost identical perception patterns to those of an age-matched virtual L1 AmE listener (Figure 6). This is rather unrealistic, since very few adult L2 learners, let alone those at an intermediate level, are expected to exhibit nativelike perceptual performance.

10.3389/flang.2023.1303511

TABLE 5 Example of featural perception grammar.

[F1 = 850, F2 = 1,400]	*/mid, central/	*/low, front/	[F2 = 1,400] */central/	[F1 = 850] */mid/	[F1 = 850] */low/	[F2 = 1,400] */front/	*/low, central/	*/mid, front/
☞ /mid, front/				*		*		*
/mid, central/	*!		*	*				
/low, front/		*!			*	*		
/low, central/			*!		*		*	

TABLE 6 Constraint updating in featural grammar.

[F1 = 850, F2 = 1,400]	*/mid, central/	*/low, front/	[F2 = 1,400] */central/	[F1 = 850] */mid/	[F1 = 850] */low/	[F2 = 1,400] */front/	*/low, central/	*/mid, front/
✗ /mid, front/				←*		←*		←*
✓ /low, central/			*!→		*→		*→	

3.3 Featural simulation

3.3.1 Cue-to-feature constraints

Our featural simulation is based on Boersma and Chládková (2011), who used cue-to-feature constraints such as 'an F1 value of *x* should not be mapped to the feature /high/' and 'an F2 value of *y* should not be mapped to the feature /back/' to model the perception of vowels in different five-vowel systems including Japanese.[8] These constraints crucially differ from cue-to-segment constraints is that, in cue-to-feature constraints, the relationships between auditory continua and featural representations are non-arbitrary. That is, the auditory continuum of F1 is tied only to height features (e.g., /high/, /mid/, and /low/), and that of F2 is tied only to backness features (e.g., /front/, /central/, and /back/), unlike cue-to-segment constraints where all auditory continua in principle can be tied to any segmental category. The features are therefore "phonetically based" (i.e., they are grounded by acoustic cues) but still "phonological" (i.e., they denote phonemic contrasts and thus are distinctive) in Boersma and Chládková (2011)'s terms, which, when used in computational modeling, seem to predict real listeners' perceptual behavior better than segment-based representations (Chládková et al., 2015a). In addition to cue constraints, our featural grammar is equipped with *structural constraints* (Boersma et al., 2003; Boersma, 2011) that prohibit the co-occurrence of certain features, such as "/low/ and /front/ features should not co-occur." Structural constraints are necessary to represent the well-formedness of the perceptual output, which is relevant to the process of new L2 category formation, as we show below.

Table 5 shows how a featural Japanese grammar perceives a vowel token with [F1 = 850 mel] and [F2 = 1,400 mel] (i.e., an [æ]-like token) through two height features (/mid/ and /low/) and two backness features (/front/ and /central/). The candidates are four logical combinations of these features, two of which are well-formed in the L1 (/mid, front/ = /e/ and /low, central/ = /a/) and the other two of which are ill-formed (/mid, central/ and /low, front/). Structural constraints against ill-formed perceptual output are usually learned to be ranked very high, as is the case in the table, thus excluding the perception of /mid, central/ and /low, front/. The cue constraint "[F2 = 1,400 mel] */central/" then outranks "[F1 = 850 mel] */mid/," making /mid, front/ the winner.

Perceptual learning in the featural grammar works in the same way as in the segmental grammar, as shown in Table 6. When the listener detects a mismatch between the intended form ("✗") and the perceived form ("✓"), GLA updates the grammar by increasing the ranking values of all constraints that led to the incorrect winner ("←") and decreasing the ranking values of the constraints that would lead to the correct form ("→") by the current plasticity. Note that both cue and structural constraints are learned.

3.3.2 L1 perception

Just like the segmental learner, our featural learner starts with a "blank" perception grammar, which has 80 cue constraints (16 F1-to-height constraints for each of two height features /mid/ and /low/, and 16 F2-to-backness constraints for each of 3 backness features /front/, /central/, and /back/) as well as 6 structural constraints (two height features × three backness features), all ranked at the same initial value of 100.0.[9] The learner then begins to receive L1 input, namely randomly generated tokens of Japanese /e/ (/mid, front/), /a/ (/low, central/), and /o/ (/mid, back/), as per Table 2. The correspondence between features and categories (e.g., /a/ = /low, central/) is based on Boersma and Chládková (2011). Whenever there is a mismatch between the perceived and intended feature combinations, GLA updates the ranking values of the relevant cue and structural constraints by the current plasticity.

Figure 7 shows the result of L1 learning, tested in the same way as the segmental grammar. A notable difference from the segmental result (Figure 4) is that the featural grammar can perceive a feature combination that does not occur in the L1 input, despite the high-ranked structural constraints against such ill-formed output. For example, a token with high F1 and F2, which the segmental grammar perceived as /a/ most of the time or as /e/ otherwise, can sometimes be perceived as /low, front/, which has no segmental equivalent in Japanese. What this means is that the featural grammar may prefer to perceive a structurally ill-formed form such as /low, front/ over well-formed forms such as /low, central/ if there is sufficient cue evidence to support the evaluation. This essentially expresses the perceptual deviance of [æ] that segmental modeling fails to capture: The vowel is too /front/ to be /low, central/ (= /a/).

3.3.3 L2 perception

The featural learner then begins to learn L2 AmE. Since the initial L2 grammar is a copy of the L1 grammar, it has 80 cue constraints and 6 structural constraints with the copied ranking values. Following the experimental results, and to make the featural simulation compatible with the segmental one, we assume that L2 /ɛ/ is represented as /mid, front/, /ʌ/ as /low, central/, and /ɑ(ɔ)/ as /mid, back/ in the grammar. No additional constraint is needed to model /æ/ (/low, front/).

Figure 8 shows the result of learning L2 AmE for six years. It can be seen that the feature combination /low, front/ is much more likely to be perceived than it was in Figure 7 because the ranking value of the structural constraint "*/low, front/" has decreased. The weakening of the constraint

8 Constraints that map acoustic cues to privative features were first introduced by Boersma et al. (2003) and incorporated into L2LP by Escudero (2005). While it is possible to employ cue constraints with a binary feature such as [±long] (Hamann, 2009), we prefer privative features such as /long/ because, at least regarding the length feature, what is not "long" is not necessarily "short" but is rather unmarked (Chládková et al., 2015b). Note also that the choice of binary features in Hamann (2009) comes from a practical purpose to reduce the number of cue constraints, rather than a theoretically motivated choice.

9 This grammar is also not truly "blank" because it already knows two height and three backness features. Boersma et al. (2003, 2022) modeled how featural representations can emerge from phonetic and lexical input, which we again do not include in our simulation for the same reason as in footnote 7.

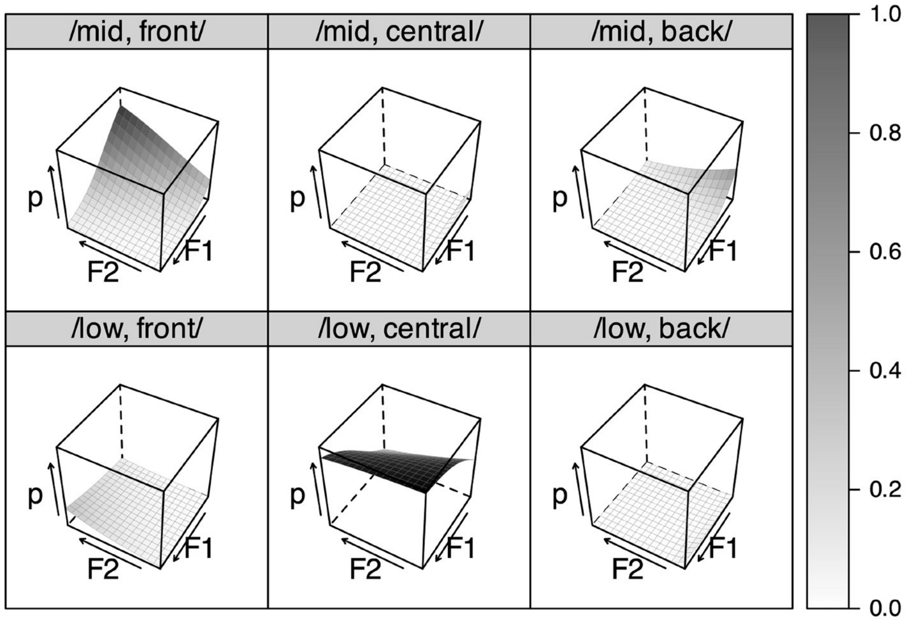


FIGURE 7
Simulated featural perception after learning Japanese as L1 for 12 years.

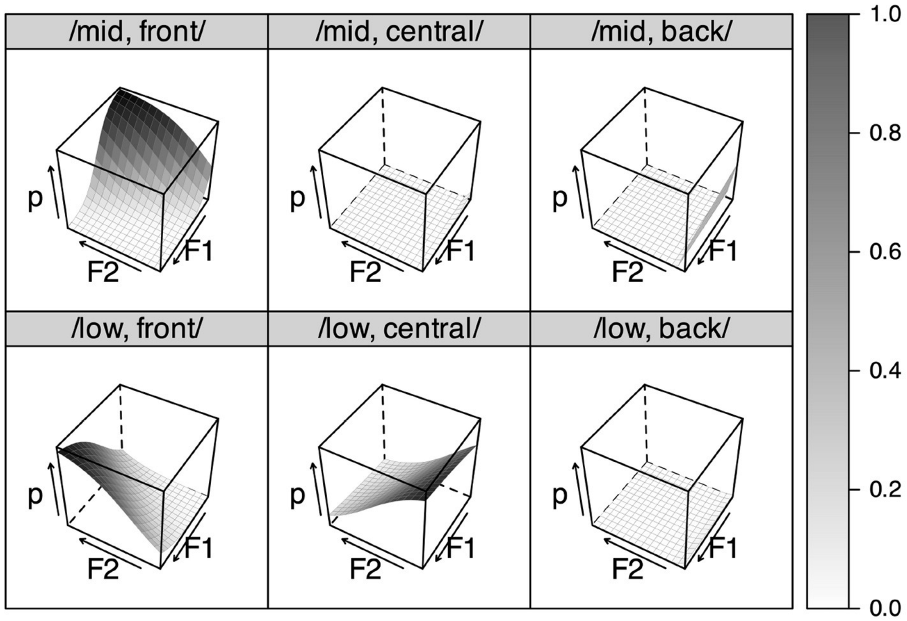


FIGURE 8
Simulated featural perception after subsequently learning AmE as L2 for 6 years.

occurred because in the L2 AmE environment, the features /low/ and /front/ often co-occur, and /low, front/ should be lexically distinguished from other feature combinations for successful communication. A new category therefore emerged from existing features by improving the well-formedness of the once ill-formed feature combination, without resorting to any L2-specific manipulation of the grammar as in the segmental modeling.

Another notable finding is that the simulated perception in Figure 8 differs from the simulated L1 AmE perception in Figure 9. One salient difference lies in /ʌ/, which was learned as /low, central/ by the learner grammar, whereas it is represented as

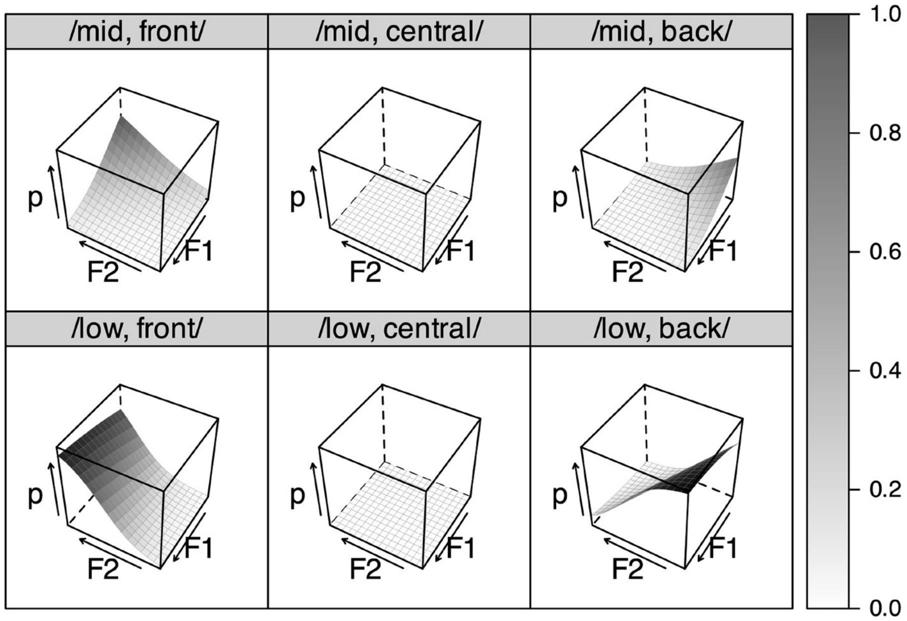


FIGURE 9
Simulated featural perception after learning only AmE as L1 for 18 years.

/mid, back/ in the native grammar.[10] The native perception is symmetrical because it reflects the production environment of AmE vowels, whereas the learner perception is asymmetrical because L2 AmE sounds are perceived through copied L1 Japanese features. The simulated learner perception actually resembles the real learners' perception, especially in the use of the F2 cue, where /ɛ/ (/mid, front/) and /æ/ (/low, front/) are perceptually more fronted, while /ɑ/ (/mid, back/) is more back, than /ʌ/ (/low, central/).

4 General discussion

This study examined how L1 Japanese learners of L2 AmE develop a new phonological representation for /æ/ by employing experimental and computational-phonological approaches. The experimental results suggested that AmE /æ/ is represented as a separate category by intermediate-level learners, distinguished from Japanese /a/ based on the F2 cue, while adjacent AmE /ɛ/, /ʌ/, and /ɑ/ are assimilated to Japanese /e/, /a/, and /o/, respectively. To explain and replicate these results with the L2LP model, segment- and feature-based versions of perceptual simulations were performed using StOT and GLA. The segmental modeling was theoretically inadequate because it failed to elucidate the mechanism for noticing the perceptual distinctness of /æ/, and was also practically implausible because the predicted overall perception patterns were too native AmE-like compared to real learners' perception. In contrast, the featural modeling explained the emergence of a new category for AmE /æ/ and the lack thereof for /ɛ/, /ʌ/, and /ɑ/ by assuming that L2 sounds are perceived through copied L1 features, i.e., */low, front/ vs. /mid, front/, /low, central/, and /mid, back/, respectively. The simulated learning outcome closely resembled real perception.

In this section, we discuss the implications of the above findings for L2LP and the other two dominant models of L2 perception, as well as directions for extending the current study in future research.

4.1 Implications for L2 perception models

4.1.1 L2LP

Our simulations have shown that, similar to the UNFAMILIAR NEW scenario in Escudero and Boersma (2004), feature-based modeling can be useful in explaining the FAMILIAR NEW scenario. While previous L2LP studies have tended to focus on the mapping of acoustic cues to segmental categories, the current study showed that representing sound categories as an integrated bundle of features can lead to more theoretically and empirically precise predictions. A salient difference observed between the segment- and feature-based simulations was in the learning outcome, which was unrealistically nativelike in the former but fairly learner-like in the latter. This is partly due to a known weakness of segmental modeling: It tends to overpredict success because GLA does not stop learning until all segmental categories are optimally perceived (unless the input data halt or the plasticity reaches zero). The featural grammar, on the other hand, remained nonnativelike because learners continued to map acoustic cues onto copied L1 features, which are organized differently from native AmE features.

10 We used the same set of feature labels in both native and learner grammars to allow for a direct comparison between them, not because we assume a universal set of features across all languages (cf. Section 4.1.1). For example, we could relabel the /mid/ feature in the native grammar as /low-mid/ and still get the same result as Figure 9.

In order for the featural learner grammar to achieve optimal AmE perception, the L1-like feature bundles (e.g., /ʌ/ = /low, central/) must be decomposed and reorganized to fit the L2 production environment (e.g., /ʌ/ = /mid, back/), perhaps with an addition of a new height or backness feature based on the FAMILIAR cue of F1 or F2 (because, after all, AmE has more vowels than Japanese). L2LP would predict that such learning is possible but challenging, as the NEW scenario is considered as more difficult than other types of learning scenarios (SIMILAR and SUBSET scenarios). It remains to be revealed, however, whether the reorganization and addition of features based on FAMILIAR cues is less difficult than the establishment of a novel feature based on UNFAMILIAR cues. According to Chládková et al. (2022), perceptual boundary shift in a SIMILAR scenario, which involves only FAMILIAR acoustic cues, is easier than creating new perceptual mappings of an UNFAMILIAR cue in a NEW scenario. Research on Japanese listeners' perception of English /ɹ/-/l/ contrast have also found that they rely persistently on unreliable but FAMILIAR acoustic cues such as F2 and duration, rather than the important but UNFAMILIAR cue of F3, to identify the NEW sound representation of /ɹ/ or, in featural terms, /rhotic/ (Iverson et al., 2003; Saito and van Poeteren, 2018; Shinohara and Iverson, 2021). It is thus predicted that the FAMILIAR NEW scenario is easier than the UNFAMILIAR NEW scenario, although more modeling and empirical testing seem necessary to verify this hypothesis (cf. Yazawa, in press).

One important point about the feature-based modeling is that the relationship between acoustic cues and phonological features is considered to be language-specific and relative. For example, while the F1 cue may be mapped to three height features (/high/, /mid/, and /low) in many languages, in some languages such as Portuguese and Italian there are four target heights (/high/, /mid-high/, /mid-low/, and /low/) and in others such as Arabic and Quechua there are only two (/high/ and /low/). Also, even if two languages share the "same" set of height features, what is perceived as /high/ in one language may not be also perceived as also /high/ in another, since the actual F1 values of high vowels varies across languages or even language varieties (Chládková and Escudero, 2012). The cue-to-feature mapping patterns are also relative within a language or language variety. This is perhaps best demonstrated by Benders et al. (2012), who showed that Spanish listeners' perceptual boundary between /i/ and /e/ (i.e., /high/ and /mid/ front vowels) was shifted by the acoustic range of the stimuli and the number of available response categories. Specifically, listeners perceived a vowel token with [F1 = 410 Hz] as /e/ when the F1 value was relatively high within the stimulus range (281–410 Hz), whereas the same token was perceived as /i/ when the F1 value was relatively low within the stimulus range (410–553 Hz). The perceptual boundary also shifted when additional response categories /a/, /o/, and /u/ were made available. These results suggest that the perception of the height feature is not determined by the absolute F1 value but rather depends largely on what other features must be considered together for the task at hand, providing useful insights into why perceptual behavior seems to vary depending on the experimental design. The above discussion has an important implication for the so-called perceptual "warping" or "magnetism" of nonnative categories to native ones (Kuhl et al., 2008), which is closely related to cross-linguistic categorical assimilation. For example, it was mentioned in Section 1 that the perceived goodness of AmE /ɛ/ as Japanese /e/ was fair despite their seemingly large spectral distance. This may be because perceived vowel height and backness are defined relatively within each language rather than between two languages. That is, Japanese listeners may perceive Japanese /e/ as /mid, front/ relative to other Japanese vowels and, in the same way, AmE /ɛ/ as /mid, front/ relative to other AmE vowels, though their judgements may depend on the task at hand. Thus, a direct comparison of raw F1 values between the two languages may not be very meaningful. The proposed language-specific feature identification is compatible with L2LP's *Full Copying* hypothesis, which claims that L1 and L2 speech perception are handled by separate grammars.

4.1.2 SLM

While the current study aimed to explain the process of new category formation within the framework of L2LP, the results also have useful implications for SLM. Specifically, it can be proposed that cross-linguistic categorical dissimilarity is defined as a mismatch of existing L1 features (e.g., */low, front/), with the caveat that the actual phonetic property of a feature is language-specific and relative as discussed above. This proposal is actually compatible with one of the hypotheses (H6) of the original SLM (Flege, 1995): "The phonetic category established for L2 sounds by a bilingual may differ from a monolingual's if: [...] the bilingual's representation is based on different features, or feature weights, than a monolingual's." In fact, many of the components of our featural modeling are compatible with the original SLM, which was full of fruitful insights into the feature-based approach. For example, it was noted in Flege (1995, p. 267) that "the features used to distinguish L1 sounds can probably not be freely recombined to produce new L2 sounds," which is essentially what our structural constraints modeled. It was also noted on the same page that "[s]ome production difficulties may arise because features used in the L2 are not used in the L1," later formalized in the model as the "feature" hypothesis (McAllister et al., 2002), which is closely related to the UNFAMILIAR NEW scenario of L2LP. Another point on the same page was that "[t]he phenomenon of "differential substitution" shows that we need recourse to more than just a simple listing of features used in the L1 and L2 to explain certain L2 production errors," meaning that L1-L2 segmental substitution patterns can vary even when two L1s share the "same" feature sets, which brings us back to the aforementioned caveat about feature relativity and forward to Section 4.2 where we discuss feature hierarchy and integration. The non-absolute nature of features also relates to yet another point on the next page of Flege (1995): "features may be evaluated differently as a function of position in the syllable."

Much of this discussion, however, did not find its way into the revised SLM (Flege and Bohn, 2021), in which the "feature" hypothesis was replaced by the "full access" hypothesis. According to the new hypothesis, L2 learners can gain unrestricted access to features not used in their L1, which aligns more closely with our segment-based modeling that overpredicted success. It is also worth noting that the term "feature" is used almost interchangeably with "cue" in the revised model, although we hope to have shown through the comparison of cue-to-segment and cue-to-feature modeling that this should not be the case. Given the compatibility of our feature-based modeling with the principles of the original SLM, perhaps separating the notions of features and cues may benefit the revised model, especially to address the perrenial issue of how L1-L2 categorical similarity should be defined. To this end, incorporating different levels of abstraction as in L2LP and PAM may be in order (cf. Section 4.1.3 below).

4.1.3 PAM

The implication of feature-based modeling for PAM is similar to that for SLM: Cross-linguistic dissimilarity can be defined as featural discrepancy. However, the implication is unique for PAM because, unlike SLM and L2LP which model speech perception as the abstraction of acoustic cues into sound representations (be them segments or features), PAM subscribes to a direct realist view that listeners directly perceive articulatory gestures of the speaker. PAM also distinguishes between phonetic and phonological levels of representations (like L2LP, in a broad sense), whereas SLM defines sound categories strictly at the phonetic or allophonic level. These differences in theoretical assumption raise a crucial question about what features really are: Are they articulatory or auditory, and phonetic or phonological? As mentioned earlier, the current study assumed what Boersma and Chládková (2011) called "phonetically based phonological features," which can be learned from perceptual input without any articulatory knowledge because perception is considered to precede production in L1 and L2 development (Escudero, 2005, 2007; Kuhl et al., 2008). The Bidirectional Phonology and Phonetics (BiPhon) framework also proposes that these auditorily learned features are used in both perception and production (Boersma, 2011), although it remains to be seen whether articulatory features are really unnecessary to account for L1 and L2 production patterns. A related topic is whether and how the features used for segmental categorization are also relevant for higher-level phonological processes, both in perception and production. The vast body of previous OT-based research provides a useful ground for testing this, because all of the components of our feature-based StOT modeling are generally compatible with the traditional OT framework.[11]

4.2 Future directions

Having discussed the theoretical implications of the feature-based approach, we now address how the current modeling can be practically extended to improve its adequacy in future research. First, as for the acoustic cues, we chose not to include duration because the participants in our experiment do not seem to have used it to categorize the target L2 AmE vowels, but it remains to be modeled why L1 listeners of Japanese with phonological vowel length would show such perception patterns. This can actually be a task effect, since duration-based categorization of AmE vowels into Japanese long and short ones was only observed when AmE vowels were embedded within a carrier sentence (Strange et al., 1998), i.e., when the target vowel duration could be compared to the duration of other vowels in the carrier sentence (cf. within-language feature relativity in Section 4.1.1). Thus, the modeling may need to incorporate some kind of temporal normalization to explain the potential task dependency. Escudero and Bion (2007) modeled formant normalization and speech perception as sequential processes by first applying an external algorithm (e.g., Z-normalization) to the raw acoustic data and then feeding the normalized input to the StOT grammar, which is a promising approach for temporal normalization as well. Second, as for the perceptual output, all target features were assumed to have equal status, which is perhaps overly simplistic. Flege (1995, p. 268) noted that "[c]ertain features may enjoy an advantage over others," and Archibald (2023) recently proposed that cross-linguistic differential substitution patterns can be explained by a contrastive hierarchy of features across languages. Greenberg and Christiansen (2019) also suggested that features are processed in a stepwise manner (e.g., voicing → manner → place) rather than all at once during online speech perception. If features are hierarchically organized and processed to be ultimately integrated into higher-level representations (Boersma et al., 2003; Escudero, 2005; Yazawa, 2020), then the perception of height and backness features may also need to be evaluated sequentially rather than simultaneously, with perhaps the height feature being processed before the backness feature (Balas, 2018). This stepwise processing can be formally modeled by assigning *stratum indices* (van Leussen and Escudero, 2015) besides the ranking values to the StOT constraints, and ordering the constraints first by stratum and then by ranking value (or selection point, to be precise) at each evaluation. Finally, in order to fully model the observed experimental results, orthographic influences must be included. Our simulations assumed a link between AmE /ɑ/ and Japanese /o/ representations without specifying its nature, but it seems likely that this link was established at the visual rather than the auditory level in real learners. Previous L2LP studies have already explored the orthographic influences on speech perception (Escudero and Wanrooij, 2010; Escudero et al., 2014; Escudero, 2015), but how exactly orthography fits into the model is yet to be seen. Hamann and Colombo (2017) proposed modeling orthographic and perceptual borrowing of English words into Italian by using orthographic constraints such as "assign a violation mark to the grapheme <t> that is not mapped onto the phonological form /t/" along with cue constraints, which is readily applicable to StOT-based modeling of L2 audiovisual perception. An ongoing collaboration aims at achieving this goal.

We also believe that further empirical testing is needed to complement the formal computational modeling. One limitation of the current experiment, or behavioral experiments in general, is that features cannot be directly observed in participant responses. To overcome this weakness, neural studies as in Scharinger et al. (2011) or Mesgarani et al. (2014) would be helpful. Of particular interest is how the locations and magnitudes of neural responses to auditory stimuli, which have been shown to be feature-based in native perception, are mediated by L1-L2 perceptual dissimilarity and the listeners' L2 proficiency level. Such investigations, combined with formal analyses, are necessary to provide a more comprehensive understanding of the mechanism of new L2 category formation, since theoretical models need empirical support whilst empirical data need theoretical interpretation.

11 Traditional OT grammars can be seen as a special case of StOT grammars, with integer ranking values and zero evaluation noise.

5 Conclusion

This study proposed that perceived (dis)similarity between L1 and L2 sounds, which is considered crucial for the process of new L2 category formation but has long remained elusive, can be better defined by assuming feature-level representations as the fundamental unit of perception, rather than segmental categories. Through our formal modeling based on L2LP and StOT, we argued that an L2 sound (e.g., AmE /æ/) whose FAMILIAR acoustic cues (e.g., F1 and F2) map to a bundle of L1 features that is structurally ill-formed (e.g., */low, front/ in Japanese) is perceived as deviant and thus is subject to category formation, whereas an L2 sound (e.g., AmE /ɛ/, /ʌ/, and /ɑ/) whose cues map to a well-formed L1 feature bundle (e.g., /mid, front/, /low, central/, and /mid, back/ in Japanese) is prone to assimilation, regardless of the ostensible acoustic distance between L1 and L2 segmental categories. The proposed feature-based modeling was consistent with our experimental results, where real L1 Japanese listeners seem to have established a distinct representation for L2 AmE /æ/ but not for /ɛ/, /ʌ/, and /ɑ/, which the segment-based modeling failed to predict and replicate. While feature-based approaches to L2 learning are still scarce compared to the vast literature on segment-based approaches, perhaps because the intangible nature of features cannot be captured without a computational platform that is currently only available to L2LP, the benefits of adopting and extending the approach are expected to go beyond the model (e.g., SLM and PAM) and beyond the current learning scenario (i.e., other sound contrasts in different language combinations), the pursuit of which should ultimately help deepen our understanding of how L2 speech acquisition proceeds as a whole.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Academic Research Ethical Review Committee, Waseda University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. JW: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Validation, Writing – original – draft. MK: Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft. PE: Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Validation, Writing – original draft.

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. KY and MK's work was funded by JSPS Grant-in-Aid for Scientific Research (grant number: 21H00533). JW's work was funded by Samsung Electronics Co., Ltd. (grant number: A0342-20220008); the funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication. PE's work was funded by ARC Future Fellowship grant (grant number: FT160100514).

Acknowledgments

The authors thank Mikey Elmers for volunteering to provide his voice for the AmE stimuli.

Conflict of interest

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

Publisher's note

References

Archibald, J. (2023). Differential substitution: a contrastive hierarchy account. *Front. Lang. Sci.* 2, 1–13. doi: 10.3389/flang.2023.1242905

Balas, A. (2018). *Non-Native Vowel Perception: The Interplay of Categories and Features*. Poznań: Wydawnictwo Naukowe UAM.

Bates, D., Mächler, M., Bolker, B. M., and Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Benders, T., Escudero, P., and Sjerps, M. J. (2012). The interrelation between acoustic context effects and available response categories in speech sound categorization. *J. Acoust. Soc. Am.* 131, 3079–3087. doi: 10.1121/1.3688512

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.

Best, C. T., and Tyler, M. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, ed O. S. Bohn, and M. J. Munro (Amsterdam; Philadelphia, PA: John Benjamins), 13–34.

Boersma, P. (1998). *Functional Phonology: Formalizing the Interactions between Articulatory and Perceptual Drives* (Ph.D. thesis), University of Amsterdam. The Hague: Holland Academic Graphics.

Boersma, P. (2009). "Cue constraints and their interactions in phonological perception and production," in *Phonology in Perception*, eds P. Boersma, and S. Hamann (Berlin: De Gruyter), 55–110.

Boersma, P. (2011). "A programme for bidirectional phonology and phonetics and their acquisition and evolution," in *Bidirectional Optimality Theory*, eds A. Benz, and J. Mattausch (Amsterdam: John Benjamins), 33–72.

Boersma, P., and Chládková, K. (2011). "Asymmetries between speech perception and production reveal phonological structure," in *Proceedings of the 17th International Congress of Phonetic Sciences*, eds W.-S. Lee, and E. Zee (Hong Kong: The University of Hong Kong), 328–331.

Boersma, P., Chládková, K., and Benders, T. (2022). Phonological features emerge substance-freely from the phonetics and the morphology. *Can. J. Linguist.* 67, 611–669. doi: 10.1017/cnj.2022.39

Boersma, P., and Escudero, P. (2008). "Learning to perceive a smaller L2 vowel inventory: an Optimality Theory account," in *Contrast in Phonology: Theory, Perception, Acquisition*, eds P. Avery, E. Dresher, and K. Rice (Berlin: de Gruyter), 271–302.

Boersma, P., Escudero, P., and Hayes, R. (2003). "Learning abstract phonological from auditory phonetic categories: an integrated model for the acquisition of language-specific sound categories," in *Proceedings of the 15th International Congress of Phonetic Sciences*, eds M. J. Solé, D. Recasens, and J. Romero (Barcelona), 1013–1016.

Boersma, P., and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguist. Inq.* 32, 45–86. doi: 10.1162/002438901554586

Boersma, P., and Weenink, D. (2023). *Praat: Doing Phonetics by Computer*. Amsterdam: University of Amsterdam.

Chen, J., and Chang, H. (2022). Sketching the landscape of speech perception research (2000–2020): a bibliometric study. *Front. Psychol.* 13, 1–14. doi: 10.3389/fpsyg.2022.822241

Chládková, K., Boersma, P., and Benders, T. (2015a). "The perceptual basis of the feature vowel height," in *Proceedings of the 18th International Congress of Phonetic Sciences*, ed The Scottish Consortium for ICPhS 2015 (Glasgow: The University of Glasgow), 711.

Chládková, K., Boersma, P., and Escudero, P. (2022). Unattended distributional training can shift phoneme boundaries. *Bilingual. Lang. Cognit.* 25, 827–840. doi: 10.1017/S1366728922000086

Chládková, K., and Escudero, P. (2012). Comparing vowel perception and production in Spanish and Portuguese: European versus Latin American dialects. *J. Acoust. Soc. Am.* 131, EL119–EL125. doi: 10.1121/1.3674991

Chládková, K., Escudero, P., and Lipski, S. C. (2015b). When "AA" is long but "A" is not short: speakers who distinguish short and long vowels in production do not necessarily encode a short long contrast in their phonological lexicon. *Front. Psychol.* 6, 1–8. doi: 10.3389/fpsyg.2015.00438

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization* (PhD thesis). Utercht University. Amsterdam: LOT.

Escudero, P. (2007). "Second-language phonology: the role of perception," in *Phonology in Context*, ed M. C. Pennington (London: Palgrave Macmillan), 109–134.

Escudero, P. (2009). "The linguistic perception of SIMILAR L2 sounds," in *Phonology in Perception*, eds P. Boersma, and S. Hamann (Berlin: de Gruyter), 151–190.

Escudero, P. (2015). Orthography plays a limited role when learning the phonological forms of new words: the case of Spanish and English learners of novel Dutch words. *Appl. Psycholinguist.* 36, 7–22. doi: 10.1017/S014271641400040X

Escudero, P., and Bion, R. (2007). "Modeling vowel normalization and sound perception as sequential processes," in *Proceedings of the 16th International Congress of Phonetic Sciences*, eds J. Trouvain, and W. J. Barry (Saarbrücken: Saarland University), 1413–1416.

Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Stud. Sec. Lang. Acquisit.* 26, 551–585. doi: 10.1017/S0272263104040021

Escudero, P., Kastelein, J., Weiand, K., and van Son, R. J. J. H. (2007). "Formal modelling of L1 and L2 perceptual learning: computational linguistics versus machine learning," in *Proceedings of the 8th Annual Conference of the International Speech Communication Association* (Antwerp: International Speech Communication Association), 1889–1892.

Escudero, P., Simon, E., and Mulak, K. E. (2014). Learning words in a new language: orthography doesn't always help. *Biling. Lang. Cogn.* 17, 384–395. doi: 10.1017/S1366728913000436

Escudero, P., and Wanrooij, K. (2010). The effect of L1 orthography on non-native vowel perception. *Lang. Speech* 53, 343–365. doi: 10.1177/0023830910371447

Escudero, P., and Yazawa, K. (in press). "The second language linguistic perception model (L2LP)," in *The Cambridge Handbook of Bilingual Phonetics and Phonology*, ed M. Amengual (Cambridge: Cambridge University Press).

Flege, J. E. (1995). "Second language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 233–277.

Flege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83.

Greenberg, S., and Christiansen, T. U. (2019). The perceptual flow of phonetic information. *Attent. Percept. Psychophys.* 81, 884–896. doi: 10.3758/s13414-019-01666-y

Hamann, S. (2009). "Variation in the perception of an L2 contrast: a combined phonetic and phonological account," in *Variation and Gradience in Phonetics and Phonology*, eds F. Kügler, C. Féry, and R. van de Vijver (Berlin: De Gruyter), 71–98.

Hamann, S., and Colombo, I. E. (2017). A formal account of the interaction of orthography and perception. *Nat. Lang. Linguist. Theory* 35, 683–714. doi: 10.1007/s11049-017-9362-3

Hirata, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *J. Acoust. Soc. Am.* 116, 2384–2394. doi: 10.1121/1.1783351

Hirata, Y. (2017). "Second language learners' production of geminate consonants in Japanese," in *The Phonetics and Phonology of Geminate Consonants*, ed H. Kubozono (Oxford: Oxford University Press), 163–184.

Hisagi, M., Higby, E., Zandona, M., Kent, J., Castillo, D., Davidovich, I., et al. (2021). "Perceptual discrimination measure of non-native phoneme perception in early and late Spanish-English & Japanese-English bilinguals," in *Proceedings of Meetings on Acoustics, Vol. 42* (The Acoustical Society of America), 1–13.

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

Kawahara, H. (2006). STRAIGHT, exploitation of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds. *Acoust. Sci. Technol.* 27, 349–353. doi: 10.1250/ast.27.349

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philos. Transact. R. Soc. B Biol. Sci.* 363, 979–1000. doi: 10.1098/rstb.2007.2154

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Labov, W., Ash, S., and Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology and Sound Change*. Berlin: De Gruyter.

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., and Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Appl. Psycholinguist.* 26, 227–249. doi: 10.1017/S0142716405050150

McAllister, R., Flege, J. E., and Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *J. Phon.* 30, 229–258. doi: 10.1006/jpho.2002.0174

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Morrison, G. S. (2007). "Logistic regression modelling for first-and second-language perception data," in *Segmental and Prosodic Issues in Romance Phonology*, eds M. J. Solé, P. Prieto, J. Mascaró, and M. J. Sol? (Amsterdam: John Benjamins), 219–236.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., and Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of Japanese and American English vowels. *J. Acoust. Soc. Am.* 124, 576–588. doi: 10.1121/1.2931949

Pajak, B., and Levy, R. (2014). The role of abstraction in non-native speech perception. *J. Phon.* 46, 147–160. doi: 10.1016/j.wocn.2014.07.001

Prince, A., and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2. New Jersey.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna.

Saito, K., and van Poeteren, K. (2018). The perception-production link revisited: the case of Japanese learners' English /ɹ/ performance. *Int. J. Appl. Linguist.* 28, 3–17. doi: 10.1111/ijal.12175

Scharinger, M., Idsardi, W. J., and Poe, S. (2011). A comprehensive three-dimensional cortical map of vowel space. *J. Cogn. Neurosci.* 23, 3972–3982. doi: 10.1162/jocn_a_00056

Schwartz, B. D., and Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Sec. Lang. Res.* 12, 40–72. doi: 10.1177/026765839601200103

Shafer, V. L., Kresh, S., Ito, K., Hisagi, M., Vidal, N., Higby, E., et al. (2021). The neural timecourse of American English vowel discrimination by Japanese, Russian and Spanish second-language learners of English. *Biling. Lang. Cogn.* 24, 642–655. doi: 10.1017/S1366728921000201

Shaw, J. A., and Kawahara, S. (2017). The lingual articulation of devoiced /u/ in Tokyo Japanese. *J. Phon.* 66, 100–119. doi: 10.1016/j.wocn.2017.09.007

Shinohara, Y., Han, C., and Hestvik, A. (2019). "Effects of perceptual assimilation: the perception of English /æ/, /ʌ/, and /ɑ/ by Japanese speakers," in *Proceedings of the 19th International Congress of Phonetic Sciences*, eds S. Calhoun, P. Escudero, M. Tabain, and P. Warren (Melbourne, VIC. Australasian Speech Science and Technology Association Inc.), 2344–2348.

Shinohara, Y., Han, C., and Hestvik, A. (2022). Discriminability and prototypicality of nonnative vowels. *Stud. Second Lang. Acquis.* 44, 1260–1278. doi: 10.1017/S0272263121000978

Shinohara, Y., and Iverson, P. (2021). The effect of age on English /r/-/l/ perceptual training outcomes for Japanese speakers. *J. Phon.* 89, 1–24. doi: 10.1016/j.wocn.2021.101108

Strange, W., Akahane-Yamada, R., Kubo, R., Trent-Brown, S. A., Nishi, K., and Jenkins, J. J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *J. Phon.* 26, 311–344. doi: 10.1006/jpho.1998.0078

Sugimoto, J., and Uchida, Y. (2020). English phonetics and teacher training: designing a phonetics course for Japanese preservice teachers. *J. Phonet. Soc. Jpn* 24, 22–35. doi: 10.24467/onseikenkyu.24.0_22

Tsukada, K. (2012). Comparison of native versus nonnative perception of vowel length contrasts in Arabic and Japanese. *Appl. Psycholinguist.* 33, 501–516. doi: 10.1017/S0142716411000452

Tsukada, K., Cox, F., Hajek, J., and Hirata, Y. (2018). Non-native Japanese learners' perception of consonant length in Japanese and Italian. *Sec. Lang. Res.* 34, 179–200. doi: 10.1177/0267658317719494

van Leussen, J.-W., and Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Front. Psychol.* 6, 1–12. doi: 10.3389/fpsyg.2015.01000

Whang, J., and Yazawa, K. (2023). Modeling a phonotactic approach to segment recovery: the case of Japanese high vowels. *Stud. Phonet. Phonol. Morphol.* 29, 271–295. doi: 10.17959/sppm.2023.29.2.271

Yazawa, K. (2020). *Testing Second Language Linguistic Perception : A Case Study of Japanese, American English, and Australian English Vowels* (Ph.D. thesis). Waseda University, Tokyo.

Yazawa, K. (in press). NEW sounds can be easier to learn than SIMILAR sounds, but only when acoustic cues are FAMILIAR. *Tsukuba Eng. Stud.* 42.

Yazawa, K., Whang, J., and Escudero, P. (2023). Australian English listeners' perception of Japanese vowel length reveals underlying phonological knowledge. *Front. Psychol.* 14, 1–14. doi: 10.3389/fpsyg.2023.1122471

Yazawa, K., Whang, J., Kondo, M., and Escudero, P. (2020). Language-dependent cue weighting: an investigation of perception modes in L2 learning. *Sec. Lang. Res.* 36, 557–581. doi: 10.1177/0267658319832645

Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 20 December 2023
DOI 10.3389/flang.2023.1286084

Check for updates

OPEN ACCESS

EDITED BY
John Archibald,
University of Victoria, Canada

REVIEWED BY
Yasaman Rafat,
Western University, Canada
Jennifer Mah,
Consultant, Calgary, AB, Canada

*CORRESPONDENCE
Paul John
paul.john@uqtr.ca

RECEIVED 30 August 2023
ACCEPTED 23 November 2023
PUBLISHED 20 December 2023

CITATION
John P and Rigoulot S (2023) On the representation of /h/ by Quebec francophone learners of English. *Front. Lang. Sci.* 2:1286084. doi: 10.3389/flang.2023.1286084

On the representation of /h/ by Quebec francophone learners of English

Paul John[1]* and Simon Rigoulot[2]

[1]CogNAC Research Group, Department of Modern Languages and Translation, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada, [2]CogNAC Research Group, Department of Psychology, Université du Québec à Trois-Rivières, Trois-Rivières, QC, Canada

The current study investigates whether some of the variation in h-production observed among Quebec francophone (QF) learners of English could follow from their at times assimilating /h/ to /ʁ/. In earlier research, we attributed variation exclusively to QFs developing an approximate ("fuzzy" or "murky") representation of /h/ that is not fully reliable as a base for h-perception and production. Nonetheless, two previous studies observed via event-related potentials differences in QF perceptual ability, which may follow from the quality of the vowel used in the stimuli: /ɑ/ vs. /ʌ/ (detection vs. no detection of /h/). Before the vowel /ɑ/, /h/ exhibits phonetic properties that may allow it to be assimilated to and thus underlyingly represented as /ʁ/. If /h/ is at times subject to approximate representation (e.g., before /ʌ/) and at others captured as /ʁ/ (before /ɑ/), we would expect production of /h/ to reflect this representational distinction, with greater accuracy rates in items containing /ɑ/. Two-way ANOVAs and paired Bayesian *t*-tests on the reading-aloud data of 27 QFs, however, reveal no difference in h-production according to vowel type. We address the consequences of our findings, discussing notably why QFs have such enduring difficulty acquiring /h/ despite the feature [spread glottis] being available in their representational repertoire. We propose the presence of a Laryngeal Input Constraint that renders representations containing only a laryngeal feature highly marked. We also consider the possibility that, rather than having overcome this constraint, some highly advanced learners are "phonological zombies": these learners become so adept at employing approximate representations in perception and production that they are indistinguishable from speakers with bona fide phonemic representations.

KEYWORDS

L2 phonological acquisition, perceptual assimilation, approximate representations, variation, h-deletion, input constraints, phonological zombies

1 Introduction

Motivated by the findings of two earlier studies on the perception of /h/ (White et al., 2015; Mah et al., 2016), the study presented here investigates the production of /h/ before the vowels /ɑ/ (*hot*) and /ʌ/ (*hut*) by Quebec francophone (QF) learners of English. Francophones, whether from Quebec or elsewhere, struggle to produce /h/, the tendency being to delete this non-native phoneme (*hair* → _*air*) or even to epenthesize it before vowel-initial forms (*ankle* → [h]*ankle*) (John and Cardoso, 2009; John and Frasnelli, 2022). H-deletion constitutes the basic QF pronunciation error, instantiated notably in loanwords to Quebec French (*hotdog* → _*otdog*; Paradis and Lebel, 1994), with loanwords corresponding to a kind of "ground zero" for L2 acquisition. H-epenthesis is a form of qualitative hypercorrection (Janda and Auger, 1992), along the lines of intrusive-r in English (Halle and Idsardi, 1997; Orgun, 2002). To narrow our scope, we focus here on the phenomenon of h-deletion.

Our position is that, at its source, h-deletion is not due to low-level articulatory difficulty nor to a phonological process that removes underlying /h/ from surface forms (e.g., resulting from constraints on output as in the Emergence of the Unmarked: Broselow et al., 1998). Instead, h-deletion follows from perceptual and representational problems. Under the perceptual reorganization that accompanies first language (L1) acquisition (Strange and Shafer, 2008), second language (L2) learners "redeploy L1 phonological knowledge" (Archibald, 2005), typically perceiving and representing novel phonemes according to L1 categories (a process referred to as "perceptual assimilation" by Best and Tyler, 2007, and auditory "equivalence classification" by Flege and Bohn, 2021). Perceptual assimilation accounts for the widespread phenomenon of substitution in L2 speech. For example, Russian learners of English realize /h/ as the L1 voiceless velar fricative /x/, and Spanish learners, depending on their variant of L1 Spanish (i.e., variants without /h/), realize /h/ as velar /x/ or uvular /χ/; similarly, QFs realize English /θ ð/ as /t d/ (*think that* → [t]*ink* [d]*at*) (Brannen, 2011). The process of h-deletion, however, suggests that QFs do not assimilate English /h/ to an L1 category. Apparently, no L1 phoneme is sufficiently similar to /h/ for assimilation to take place; instead, QFs fail to detect /h/ in speech output and consequently leave the segment out of underlying representations (URs). As a result, /h/-vowel minimal pairs such as *hair-air* are represented as homophonous /ɛr/, and the error of h-deletion in fact constitutes an accurate realization of the stored form.

The situation is, however, more complicated than the above scenario implies. First, QFs typically exhibit variable h-production rather than categorical deletion in English, which is hard to reconcile with the absence of /h/ in URs. If /h/ is missing from lexical entries, how do learners generate [h] in output at all? Indeed, learners manage to generate higher rates of [h] in items that should contain it (correct h-production) vs. items that should not contain it (hypercorrect h-epenthesis) (John and Cardoso, 2009). This suggests that learners must somehow lexically mark the distinction between h-ful and h-less items. One possibility is that they develop approximate (i.e., "fuzzy," as in Darcy et al., 2013; or "murky," as in John, 2006) representations for /h/ using non-linguistic diacritics rather than actual distinctive features (John and Frasnelli, 2022). That is, the minimal pair *hair-air* may be distinguished via an *ad hoc* marking that we capture graphically with a superscript question mark: /$^{?}$ɛr/ vs. /ɛr/. Learners are thought to develop approximate representations when they become aware (e.g., due to feedback) that their output diverges from that of native speakers (NSs). That is, when learners recognize their own tendency to delete /h/ and yet are at a loss to match this elusive speech sound to a phoneme category, they compensate by marking h-ful items as requiring special implementation (and h-epenthesis, incidentally, would simply be an instance of over-application of this special implementation to h-less items). Likewise, in perception, approximate representations correspond to an auditory level of processing: while QFs may fail to perceive /h/ phonetically/phonemically, they can detect acoustic differences between h-ful and h-less items (see the discussion below of Mah et al., 2016). Acoustic / auditory perception usually decays rapidly (Werker and Tees, 1984; Werker and Logan, 1985); approximate representations are an attempt to preserve these low-level perceptual distinctions. Such markings are not part of the toolkit supplied by Universal Grammar; instead they are add-ons introduced from outside the Language Faculty when normal phonological acquisition fails. It is not surprising then that they are less reliable than bona fide feature-based representations both in enabling h-perception and in cuing h-production. This explains the considerable variability in learner performance. In essence, the marking merely reflects auditory processing and signals how a form should be phonetically implemented.

It is worth noting that this representational view of L2 variation runs counter to established sociolinguistic accounts of L1 variation. These usually situate variation in the derivational grammar, whether due to variable rules (Cedergren and Sankoff, 1974), to shifts between categorical grammars (Kroch, 1989), or to partially ordered (Anttila, 1997) or overlapping constraints (Boersma, 1997); representational explanations are restricted to subsets of the lexicon (e.g., lexical exceptions in Guy, 2007). Studies of L2 variation have tended to adopt a similar derivational approach (Dickerson, 1975; Preston, 1989; Cardoso, 2007). Attributions of L2 variation to a phonological process, however, rely on the problematic assumption of accurate representation, as well as failing to account for the parallel phenomenon of variable perception. A lexical account of variation, based on approximate representations, avoids these problems. Second, two electroencephalography (EEG) studies show contradictory results regarding QF h-perception. On the one hand, an *unattended oddball paradigm* of auditory *linguistic* stimuli corresponding to syllables with and without [h] ([hʌm]-[ʌm]) failed to generate Mismatch Negativity [MMN—an event-related potential (ERP) associated with a deviant stimulus after a series of standard stimuli] among QF participants; only NSs exhibited this ERP (Mah et al., 2016). This finding suggests QFs neither perceive nor lexically record the distinction between h- and vowel-initial forms. Furthermore, *non-linguistic* noise burst stimuli as in [f], [hf], and [θf] elicited comparable MMN responses among NSs and QFs, which suggests that QFs have no problem with low-level processing of the acoustic properties of [h]. On the other hand, an *attended oddball paradigm* targeting [h] ([hɑ]-[ɑ]) led to MMN for both NSs and QFs (White et al., 2015). This finding suggests QFs are able to perceive, and potentially record in lexical entries, differences between h- and vowel-initial items. The nature of this possible representation is open to debate: it could be an actual /h/, an approximate representation as in John and Frasnelli (2022), or something else entirely, as we consider next.

The perceptual tasks in the two studies differ in whether participants attended to the input and in the quality of the vowel used in the stimuli: [ʌ] vs. [ɑ]. Mah et al. (2016) attribute the difference in perceptual accuracy not to the absence vs. presence of attention but to the phonetic realization of /h/ before the two vowels. In phonetic implementation, /h/ undergoes considerable contextual conditioning such that, particularly when preceded by a pause, it takes on the oral articulatory properties of the following vowel (Keating, 1988; Ladefoged and Maddieson, 1996). That is, /h/ is realized as a voiceless version of the subsequent vowel sound.

Before the low vowel [ɑ], [h] consequently resembles a voiceless uvular continuant. Since QFs have the devoiced uvular rhotic continuant [ʁ̥] in their repertoire (Tousignant, 1987; Walker, 2001; Sankoff and Blondeau, 2007), this might enable them to distinguish [hɑ] from [ɑ]: the tokens sound like [ʁ̥ɑ] vs. [ɑ]. This perceptual assimilation could explain the QF participants' MMN responses in White et al. (2015).

If Mah et al. (2016)'s explanation of the different findings holds, QFs assimilate instances of [h] before [ɑ] to the L1 category /ʁ/: it is for this reason that [h] in [hɑ] is detected rather than falling under the perceptual radar as in [hʌm]. By extension, instead of always omitting /h/, we might expect QFs to replace /h/ with /ʁ/ in URs containing the vowel /ɑ/. That is, just as QFs perceive [θ] as being an instance of /t/, thus storing the item *thank* as /tæŋk/, [h] before [ɑ] in *hot* would be heard and stored as /ʁɑt/. The subsequent realization of *hot* as [ʁ̥ɑt] could also sound sufficiently convincing to NS ears to pass as accurate. That is, perceptual assimilation can go both ways: NSs would interpret [ʁ̥] as /h/. Certainly, a listener would not process the output as an instance of h-deletion. To elaborate, by virtue of /t/ being a segment of English, NSs immediately recognize the substitution of [t] for /θ/; by virtue of the absence of /ʁ/ from the English inventory (typically, the English rhotic is alveolar /ɹ/ or retroflex /ɻ/; McMahon, 2002), however, NSs would not so readily detect the substitution of [ʁ̥] for /h/.

In sum, under Mah et al.'s (2016) proposal, we expect items with /ɑ/ (e.g., *hot, hop, hall, hard*) to be associated with higher rates of h-production than items with /ʌ/ (e.g., *hut, hug, hulk, honey*). That is, part of the variation observed in QF h-production may follow from how /h/ is at times included in URs (albeit as /ʁ/) and at others omitted or assigned an approximate representation. Just as Russian and Spanish learners may represent /h/ as /x/ or /χ/, francophone learners potentially represent a subset of items containing /h/ as /ʁ/, which should be reflected in production. The current study sought to establish whether QFs in fact show such variation in producing /h/ before /ɑ/ and /ʌ/. After the literature review in the next section, we outline the method used to test this prediction.

2 Background

In what follows, we consider the status and distribution of /h/ in both English and French and review previous work on QF perception and production of /h/.

2.1 On /h/ in English

In English, the voiceless glottal fricative /h/ is special in having only a single feature in its representation and a distribution limited to positions of prosodic prominence. Following arguments in Davis and Cho (2003), the representation for /h/ contains only [spread glottis], a laryngeal feature also associated with aspiration. The glottal fricative is thus unusual in having neither place of articulation nor manner features (Figure 1).

The restricted distribution of English /h/ runs parallel to that of aspirated stops. Barred from coda position, aspirated stops and /h/ appear only in word-initial onsets or, word-internally, in

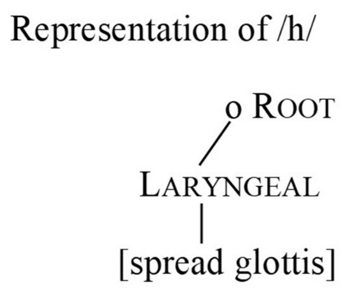


FIGURE 1
Representation of /h/.

onsets of stressed syllables (Table 1). At the beginning of word-internal unstressed syllables, /pʰ tʰ kʰ/ are de-aspirated (compare *pre*′[pʰ]*are* and *pre*[p]*a*′*ration*) and /h/ is deleted (compare *pro*′[h]*ibit* and *pro*[_]*i*′*bition*). If we assume that, like /h/, aspiration is underlying in English (on this, see Harris, 1994), the phenomena of de-aspiration and h-deletion in prosodically weak contexts can be captured via a unified process delinking [spread glottis] (or even the entire laryngeal node: Lombardi, 1995). Function words also undergo optional delinking of [spread glottis] such that *give her* and *give to* surface variably with and without initial [h] in *her* and aspirated [tʰ] in *to* (variation is shown as √ ∼ X in Table 1).

In a further distributional limitation, English /h/ forms a branching onset only with the glide /j/ (e.g., *huge* =/hjudʒ/) and, in some varieties of English, with the glide /w/ (e.g., *where* =/hwɛr/). That is, unlike most obstruents, /h/ never combines with the liquids /r/ or /l/. One means of accounting for this restriction is to argue that, being a single-feature segment, /h/ does not have sufficient strength to license a dependent segment of greater complexity than a glide (Harris, 1997).

Finally, speakers of h-dropping varieties of English in Britain and Newfoundland routinely leave /h/ out, variably or categorically deleting it regardless of position in the word (Wells, 1982; Milroy, 1983). That is, /h/ alternates with absence in the output of speakers of these varieties, much as it does in the speech of QFs. Indeed, h-droppers who try to emulate h-ful speech sometimes produce epenthetic [h] (Häcker, 2002), just like francophone learners of English. We address the status of /h/ in French in the next section.

2.2 On /h/ in French

Absent from the French inventory, /h/ constitutes a new phoneme that QF learners of English need to acquire. In addition, no native phoneme is perceptually close to /h/: rather than being assimilated to an L1 phoneme, /h/ is generally not detected in input (LaCharité and Prévost, 1999; Melnik and Pepercamp, 2019). As a result, QFs seem initially to leave /h/ out of URs, although they may eventually construct an approximate representation employing *ad hoc* diacritics (John and Frasnelli, 2022). QFs also typically exhibit variable h-deletion, alternating between _*appy* and [h]*appy* as a realization of *happy*. All the same, it is not entirely clear why /h/ poses such a considerable challenge for francophones. As we discuss

10.3389/flang.2023.1286084

TABLE 1 Distribution of [h] and aspiration in English.

	Word-initial		Word-internal		
	Onset (*content wd*)	Onset (*function wd*)	Onset (*stressed*)	Onset (*unstressed*)	Coda
[h]	√	√ ~ X	√	X	X
$[p^h\ t^h\ k^h]$	√	√ ~ X	√	X	X

next, [h] can occur in French epenthetically, so it is not beyond learners' articulatory abilities. More significantly, [h] appears as an allophone of /ʃ/ in some varieties of Quebec French. This further confirms that physical production of the sound is unproblematic. Additionally, as discussed below, the occurrence of allophonic [h] suggests that the French phoneme inventory employs [spread glottis], the sole feature required to represent /h/. This view on the status of [spread glottis] in French is based on Harris's (1994) position that URs are fully specified, and phonological processes are limited to operations of spreading, delinking and default insertion. Since [spread glottis] is not an unmarked feature eligible for default insertion, we assume it is present in URs, emerging as allophonic [h] in a lenition process which delinks all other features. Arguably, the presence of [spread glottis] in L1 representations should make it easier to access for representing L2 phonemes such as /h/. This follows from Brown's (1998) position that L2 phonemes *per se* are not problematic for acquisition, only L2 distinctive features (i.e., features not employed in L1 representations). Admittedly, Brown's proposal is made within the model of Minimally Contrastive Underspecification (Avery and Rice, 1989), which postulates that only those features required to establish contrasts are specified in URs. In this model, since [spread glottis] is not employed contrastively in French, the feature may be left out of URs. However, if we omit [spread glottis] from French representations, the lenition process which generates [h] in Quebec French is hard to account for. This brings us back to the question of why QFs cannot readily access [spread glottis] to represent English /h/. To resolve the conundrum, we propose the presence in the phonological component of a constraint on phonemic representations that are exclusively composed of laryngeal features. This Laryngeal Input Constraint makes /h/ a marked phoneme and accounts for why it is so problematic for francophones.

The sound [h] is at times realized in interjections in French—Walker (2001) mentions *hop!* and *hein?*—and even before vowel-initial content words instead of a glottal stop. While, as in many languages (Lombardi, 2002), glottal stops are the preferred epenthetic consonant in French (notably for marking *h-aspiré* words: Gabriel and Meisenburg, 2009), an epenthetic [h] can also occur.[1] Although epenthetic [h] may be more common in song (see footnote 1), its occurrence even there suggests that the francophone problem with English /h/ is not superficially phonetic, related to articulatory difficulty; francophones are physically capable of producing the sound. Instead, /h/ appears to constitute a phonological problem for L2 learners of English, but the precise nature of the problem is hard to pinpoint.

Arguably, instances of epenthetic [h] are merely added during phonetic implementation. That is, they may be like excrescent stops in nasal-fricative sequences such as *prin*[t]*ce* or *Chom*[p]*sky* in English, which seem to be articulatory effects generated phonetically rather than phonologically (Ohala and Ohala, 1993; Feldscher and Durvasula, 2017). It is of course difficult to determine definitively whether a speech phenomenon has its source in the phonological system or in phonetic implementation. Categorical phenomena tend to be phonological, whereas variable / optional phenomena could be either. For example, epenthetic glottal stops (a variable phenomenon in English and many other languages) may result from a phonological preference for syllables with onsets (e.g., the Onset constraint in Prince and Smolensky, 1993/2004), or they could simply reflect ease of articulation, emerging like excrescent stops during phonetic implementation. The same applies to epenthetic [h] in French. Under the analysis of epenthetic [h] as a product of the phonetic system, its occurrence in French would not necessarily aid in the perception and eventual phonological acquisition of /h/.

When [h] is generated as an allophone of an underlying segment, however, this is clearly a phonological process, which can conceivably facilitate acquisition of underlying /h/. Some varieties of Quebec French instantiate a process of debuccalization, whereby /ʃ/ is reduced to [h] in onset position: *chocolat chaud*, for example, is realized as [h]*ocolat* [h]*aud* (Bittner, 1995; Paradis and LaCharité, 2001; Morin, 2002). Similar processes are observed in other languages, [h] being in Spanish a well-attested product of lenition of coda /s/ and a pre-deletion segment: /s/ → [h] → Ø (File-Muriel and Brown, 2010; Núñez-Méndez, 2022). Brazilian Portuguese instantiates a similar pattern for coda rhotics: /ʁ/ → [h] → Ø (Rennicke, 2015). Interestingly, QFs seem to readily process [h] as an allophone of /ʃ/ even if this pronunciation does not occur in their own variety. There is no evidence that QFs without this allophonic process struggle to understand speakers who do realize /ʃ/ as [h]; indeed, although to our knowledge the matter has not been formally investigated, informal observations suggest that non-debuccalizing QFs are capable of imitating speakers who realize /ʃ/ as [h]. Again, articulation of /h/ is resolutely not the problem.

Under the view that lenition involves feature loss (Harris, 1990, 1994, 1997; Honeybone, 2008), debuccalization consists of suppression of all other features in the representation except [spread glottis]. The process can be captured in feature geometry (e.g., Clements, 1985) via delinking of the supralaryngeal node, leaving only the laryngeal node with [spread glottis] as dependent

1 An example of [h] in an interjection occurs in commercials for *Familiprix* where *aha* is unmistakably realized as [aha]: https://www.youtube.com/watch?v=aURNJoNgUIQ. An epenthetic [h] before a vowel-initial form occurs in the introductory song for the animated series *Wakfu*, where *héros malgré lui* is realized as [h]*éros*: https://www.youtube.com/watch?v=_7TvSNdgKik. Likewise, Edith Piaf clearly sings [h]*Allez, venez, Milord!* on two occasions in the following version of her well-known song: https://www.youtube.com/watch?v=nZcdl1u_9o8.

feature. The cross-linguistic occurrence of [h] as a product of lenition, as well as its tendency to alternate with zero, is thus consistent with the view of /h/ as a single-feature phoneme. It also indicates that [spread glottis] must be part of the underlying representation for Spanish /s/, Brazilian Portuguese /ʁ/ and Quebec French /ʃ/, even though none of these languages have /h/ in their phoneme inventory. This is hard to reconcile with Brown's (1998) claim that L2 phonemes using distinctive features employed in L1 representations should be relatively easy to acquire. If QFs already require [spread glottis] in their L1 (albeit as a non-contrastive feature), acquisition of /h/ should be relatively straightforward. However, production data from previous studies suggest that acquiring /h/ is highly challenging: despite extensive English-language studies (mean: 12.06 years), only 12 of the 50 QF participants in John and Frasnelli (2022) showed no h-deletion even in a limited reading-aloud task. QFs who perform on a par with NSs of English in the perception and production of /h/ are thus the exception.

Despite [spread glottis] potentially being available to QFs from their L1, learners have difficulty acquiring a phonemic representation containing only this distinctive feature. Possibly something other than the absence of the feature from the L1 is at work to block the acquisition of phonemic /h/. We propose the existence of an input constraint. Phonological constraints in recent decades have been construed in Optimality Theory (OT; Prince and Smolensky, 1993/2004; see also McCarthy, 2002) as applying purely to output. Although, under Lexicon Optimization, input representations generally reflect surface forms, the principle of Richness of the Base (Smolensky, 1996; Davidson et al., 2004) considers that input forms are entirely unconstrained (for challenges to Richness of the Base, cf. Vaysman, 2002; Gouskova, 2023). In theory, this means that L2 learners should have no problem developing URs that contain novel segments, including highly marked phonemes; their only challenge should be to re-rank markedness constraints such that the segments can emerge in output.

Conceivably, however, the phonological component incorporates input constraints that make certain URs dispreferred (i.e., a Restriction on the Base). For example, a constraint on underlying segments comprised exclusively of laryngeal features would favor phoneme inventories that lack /h/; it would likewise exclude /ʔ/, glottal stops being comprised solely of the feature [constricted glottis]. A [spread glottis] Laryngeal Input Constraint would account for the absence of /h/ from the phoneme inventories of Spanish, Brazilian Portuguese and Quebec French even though [h] appears in the output of these languages. The same case can be made for languages such as English that lack underlying /ʔ/ but employ glottal stops as a surface allophone: they could contain a [constricted glottis] Laryngeal Input Constraint. Interestingly, the proposed constraint contradicts the claim that glottal is an unmarked feature. For example, according to Lombardi (2002) and de Lacy (2006), glottal constitutes the least marked place of articulation. Our position is that, while this unmarked status may hold at the surface, it seems not to apply underlyingly.

While diverging from OT-based output constraints, the notion of restrictions on input recalls Morpheme Structure Constraints. For example, the Obligatory Contour Principle (OCP) bars the presence in a UR of adjacent identical features (e.g., in the initial conception, tones: Leben, 1973). Though the OCP was later expanded to apply to output, triggering and blocking phonological derivations (McCarthy, 1986; Yip, 1988), it was originally conceived of as a constraint on URs much like the Laryngeal Input Constraint. The OCP has also been depicted as a soft constraint that is not always respected (Odden, 1986); underlying structures that violate the OCP are avoided since more marked, but not strictly ruled out. The same seems to be the case with the Laryngeal Input Constraint: while /h/ and /ʔ/ are dispreferred, phoneme inventories can nonetheless contain these laryngeal segments.

The distinction between phonological constraints on output (as in OT) and input (as proposed here) is intriguing insofar as the content of the two constraint types may be contradictory. Output constraints target either *faithfulness* ("output is identical to input") or *markedness* ("output is less marked than input"). Consequently, output that diverges from input is necessarily less marked. Following this argument, [h] should be an unmarked segment, as it is an allophone of underlying /ʃ/, /s/, and /ʁ/ in Quebec French, Spanish, and Brazilian Portuguese. According to the Laryngeal Input Constraint, however, /h/ and /ʔ/ are marked structures in URs. Apparently, the markedness status of a segment can differ between underlying and surface levels of representation.

Concentrating on surface realizations of laryngeal segments, an interesting parallel can be made between the glottal fricative [h] in QF and the glottal stop [ʔ] in English, both of which have epenthetic as well as allophonic status. As mentioned above, QFs arguably process epenthetic [h] in French as a purely phonetic effect, filtering it out as a linguistically irrelevant segment unrelated to any phoneme. In this sense, instances of epenthetic [h] (in *hop!* and *hein?* and elsewhere) are processed in similar fashion to inserted glottal stops in English (Garellek, 2012). In our experience (e.g., the first author's, as a NS of English), anglophones have difficulty detecting epenthetic [ʔ] in the speech signal. To notice the presence of glottal stops, listeners have to attune their ears to a phonetic level of processing, which degrades rapidly under normal speech perception (Werker and Tees, 1984; Werker and Logan, 1985). Our impression is that neither epenthetic [h] in French nor epenthetic [ʔ] in English is particularly salient since neither corresponds to an underlying segment. Consequently, the presence of epenthetic [h] and [ʔ] in the L1 would not necessarily aid in the acquisition of phonemic /h/ and /ʔ/ in an L2.

However, like [h] in Quebec French, [ʔ] in some varieties of English can also be an allophone of an underlying segment. That is, one and the same form in output can have either epenthetic or allophonic status. Just as [h] can be an allophone of /ʃ/ in Quebec French, intervocalic /t/ can be realized as [ʔ] in British English as in *butter* → *bu*[ʔ]*er* (Harris and Kaye, 1990). Interestingly, while English speakers are largely unaware of epenthetic [ʔ] (i.e., it passes under the perceptual radar), allophonic [ʔ] is readily perceived. While processing allophonic [ʔ] is effortless and automatic, detecting epenthetic [ʔ] requires concentration on the signal. Despite epenthetic vs. allophonic [h] and [ʔ] having comparable phonetic properties, the segments are perceived quite differently according to their status. When [h] and [ʔ] are epenthetic, they tend not to be noticed; but when [h] and [ʔ] are allophonic, they are readily perceived.

This distinction, albeit speculative, strikes us as insightful for how QFs perceive and represent /h/ in English. As the studies reviewed in the next section show, it seems that QFs often fail to detect English [h] in speech, leaving it out of URs or at some point developing an approximate representation that allows for variable detection of this sound. In this sense, English [h] is processed like epenthetic [h] in French or epenthetic [ʔ] in English. Nonetheless, not all instances of [h] in English are necessarily equal: depending on the type of adjacent vowel, which affects how [h] is articulated, [h] may have phonetic properties that allow QFs to perceive it as a surface form of French /ʁ/. In this case, input [h] is associated to an underlying segment, much as [ʔ] in English at times corresponds to underlying /t/. It is just that this underlying segment is /ʁ/ rather than /h/. While the latter representation, according to the Laryngeal Input Constraint, is hard to acquire, QFs will easily construct the former representation, given that the phoneme is already available in the L1 inventory.

2.3 QF perception and production of English /h/

The QF tendency to delete /h/ in English conceivably derives from a difficulty in distinguishing h-initial and vowel-initial forms. From a bottom-up perspective, if QFs cannot hear the difference between *heat* and *eat* any more than an anglophone distinguishes a realization of *eat* with or without a glottal stop, they will not record any distinction in URs. The minimal pair will consequently be stored as homophonous /it/. From a top-down perspective, the presence of a particular phoneme category in a listener's inventory leads to automatic detection of instances of the category—this is what makes speech comprehension so effortless. QF learners of English are disadvantaged in not having /h/ in their phoneme arsenal, thus impeding their ability to detect the speech sound and to record it in URs. As we see below, earlier studies on the perception and production of /h/ by francophones have findings either consistent with this "defective UR" account or pointing to the need for a more nuanced view.

The ERP findings in Mah et al. (2016) largely support the notion that QFs fail to record /h/ in URs. In an unattended oddball paradigm using both linguistic and non-linguistic auditory stimuli, participants listened to a series of repeated stimuli (standards) interspersed with different stimuli (deviants). Detection of a deviant in a stream of standards is associated with the ERP MMN. Both QFs and NSs of English exhibited MMN with *non-linguistic* noise burst stimuli ([f], [hf], [θf]), which suggests that QFs have no trouble with auditory detection of [h]. Only NSs exhibited MMN, however, with *linguistic* stimuli containing [h] ([ʌm], [hʌm]). This suggests that QFs fail to perceive [h] when processing the signal as speech, consistent with their leaving /h/ out of URs.

Nonetheless, there is a mismatch between what the perception data indicate about URs (apparently inaccurate) and the production data (surprisingly accurate): in a reading-aloud task, more than half the participants realized all 8 tokens of /h/. Mah et al. attribute this production accuracy to the influence of orthographic cues that guide the realization of h-initial forms. While previous studies have found some evidence that exposure to orthographic forms can scaffold acquisition of confusable phoneme contrasts (e.g., in a non-word learning task with Dutch participants targeting the /ε/-/æ/ contrast in English: Escudero et al., 2008), the role of orthography in guiding speech production is far from clear. Indeed, QFs were found to produce more instances of h-epenthesis in reading aloud than in spontaneous speech (John and Cardoso, 2009); that is, despite the evidence for h-initial vs. vowel-initial forms being in full view, orthography failed to promote more accurate output. We are thus skeptical of claims that learners can reliably use written forms to guide production.

It is not that orthography plays no role in L2 phonological acquisition (see Bassetti et al., 2015; Hayes-Harb and Barrios, 2021, for overviews). For example, when learners transfer grapheme-phoneme correspondences from the L1 to the L2, this may interfere with production (e.g., Rafat, 2016). Indeed, the grapheme < *h* > does not correspond to any speech sound in French, as shown in homophones such as *aine-haine* /εn/ ("groin"-"hate"), *eau-haut* /o/ ("water"-"high"), and *ache-hache* /aʃ/ ("wild celery"-"axe"). Certainly, French contains orthographically h-initial items referred to as *h-aspiré* that act as though they are consonant-initial (Charette, 1991; Tranel, 1995). These forms block linking processes such as *liaison* that supply an onset to otherwise onsetless forms. Nonetheless, the beginning of an *h-aspiré* item does not itself correspond to an actual speech sound. Consequently, QFs are used to processing < *h* > as a silent letter. To complicate matters, < *h* > is not even a reliable indicator of /h/ in English since it remains unpronounced in a few high-frequency words (e.g., *hour, honor, honest*). A recent pilot project indicates that the inconsistency of the grapheme-phoneme correspondences for English /h/ could have a confounding effect on the development of accurate lexical representations (Jackson and Cardoso, 2023). In brief, the fact that learners do not associate < *h* > with any speech sound in French (i.e., unlike letters such as < *j* > or < *s* >) and that < *h* > is not always pronounced in English makes the grapheme more likely to generate h-deletion than to promote h-production.

It is also not clear that QFs always fail to perceive the distinction between h-initial and vowel-initial forms. While the results of the *unattended oddball paradigm* in Mah et al. (2016) point in this direction, the findings in White et al. (2015) are not consistent with this view. In an *attended oddball paradigm*, the latter researchers found similar MMN responses among QFs and NSs with auditory stimuli using the syllables [hɑ] and [ɑ]. This suggests that QFs are able to perceive [h] under certain conditions, whether they can record /h/ in URs or not.

Possibly, the difference in the findings of the two EEG studies is due to the presence or absence of attention; that is, QFs can perceive the [h]-Ø contrast as long as they are attending to the phonetic input. In our view, approximate representations certainly require special effort for learners to draw on them in perception and production. If such *ad hoc* markings are associated with /h/, it comes as no surprise for attention to assist perception. Mah et al. (2016), however, intriguingly attribute the difference to the vowel type used in the two studies: [ɑ] vs. [ʌ]. From our perspective, their position deserves further consideration and investigation. The absence of phonological place features means /h/ undergoes considerable contextual conditioning during phonetic implementation. The low vowel [ɑ] influences the realization of /h/, creating an acoustic effect that resembles a voiceless uvular continuant (Keating, 1988; Ladefoged and Maddieson, 1996). Since QFs often devoice the uvular rhotic continuant /ʁ/ in their inventory, such that /ʁ/→ [ʁ̥] (Tousignant, 1987; Walker, 2001; Sankoff and Blondeau, 2007), this could facilitate their ability to distinguish [hɑ] from [ɑ]. For QFs, this pair conceivably sounds like [ʁ̥ɑ]-[ɑ], a contrast that is easy for them to process, unlike the [hʌm]-[ʌm] contrast.

By extension, if QFs assimilate [h] to /ʁ/ before [ɑ], they should also substitute /ʁ/ for /h/ in URs of items containing /ɑ/. Consequently, while an item such as *hut* might be stored as /ʌt/ (/h/ omitted), an item such as *hot* would be stored as /ʁɑt/ (/ʁ/ substituted for /h/). If /ʁɑt/ is then realized as [ʁ̥ɑt], it could strike NSs' ears as sufficiently close to /h/; certainly listeners would not have the impression /h/ has been deleted. Indeed, when Spanish or Russian speakers substitute [x] for /h/, NSs of English automatically classify the input as a realization of /h/, even if it sounds phonetically unconventional. Essentially, NSs themselves assimilate [x], a sound missing from their inventory, to the closest L1 phoneme, namely /h/. Similarly, NSs could process the QF realization [ʁ̥] as /h/. This is not the only possibility, however, for how QFs might represent /h/; they may instead develop an approximate representation using *ad hoc* diacritics, as we consider next.

While the two EEG studies show either presence or absence of MMN, consistent with ability/inability to perceive and possibly represent /h/, John and Frasnelli (2022) found highly variable QF perception and production of /h/; they also observed considerable inter-participant variation. Indeed, one of the advantages of behavioral over ERP data is that precisely this kind of variation is easier to discern. Because ERP responses are quite subtle, extensive data are required from numerous participants for patterns to emerge. Consequently, while differences between larger groups (e.g., QFs vs. NSs) are detectable, differences between individuals are typically lost. The intra- and inter-participant variation observed in John and Frasnelli (2022) is consistent with varying degrees of gradient perception across learners rather than simple ability/inability to perceive /h/. Perception was tested via an attended oddball paradigm task with trials where the fourth item was either the same or different from the preceding three. Stimuli were mono- or disyllabic real and non-words involving a variety of vowel sounds (e.g., *heat-heat-heat-eat, old-old-old-hold, hice-hice-hice-ice, enk-enk-enk-henk*), although the analysis did not include vowel quality as a condition. At the end of each trial, participants indicated via keyboard press whether the final item was the *same* or *different*. For the condition targeting /h/, QFs showed lower mean accuracy rates than NSs (64 vs. 96%) as well as considerably wider ranges (0.8–100% vs. 91–100%), with individual QF participant rates distributed evenly across the broad range. That is, QFs did not only show either poor or nativelike perception but everything in-between. QF h-production rates also covered a wide range (23.81–100%), and accuracy rates in perception and production were highly correlated.

QFs thus generally struggle to perceive and produce /h/ but nonetheless show gradient differences between individuals. Some QFs exhibit very poor perception and production, consistent with failure to record /h/ in URs; others perform on a par with NSs, consistent with having acquired /h/. Most, however, perform somewhere between these two poles, a distribution consistent with their having developed some distinction between h- and vowel-initial forms but falling short of full acquisition of phonemic /h/. Instead, John and Frasnelli argue for approximate ("fuzzy" or "murky") representations. These are indicated in URs with a diacritic such as a superscript question mark that reflects the murky status of the representation. Items such as *hut* and *hot* are thus presumed to be stored as /ˀʌt/ and /ˀɑt/. It may, however, be more appropriate to think of these markings as separate from actual phonological representations. Similar to orthographic information associated with a lexical entry, approximate representations may constitute extra-phonological add-ons. These reflect a special phonetic quality that is both detectable when the input is attended to and reproducible when output is formulated with sufficient control and effort. Unlike representations employing features supplied by Universal Grammar, approximate representations would allow for only variable perception and production of /h/. As a function of experience, learners should show improvement in performance, thus accounting for the wide distribution in their perceptual and productive abilities.

Nonetheless, not all of the variation in h-production and perception observed among QFs is necessarily due to approximate representation of /h/; some variation may be due to assimilation of /h/ before /ɑ/ to /ʁ/. That is, as well as representing an item such as *hut* as /ˀʌt/, QFs may represent an item such as *hot* as /ʁɑt/, substituting /ʁ/ for /h/. Although inaccurate in terms of the target, the phonemic representation /ʁ/ should permit QFs to distinguish consistently between h- and vowel-initial items in perception and production, but only h-initial items containing /ɑ/. Additionally, while development of an approximate representation constitutes a strategy for getting round the Laryngeal Input Constraint, assimilation of /h/ to /ʁ/ means the constraint does not even apply to the input.

2.4 Research question and hypothesis

In the current study, we test the prediction that QFs assimilate /h/ before /ɑ/ to /ʁ/ and thus produce lower rates of h-deletion in items such as *hot* than *hut*. Our aim is to answer the following research question and verify the hypothesis given below.

Research question: Do QFs differ in their production of /h/ before /ɑ/ and /ʌ/? Hypothesis: QFs will show high accuracy in the production of /h/ in items where the following vowel is /ɑ/, but only variable production of /h/ in items where the following vowel is /ʌ/.

More graphically, h-production should show:

i) Accuracy: /hɑt/ > /hʌt/ (where the symbol ">" means "greater accuracy than").
ii) Variation: /hʌt/ > /hɑt/ (where the symbol ">" means "greater variation than").

The method used to test our hypotheses is outlined next.

3 Materials and methods

3.1 Participants

A total of 34 QFs (9_{males}, M age = 28.00 yrs, range = 18–57; $25_{females}$, M age = 30.79 yrs, range = 21–53) were recruited mainly among the student bodies of francophone universities in Quebec, a majority French-speaking region of Canada. For context, we should explain that French is the sole official language in Quebec, although anglophones constitute a significant minority in certain regions, notably in Montreal, the largest city in the province. School boards in Quebec are divided along linguistic lines, with francophones attending schools run by the French school board, where English language instruction is usually introduced in the later years of primary school. Consistent with this situation, language background questionnaires administered before the production task established that the participants generally started learning English in a classroom setting from an early age (M age = 9.19 yrs; range = 5–18 yrs) and for an extended duration (M = 12.93 yrs). It should be noted that the questionnaires revealed considerable variation in both degree and type of exposure, as well as in age of initial exposure, such that it would be difficult to investigate any correlation between age or degree of exposure and production accuracy. Hypothetically, we might anticipate a divide between participants in Montreal and outlying regions, since the potential for contact with NSs is greater in Montreal. In practice, however, contact and English language use in Montreal can vary wildly, from virtually none to occasional or sustained contact, whether with neighbors, friends or colleagues/customers, and this can change considerably from one period in a person's life to another. Conversely, residing in a region with few anglophones does not preclude contact in diverse settings such as the workplace, foreign travel, immersion exchanges, and online environments (e.g., input from Netfix shows, virtual exchanges in gaming). Likewise, while we might presume participants who started learning English in the classroom from age 5 or 6 would have an advantage, such early exposure in the Quebec context is typically limited to 1 h per week, in which case the amount of exposure is too limited to provide an edge in acquisition. In brief, in the face of such diversity of experience, we did not attempt to establish correlations with h-production.

3.2 Materials

In an online environment using Zoom, participants were recorded reading aloud a series of 36 expressions (e.g., *some* ***hot*** *apple pie*; *a big* ***hug***) and 25 sentences (e.g., *True love is* ***hard*** *to find; She lives in a mud* ***hut***) presented one-by-one on PowerPoint slides. In all, the task contained 80 target items containing /h/ followed by either /ɑ/ (*hot, hard*) or /ʌ/ (*hug, hut*)—see Appendix A for the full list of expressions and sentences. Equal numbers of target items with /ɑ/ and /ʌ/ appeared in both the phrases and sentences.

3.3 Data analysis

The recordings were coded impressionistically by a NS of English, who indicated whether /h/ was deleted or preserved in the target item, as well as noting instances of h-epenthesis, although these were not included in the actual analysis. If the rater judged that /h/ was preserved in a given item, this was coded as an instance of accurate h-production. Using R software for statistical computing (R Core Team, 2021), we carried out two-way ANOVAs with accuracy of h-production as dependent variable and vowel type (/ɑ/ vs. /ʌ/) and stimulus type (phrase vs. sentence) as intra-participant independent variables.

Effects for vowel type were anticipated, with items containing /ɑ/ (*hot*) expected to show higher rates of accuracy than items containing /ʌ/ (*hut*). Although the "stimulus type" variable was included more for exploratory purposes, we might also expect higher rates of h-production to be associated with phrase stimuli, given that short phrases are less cognitively challenging to process and articulate than full sentences. This effect is particularly anticipated if /h/ is captured via approximate representations in the QF lexicon, since such representations are thought to entail greater effort than true phonemic representations.

Since classical ANOVAs cannot be used to support the null hypothesis (absence of difference in h-production), we subsequently ran Bayesian *t*-tests to directly compare the amount of evidence in favor of the null hypothesis for both vowel type and stimulus type. These analyses yielded a Bayes Factor (BF_{10}) that corresponds to the ratio of evidence in favor of the alternate model (i.e., where there is a difference as a function of the parameter considered) vs. the evidence in favor of null model (where there is no difference). Specifically, BF values >1 indicate the strength of evidence in favor of the alternate model or, should BF be <1, the lower the value the stronger the evidence in favor of the null hypothesis (Rouder et al., 2009, 2012). Benchmark scores are: BF_{10} between 1 and 1/3 are considered weak (barely worth mentioning); between 1/3 and 1/10, they are considered substantial; and <1/10, they are considered strong evidence in favor of the null hypothesis (Jeffreys, 1961). The higher this value, the greater the evidence in favor of the alternative hypothesis, with benchmarks BF_{10} between 1 and 3 considered weak evidence and between 3 and 10 substantial (Jeffreys, 1961).

TABLE 2 QF accuracy rates for h-production (%).

		Descriptive statistics		
Independent variables		Mean	Std deviation	Range
Vowel type	/ɑ/	76.82	29.07	0-100
	/ʌ/	77.12	26.89	0-100
Stimulus type	Phrases	77.73	26.57	5-100
	Sentences	76.21	31.46	0-100

4 Results

Table 2 shows the mean accuracy rates in h-production by the 34 QF participants (2,720 tokens in all) according to the key independent variable of vowel type (/ɑ/ vs. /ʌ/) and the further variable of stimulus type (phrase vs. sentence).

It should be mentioned that considerable inter-participant variation was observed, with the accuracy rates for h-production of individual participants distributed across a range from near-categorically inaccurate to categorically accurate (2.5–100%). We can also report that 13 participants produced between 1 and 4 instances of h-epenthesis (e.g., *a broken* [h]*arm, the front* [h]*office*). For the purposes of the two-way ANOVAs and *t*-tests, the data from the 7 categorically accurate h-producers were removed, since the purpose is to analyze variation. Note that including these 7 participants did not change the results of the ANOVA and Bayesian *t*-tests indicated next.

Classical ANOVAs run on participant accuracy rates could not reveal any effect for vowel type ($F_{[1,26]} = 1.00$, $p = 0.327$, $\varepsilon^2 = 0.001$) and stimulus type ($F_{[1,26]} = 0.17$, $p = 0.685$, $\varepsilon^2 < 0.001$), nor any interaction between these factors ($F_{[1,26]} = 1.35$, $p = 0.256$, $\varepsilon^2 = 0.003$). The results suggest that we cannot rule out the null hypothesis (i.e., an absence of difference in accuracy rates related to these factors). Paired Bayesian *t*-tests were run to compare the accuracy rates of participants as a function of vowel type and of sentence type. Following previously indicated benchmarks, the Bayesian paired *t*-tests provide moderate evidence for the null hypotheses, that is, that vowel type and sentence type had no influence on accuracy ($BF_{10} = 0.220$ and $BF_{10} = 0.236$, respectively).

5 Discussion

While QFs typically struggle to detect [h] in input, Mah et al. (2016) proposed that the phonetic quality of [h] before the vowel [ɑ] leads QFs to hear [h] as [ʁ̞], a common realization of the L1 segment /ʁ/. This would explain why, in two EEG studies using an oddball paradigm to target the ERP MMN, White et al. (2015) detected MMN responses among QF participants, whereas Mah et al. did not: the former study employed stimuli containing [ɑ], and the latter stimuli containing [ʌ]. We extrapolated that the presumed perceptual assimilation that facilitates h-perception before [ɑ] should lead QFs to represent /h/ as /ʁ/, but only for items where the following vowel is /ɑ/ (*hot*) and not /ʌ/ (*hut*). If such is the case, the distinction should be clearly reflected in h-production: QFs should show decidedly higher rates of h-production for items containing /ɑ/ than /ʌ/. Indeed, we expected QFs to show only variable h-production for items such as *hut,* represented in our view as /ˀʌt/ (John and Frasnelli, 2022). The approximate representation of /h/ via a superscript question mark (more properly an add-on external to the actual phonological representation) is what permits QFs, despite the absence of /h/ from their segmental repertoire, to distinguish h-initial from vowel-initial items such as *heat-eat* in their lexicon. The lexical distinction leads QFs to realize higher rates of h-production in actual h-initial items than hypercorrect h-epenthesis in vowel-initial items (John, 2006; John and Cardoso, 2009). In brief, according to our hypothesis, items such as *hot,* by virtue of /h/ being replaced with /ʁ/, should exhibit essentially categorical h-production, whereas items such as *hut,* with murkily specified /h/, should only exhibit variable h-production.

With data from a reading-aloud task involving 27 QFs, however, two-way ANOVAS and paired Bayesian *t*-tests revealed no difference in the realization of /h/ before /ɑ/ vs. /ʌ/ (vowel type) nor in phrases vs. sentences (stimulus type). QFs showed comparably variable h-production regardless of vowel, and the hypothesized greater accuracy in h-production for *hot* vs. *hut* failed to materialize. Instead, the null hypothesis (i.e., vowel type has no effect on h-production) was confirmed. This suggests that /h/ has the same representation for QFs regardless of the adjacent vowel, and the variation found in QF h-production in no way derives from learners' at times assimilating /h/ to /ʁ/. In this case, the fact that MMN was observed among QFs in the oddball paradigm task in White et al. but not Mah et al. remains to be explained. We suggest that the design difference involving presence/absence of attention is responsible: in the former study, participants attended to the stimuli, whereas in the latter, they did not. Interestingly, attention is not an absolute requirement for h-perception, since MMN was observed among NS participants in Mah et al. It seems only QF learners of English need to attend to the signal in order for /h/ to be detected in the linguistic input.

In our view, the need for attention in order for QFs but not NSs to detect /h/ is consistent with a difference in the representation of this segment: unlike NSs, QFs do not record in URs an actual phonemic representation for /h/, involving the laryngeal feature [spread glottis]; instead, they use an approximate representation that bypasses distinctive features. Such *ad hoc* markings are less reliable and require greater effort than representations using UG-based features in supporting speech perception and production. Although we might have expected the association between approximate representations and effort/attention to result in greater accuracy in short phrases than full sentences, this variable failed to emerge as an influence on h-production. Nonetheless, considerable differences are consistently encountered between QFs and NSs in h-perception and production. In addition, QFs exhibit wide ranges in performance (e.g., John and Frasnelli, 2022): while learners initially have difficulty using approximate representations to support perception and production of /h/, over time they get better at the task, such that performances may rival those of NSs, with their actual feature-based representation of /h/.

Conceivably, the reading-aloud task, compared with spontaneous speech, may be particularly conducive to the kind of focused attention required for learners to draw on approximate representations as a cue for h-production. QFs are acutely aware of their difficulties with English /h/, so part of their success in h-production may be due to the task facilitating efforts to control articulatory behavior. Attention involves concentrated awareness directed toward a particular input, as with White et al.'s (2015) attended oddball paradigm, or output, as with our reading-aloud task (for a review of the concept of attention, see Lindsay, 2020). It may be that the 7 participants who showed categorical h-production were particularly skilled at the kind of heightened vigilance required to produce a speech sound that lacks a phonemic representation in their lexicon.

That QFs generally fail to develop an accurate representation for /h/ is unusual, given that [spread glottis], as we argued in the Background section, is present in L1 representations. Although [spread glottis] is not a contrastive feature in French, we considered, contra Brown (1998), that it should be available for developing L2 representations, making /h/ relatively easy to acquire. Since QFs have considerable difficulty acquiring /h/, we suggested that the phonological component contains a Laryngeal Input Constraint that renders highly marked any representations based exclusively on laryngeal features. Such a proposal runs counter to the OT view that constraints apply only to output, URs being entirely unconstrained. The principle of Richness of the Base (Smolensky, 1996; Davidson et al., 2004) seems to be disproved by L2 phenomena such as QF acquisition of /h/; the base itself is apparently subject to markedness constraints. When it comes to /h/, QFs circumvent these by constructing an approximate representation that employs a non-feature-based diacritic. The proposal is important because it provides us with a means of characterizing certain challenges in L2 phonological acquisition. These are not problems of phonological or phonetic output nor even perceptual problems *per se*, but problems of underlying representation. Approximate representations also permit us to understand the considerable variation that characterizes L2s: variation follows directly from the nature of the *ad hoc* status of the representation.

It remains an open question as to whether some QF learners of English overcome the Laryngeal Input Constraint and eventually develop an accurate representation for /h/. The presence of 7 participants with categorically accurate h-production points to the possibility that these learners have in fact acquired /h/. Nonetheless, it is also possible that, given a more extended task or a task less conducive to attention, these highly proficient h-producers would have eventually slipped up and exhibited occasional instances of h-deletion. More intriguingly, another possibility is that some learners become so adept at drawing on approximate representations to ensure h-production that their performances are indistinguishable from those of NSs, despite the differences in representation of /h/. As such, some learners may constitute what John and Frasnelli (2022) refer to as "phonological zombies." The term "zombie" is borrowed from debates on the philosophy of mind (Chalmers, 1996) and is in no way intended to disparage L2 learners. The point of the zombie concept is that, while we each have privileged access to our own internal worlds, including subjective thoughts and feelings that confirm that we are personally in possession of consciousness, we can never be entirely sure about those around us. Despite exhibiting behavior consistent with a similar inner life, others may not experience consciousness exactly as we do. Indeed, we cannot be sure others are endowed with consciousness at all: we may be surrounded by zombies who only show the outward signs of consciousness. By extension, some L2 learners may exhibit an ability to perceive and produce /h/ that is consistent with their having acquired an accurate representation, but we cannot be sure that this is the case: they may be phonological zombies who perform on a level with NSs, despite /h/ being captured in their lexicon by approximate rather than phonemic representations. The challenge for future research is to design an experiment able to distinguish between phonological zombies and L2 learners who have in fact acquired an accurate representation for /h/ or other L2 sounds.

Finally, the current study contains certain limitations. Notably, while our analysis of the reading-aloud data investigates the influence of the following vowel on h-production, we did not include an examination of preceding sounds. As can be seen in Appendix A, items beginning with /h/ were preceded by different consonant or vowel sounds and at times by no sound at all. Since the preceding phonetic environment has been shown to influence rates of h-epenthesis (John and Cardoso, 2009), it remains possible that this variable also affects h-production. Furthermore, the information participants provided in the Language Background Questionnaire did not permit us to develop a clear portrait of age of acquisition and degree of exposure to L2 English. For example, although some participants reported starting English instruction as early as 5 or 6 years old, the exposure was not necessarily sufficient to provide an advantage over learners who started at an older age. Likewise, exposure is hard to quantify and compare. To demonstrate, one participant reported "occasional interactions with anglophone friends" while another indicated having spent "8 months working with anglophones." Determining which situation constitutes greater exposure proved impossible, thus preventing us from exploring the influence of this variable in our data. Finally, we did not perform a fine-grained phonetic analysis of QF tokens of /h/, which could further elucidate whether /ʁ/ is at all substituted for /h/ underlyingly. In future research, it would be worthwhile to use a tool such as PRAAT for an acoustic analysis of QF and NS productions of [h] in English, comparing these with QF productions of [ʁ̥] in French.

6 Conclusion

Phenomena that involve L2 segments, such as the deletion of /h/ by QFs investigated here, are frequently variable (John and Frasnelli, 2022). This raises the question of how L2 segments are represented. If L2 representations are simply accurate, it remains to be seen why learners struggle to produce the target sounds and, more crucially, why learners have parallel problems in perception. QFs not only delete /h/, but they also fail to detect /h/ in the speech signal, the implication being that the state of the intervening representation is responsible for both. If, conversely, L2 representations are inaccurate, that is, if /h/ is simply left out of URs, it remains to be seen how QFs are able at times to produce and perceive /h/. Previous research has suggested that learners resort to an approximate representation of L2 phonemes (John and Frasnelli, 2022). Such diacritic representations are not as reliable as actual feature-based representations, but they allow learners to perceive and produce /h/ to varying degrees. Two earlier EEG studies, however, suggested another possible source for variation in h-production and perception: while [hʌm]-[ʌm] stimuli in an oddball paradigm failed to generate MMN among QFs (Mah et al., 2016), consistent with /h/ being left out of URs, [hɑ]-[ɑ] stimuli were accompanied by MMN (White et al., 2015), possibly because the phonetic properties of [h] before [ɑ] allow it to be assimilated to, and consequently represented as, the L1 phoneme /ʁ/. If /h/ is represented as /ʁ/ in items containing /ɑ/ (*hot*) but not /ʌ/ (*hut*), we hypothesized that QFs should show lower rates of h-deletion in such items. While QFs showed considerable variation in h-production in a reading-aloud task, none of this variation, however, could be attributed to the type of vowel occurring in an item. Our conclusion is that /h/ must be represented identically in the QF lexicon regardless of vowel type. While some QFs may simply leave /h/ out of URs such that *hot* and *hut* are stored as /ɑt/ and /ʌt/ and others may overcome the Laryngeal Input Constraint to develop the accurate representations /hɑt/ and /hʌt/, most of the participants in our study seem to have developed approximate representations based on non-linguistic markings: /ˀɑt/ and /ˀʌt/.

By extension, instead of attributing the different findings in the two EEG studies to the quality of the vowel, we conclude that the presence vs. absence of attention during the task was responsible for whether /h/ was detected: only when QFs pay attention to the signal does MMN emerge. Interestingly, this observation is consistent with QFs having an approximate representation for /h/. While actual phonemic representations are associated with automatic and effortless processing of speech, whether in perception or production, approximate representations require effort and attention. It is unusual that QFs should resort to an approximate representation for /h/, given that [spread glottis], the sole feature required to represent this phoneme, is available from their L1 inventory. According to Brown (1998), this should make /h/ relatively easy for QFs to acquire. To resolve this conundrum, we suggest the presence of a Laryngeal Input Constraint that makes representations composed exclusively of laryngeal features particularly marked and hence dispreferred. This and other constraints on underlying representation are potentially what make certain L2 segments so difficult to acquire and what make L2 learners turn to alternate forms of non-feature-based representation to solve the puzzle of L2 phonological acquisition.

Data availability statement

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

Ethics statement

The studies involving humans were approved by Comité d'éthique de la recherche de l'Université du Québec à Trois-Rivières. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

PJ: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing—original draft, Writing—review & editing. SR: Formal analysis, Methodology, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by a grant from the Entente Canada-Québec relative à l'enseignement dans la langue de la minorité et à l'enseignement des langues secondes program.

Acknowledgments

We are grateful to our research assistant, Ludovik Béchard-Tremblay, for data collection. We would also like to acknowledge the support of our colleagues in the CogNAC Research Group at UQTR.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2023.1286084/full#supplementary-material

References

Anttila, A. (1997). "Deriving variation from grammar: a study of Finnish genitives," in *Variation, change and phonological theory,* eds. F. Hinskens, R. Van Hout, and L. Wetzels (Amsterdam: John Benjamins), 935–68.

Archibald, J. (2005). Second language phonology as redeployment of LI phonological knowledge. *Canad. J. Linguist.* 50, 285–314. doi: 10.1017/S0008413100003741

Avery, P., and Rice, K. (1989). Segment structure and coronal underspecification. *Phonology* 6, 179–200. doi: 10.1017/S0952675700001007

Bassetti, B., Escudero, P., and Hayes-Harb, R. (2015). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholing.* 36, 1–6. doi: 10.1017/S0142716414000393

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

Bittner, M. (1995). *Réalisation des constrictives/ʃ/et/ʒ/en parler saguenéen: étude acoustique.* Unpublished Master's thesis, Université du Québec à Chicoutimi, Chicoutimi. doi: 10.1522/1514294

Boersma, P. (1997). "How we learn variation, optionality, and probability," in *Proceedings of the Institute of Phonetic Sciences 21* (Amsterdam: University of Amsterdam), 43–58.

Brannen, K. (2011). *The perception and production of interdental fricatives in second language acquisition.* Unpublished doctoral dissertation, McGill University.

Broselow, E., Chen, S.-I., and Wang, C. (1998). The emergence of the unmarked in second language phonology. *Stud. Second Lang. Acquisit.* 20, 261–280. doi: 10.1017/S0272263198002071

Brown, C. (1998). The role of the L1 grammar in the acquisition of L2 segmental structure. *Second Lang. Res.* 14, 136–193. doi: 10.1191/026765898669508401

Cardoso, W. (2007). The variable development of English word-final stops by Brazilian Portuguese speakers: a stochastic optimality theoretic account. *Lang. Variat. Change* 19, 219–248. doi: 10.1017/S0954394507000142

Cedergren, H. J., and Sankoff, D. (1974). Variable rules: performance as a statistical reflection of competence. *Language* 50, 333–355. doi: 10.2307/412441

Chalmers, D. J. (1996). *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Charette, M. (1991). *Conditions on Phonological Government*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511554339

Clements, G. N. (1985). The geometry of phonological features. *Phonol. Yearb.* 2, 225–252. doi: 10.1017/S0952675700000440

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

Davidson, L., Smolensky, P., and Jusczyk, P. (2004). "The initial and final states: theoretical implications and explorations of Richness of the Base," in *Constraints in phonological acquisition,* eds. R. Kager, J. Pater, and W. Zonneveld (Cambridge: Cambridge University Press), 321–368. doi: 10.1017/CBO9780511486418.011

Davis, S., and Cho, M.-H. (2003). The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics* 41, 607–652. doi: 10.1515/ling.2003.020

de Lacy, P. (2006). *Markedness – Reduction and Preservation in Phonology*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511486388

Dickerson, L. J. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quart.* 9, 401–408. doi: 10.2307/3585624

Escudero, P., Hayes-Harb, R., and Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *J. Phonet.* 36, 345–360. doi: 10.1016/j.wocn.2007.11.002

Feldscher, C., and Durvasula, K. (2017). Excrescent stops in American English. *Proc. Linguistic Soc. Am.* 2, 1–15. doi: 10.3765/plsa.v2i0.4064

File-Muriel, R. J., and Brown, E. K. (2010). The gradient nature of s-lenition in Caleño Spanish. *Lang. Variat. Change* 23, 223–243. doi: 10.1017/S0954394511000056

Flege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress,* ed. R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002

Gabriel, C., and Meisenburg, T. (2009). "Silent onsets? An optimality-theoretic approach to French *h aspiré* words," in *Variation and Gradience in Phonetics and Phonology, eds.* F. Kügler, C. Féry, and R. van de Vijver (Berlin: Mouton de Gruyter), 163–184. doi: 10.1515/9783110219326.163

Garellek, M. (2012). Glottal stops before word-initial vowels in American English: distribution and acoustic characteristics. *UCLA Work. Papers Phonet.* 110, 1–23. Available online at: https://escholarship.org/uc/item/7m55b8bb

Gouskova, M. (2023). MSCs in positional neutralization: the problem of gapped inventories. *Phonology*.

Guy, G. R. (2007). Lexical exceptions in variable phonology. *Univ. Pennsylvania Work. Papers Linguist.* 13, 109–119. Available online at: https://repository.upenn.edu/handle/20.500.14332/44657

Häcker, M. (2002). "Intrusive [h] in present-day English accents and <h>-insertion in medieval manuscripts," in *New perspectives on English historical linguistics, volume II: Lexis and transmission,* eds. C. Kay, C. Hough, and I. Wotherspoon (Amsterdam: John Benjamins), 109–123. doi: 10.1075/cilt.252.09hac

Halle, M., and Idsardi, W. J. (1997). "*r*, hypercorrection, and the Elsewhere Condition," in *Derivations and constraints in phonology,* ed. I. Roca (Oxford: Clarendon), 331–348. doi: 10.1093/oso/9780198236894.003.0010

Harris, J. (1990). Segmental complexity and phonological government. *Phonology* 7, 255–300. doi: 10.1017/S0952675700001202

Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell.

Harris, J. (1997). Licensing inheritance: an integrated theory of neutralisation. *Phonology* 14, 315–370. doi: 10.1017/S0952675797003479

Harris, J., and Kaye, J. (1990). A tale of two cities: London glottalling and New York City tapping. *Linguist. Rev.* 7, 251–274. doi: 10.1515/tlir.1990.7.3.251

Hayes-Harb, R., and Barrios, S. (2021). The influence of orthography in second language phonological acquisition. *Lang. Teach.* 54, 297–326. doi: 10.1017/S0261444820000658

Honeybone, P. (2008). "Lenition, weaking and consonantal strength: tracing concepts through the history of phonology," in *Lenition and Fortition,* eds. J. Brandão de Carvalho, T. Scheer, and P. Ségéral (Berlin: Mouton de Gruyter), 9–92. doi: 10.1515/9783110211443.1.9

Jackson, S., and Cardoso, W. (2023). "Orthographic interference in the acquisition of English /h/ by Francophones," in *Second language pronunciation,* eds. U. K. Alves and J. I. A. de Albuquerque (Berlin/Boston: De Gruyter Mouton), 229–248. doi: 10.1515/9783110736120-009

Janda, R. D., and Auger, J. (1992). "Quantitative evidence, qualitative hypercorrection, sociolinguistic variables – and French speakers' 'eadhaches with English h/Ø. *Lang. Commun.* 12, 195–236. doi: 10.1016/0271-5309(92)90015-2

Jeffreys, H. (1961). *Theory of Probability (3rd edition)*. Oxford: Oxford University Press.

John, P. (2006). *Variable h-epenthesis in the interlanguage of francophone ESL learners*. Unpublished Master's thesis, Concordia University.

John, P., and Cardoso, W. (2009). "Francophone ESL learners' difficulties with English /h/," in *Recent research in second language phonetics/phonology: Perception and production,* eds. M. A. Watkins, A. S. Rauber, and B. O. Baptista (Newcastle upon Tyne: Cambridge Scholars Publishing), 118–140.

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

Keating, P. (1988). Underspecification in phonetics. *Phonology* 5, 275–292. doi: 10.1017/S095267570000230X

Kroch, A. (1989). Reflexes of grammar in patterns of language change. *Lang. Variat. Change* 1, 199–244. doi: 10.1017/S0954394500000168

LaCharité, D., and Prévost, P. (1999). "The role of L1 and of teaching in the acquisition of English sounds by francophones," in *Proceedings of the 23rd Annual Boston University Conference on Language Development,* eds. A. Greenhill, H. Littlefield, and C. Taro (Somerville, MA: Cascadilla Press), 373–385.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Leben, W. (1973). *Suprasegmental phonology*. PhD dissertation, MIT. Distributed by Indiana University Linguistics Club.

Lindsay, G. W. (2020). Attention in psychology, neuroscience, and machine learning. *Font. Comput. Neurosci.* 14, 29. doi: 10.3389/fncom.2020.00029

Lombardi, L. (1995). Laryngeal neutralization and syllable wellformedness. *Nat. Lang. Lingu. Theory* 13, 39–74. doi: 10.1007/BF00992778

Lombardi, L. (2002). Coronal epenthesis and markedness. *Phonology* 19, 219–251. doi: 10.1017/S0952675702004323

Mah, J., Goad, H., and Steinhauer, K. (2016). Using event-related brain potentials to assess perceptibility: the case of French speakers and English. *Front. Psychol.* 7, 1469. doi: 10.3389/fpsyg.2016.01469

McCarthy, J. J. (1986). OCP effects: gemination and antigemination. *Ling. Inquiry* 17, 207–263.

McCarthy, J. J. (2002). *A Thematic Guide to Optimality Theory*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511613333

McMahon, A. (2002). *An Introduction to English Phonology*. Edinburgh: Edinburgh University Press.

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

Milroy, J. (1983). "On the sociolinguistic history of /h/ -dropping in English," in *Current topics in English historical linguistics,* eds. M. Davenport, E. Hansen, and H. Nielsen (Odense: Odense University Press), 37–53.

Morin, Y.-C. (2002). Les premiers immigrants et la prononciation du français au Québec. *Rev. Québécoise Linguist.* 31, 39–78. doi: 10.7202/006844ar

Núñez-Méndez, E. (2022). Variation in Spanish /s/: overview and new perspectives. *Languages* 7, 1–50. doi: 10.3390/languages7020077

Odden, D. (1986). On the role of the obligatory contour principle in phonological theory. *Language* 62, 353–383. doi: 10.2307/414677

Ohala, J., and Ohala, M. (1993). The phonetics of nasal phonology: theorems and data. *Phonet. Phonol* 5, 225–249. doi: 10.1016/B978-0-12-360380-7.50013-2

Orgun, C. O. (2002). English *r*-insertion in optimality theory. *Nat. Lang. Ling. Theory* 19, 737–749. doi: 10.1023/A:1013313827798

Paradis, C., and LaCharité, D. (2001). Guttural deletion in loanwords. *Phonology* 18, 255–300. doi: 10.1017/S0952675701004079

Paradis, C., and Lebel, C. (1994). Contrasts from segmental parameter settings in loanwords: core and periphery in Quebec French. *Toronto Work. Papers Ling.* 13, 75–94.

Preston, D. R. (1989). *Sociolinguistics and Second Language Acquisition*. Oxford: Blackwell.

Prince, A., and Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Hoboken: Blackwell Publishers. doi: 10.1002/9780470759400

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available online at: https://www.R-project.org/ (accessed August 23, 2023).

Rafat, Y. (2016). Orthography-induced transfer in the production of English-speaking learners of Spanish. *Lang. Learn. J.* 44, 197–213. doi: 10.1080/09571736.2013.784346

Rennicke, I. E. (2015). *Variation and change in the rhotics of Brazilian Portuguese*. Unpublished doctoral dissertation, Universidade Federal de Minas Gerais.

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Mathem. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16, 225–237. doi: 10.3758/PBR.16.2.225

Sankoff, G., and Blondeau, H. (2007). Language change across the lifespan: /r/ in montreal French. *Language* 83, 560–588. doi: 10.1353/lan.2007.0106

Smolensky, P. (1996). *The initial state and 'Richness of the Base' in optimality theory.* Technical Report JHU-CogSci-96–4, John Hopkins University, ROA-154.

Strange, W., and Shafer, V. L. (2008). "Speech perception in second language learners," in *Phonology and Second Language Acquisition,* eds. J. G. Hansen-Edwards and M. L. Zampini (Amsterdam: John Benjamins), 153–191. doi: 10.1075/sibil.36.09str

Tousignant, C. (1987). Les variantes du/R/montréalais: Contextes phonologiques favorisant leur apparition. *Rev. québécoise Linguist. Théor. Appl.* 6, 73–113.

Tranel, B. (1995). "Current issues in French phonology: liaison and position theories," in *The Handbook of Phonological Theory,* ed. J.A. Goldsmith (Oxford: Blackwell), 798–816.

Vaysman, O. (2002). Against richness of the base: evidence from Nganasan. *Ann. Meet. Berkeley Lingu. Soc.* 28, 327–338. doi: 10.3765/bls.v28i1.3848

Walker, D. C. (2001). *French Sound Structure.* Calgary, University of Calgary Press.

Wells, J. C. (1982). *Accents of English 1: An Introduction.* Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511611759

Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percept. Psychophys.* 37, 35–44. doi: 10.3758/BF03207136

Werker, J. F., and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1866–1878. doi: 10.1121/1.390988

White, E. J., Titone, D., Genesee, F., and Steinhauer, K. (2015). Phonological processing in late second language learners: the effects of proficiency and task. *Bilingualism* 20, 162–183. doi: 10.1017/S1366728915000620

Yip, M. (1988). The Obligatory Contour Principle and phonological rules: a loss of identity. *Lingu. Inquiry* 19, 65–100.

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 11 January 2024
DOI 10.3389/flang.2023.1304666

Check for updates

OPEN ACCESS

EDITED BY
Baris Kabak,
Julius Maximilian University of Würzburg, Germany

REVIEWED BY
Peter Avery,
York University, Canada
Romana Kopeckova,
University of Münster, Germany

*CORRESPONDENCE
Zsuzsanna Bárkányi
✉ zsuzsanna.barkanyi@open.ac.uk

RECEIVED 29 September 2023
ACCEPTED 06 December 2023
PUBLISHED 11 January 2024

CITATION
Bárkányi Z and G. Kiss Z (2024) Learning and unlearning voicing assimilation.
Front. Lang. Sci. 2:1304666.
doi: 10.3389/flang.2023.1304666

Learning and unlearning voicing assimilation

Zsuzsanna Bárkányi[1,2]* and Zoltán G. Kiss[3]

[1]School of Languages and Applied Linguistics, The Open University, Milton Keynes, United Kingdom, [2]Hungarian Research Centre for Linguistics, Budapest, Hungary, [3]School of English and American Studies, ELTE Eötvös Loránd University, Budapest, Hungary

This study investigates how postlexical phonological processes are acquired in multilingual speech, namely, how learners cope with conflicting demands in the production and perception of the voicing patterns in their non-native languages, what impact lexical knowledge has on learner behavior, and to what extent existing speech learning models can account for it. To investigate this, 14 Hungarian native speakers, proficient sequential learners of Spanish and English, took part in two types of experiment. The production experiments examined regressive voicing assimilation between obstruents and when the trigger was a sonorant consonant (presonorant voicing) word-internally and across word-boundary. At word level, we compared various lexical groups: non-cognates, double cognates and triple cognates (inhibitory, facilitative, and cognates with conflicting information). The perception experiments aimed to find out whether learners notice the voicing assimilations mentioned. The results showed that participants failed to learn presonorant voicing and failed to block regressive voicing assimilation despite perceiving the latter as linguistically relevant. Data also revealed that there is no direct link between perception and production, and that cognate status had a limited effect, but in triple cognates the primacy of the native language was dominant. Thus, it is concluded that in laryngeal postlexical processes the native language plays the primary role, neither the other non-native language, nor linguistic proximity seems to be decisive. Our data can be best accounted for by the Scalpel Model extended to phonological acquisition.

KEYWORDS

L3 phonology, cross-linguistic influence, L2 speech acquisition, postlexical processes, regressive voicing assimilation, presonorant voicing, Spanish, Hungarian

1 Introduction

Studies on third language (L3) acquisition aiming to determine the source and direction of cross-linguistic influence (CLI) mostly focus on morphosyntactic features and typically the early stages of acquisition. Several models have been proposed to account for the attested phenomena, but they do not agree on which language has a privileged role as a source of transfer: the native language (e.g., Hermas, 2015), or rather the second language—especially in sequential bilinguals—, which is acquired later, often in adulthood, and as such is cognitively more similar to L3 (Bardel and Sánchez, 2017); or perhaps the typologically more similar language (Rothman, 2015). It is widely accepted by now that all previously acquired languages are available for transfer (Berkes and Flynn, 2012). The Scalpel Model (Slabakova, 2017) and the Linguistic Proximity Model (Westergaard et al., 2017) advocate for both positive and negative transfer and claim that it occurs property-by-property rather than wholesale depending on which aspects of the native language (L1) or the second (non-native) language (L2) are perceived to be more similar. Discussion regarding wholesale or piecemeal transfer is still ungoing (see the 2021 special issue to *Linguistic Approaches to Bilingualism*). It is also debatable how complete the full transfer is and what exactly constitutes a property or a block of properties. The question also arises how these models can be extended to multilingual phonologies.

Previous studies on L3 phonological acquisition have identified several factors that might contribute to CLI (see Wang and Nance, 2023 for an overview), such as (perceived) typological similarity (Llama et al., 2010; Cabrelli and Pichan, 2021), and experience with L3 (Cal and Sypiańska, 2020). Many studies argue in favor of property-by-property transfer in L3 phonology. Benrabah (1991) reports that in the speech of Arabic learners of English, consonants are transferred from Arabic, while vowels form French, and this is due to the respective similarity of these subsystems. Archibald (2022) shows that stress patterns follow a mixture of influence from Arabic and French. He claims that the data he gathered can be explained by adopting a contrastive feature hierarchy model which can formally capture linguistic proximity. Wrembel et al. (2020) in a speech perception study with L1 Polish speakers, focusing on the acquisition of rhotics and final (de)voicing, also conclude that acquisition is feature-dependent. Kopečková et al. (2022) in production studies also with L1 Polish learners show that transfer comes from both previously acquired languages based on the perceived structural similarity of the examined features. Wrembel (2021) advocates for a dynamic account of CLI in L3 phonology rather than transfer from L1 or L2 only, wholesale or feature-based, since a multilingual speaker has continuous access to the previously acquired language systems.

Although there is no widely applied L3 phonological acquisition model, current well-established L2 phonology models can potentially be extended to account for L3 speech acquisition (Wrembel et al., 2019). The L2 Perception Model (Escudero, 2005) and its revised version (van Leussen and Escudero, 2015), as well as the Perceptual Assimilation Model (Best, 1994; Best and Tyler, 2007) focus on the acquisition of phonemic contrasts (a less relevant aspect in the current study). Flege's Speech Learning Model (SLM, Flege, 1995) and its revised version (SLM-r; Flege and Bohn, 2021) focus on the acquisition of single sounds, and argue that L2 speech learning is shaped by perceptual biases induced by the L1 phonetic system. The model sees CLI as an equivalence classification where perceptual objects are compared to existing L1 categories. The comparison occurs at the level of position-sensitive allophones. The model predicts that categories that are similar in L1 and L2 are more difficult to acquire as a "new category" because they are equated to an existing L1 category. Learners therefore must discover the phonetic differences and break the L2-to-L1 perceptual link in order to form a new phonetic category. Although sounds are categorized at phonetic level, representations in long-term memory are abstract, consequently, phonetic categories are used to access segment-sized units that are used to activate words (or word candidates) during lexical access. The delinking process can be speeded up by the growth of the L2 lexicon.

While SLM claimed that accurate perception precedes accurate production, in SLM-r this has been revised, and the authors claim that production and perception co-evolve, they are closely linked, and a bidirectional relationship is assumed between the two domains. Some researchers bring evidence for a third scenario, namely, that it is accurate production that precedes accurate perception rather than the other way round (e.g., Baker and Trofimovich, 2006), yet others did not find a direct link between these two domains (Derwing and Munro, 2015). Research into the relationship between perception and production for multilingual speakers is scant. Wrembel et al. (2022) in a study with 12 L1 German and 12 L1 Polish adolescent learners found that accurate perception overall precedes accurate production, but linguistic competence, the learnability of segments (e.g., articulatory difficulty of rhotics), and individual differences also play a role.

Similarly to research on L2 speech, most previous studies on L3 phonology focus on the acquisition of phonemic categories and contrasts and how beginner L3 learners categorize speech sounds based on their phonetic properties. Research on the acquisition of allophonic alternations and dynamic (postlexical) processes that create neutralisations is scarce. The present study hopes to reduce this gap by examining regressive voicing assimilation (RVA) in the speech of Hungarian learners of English and Spanish. RVA works on adjacent obstruent consonants in the speakers' L1 (Hungarian), extends to sonorant triggers in Spanish, and does not operate in English. The question is how Hungarian learners cope with these (partly) conflicting demands in their English and Spanish. Unlike in previous studies, the participants of this study are proficient in both their non-native languages. We also explore the link between speech production and perception in these laryngeal processes, and examine how existing speech learning models can account for our data.

2 Background

2.1 Voicing in Hungarian, Spanish, and English

2.1.1 Regressive voicing assimilation in Hungarian

Hungarian is a true voice language (Beckman et al., 2013) where voicing contrast of obstruents is based on negative Voice Onset Time (VOT) or voice lead in voiced stops vs. zero/short-lag VOT in voiceless stops. The language displays RVA: adjacent obstruents must agree in their voicing feature, that is, voiced obstruents voice preceding voiceless obstruents (1a); voiceless obstruents devoice preceding voiced obstruents (1b); and RVA is right-to-left iterative (1c). Hungarian has a symmetrical obstruent system with contrastive voiceless–voiced pairs at each place of articulation, thus /s/ and /z/ contrast word-initially (2a), word-finally (2b), and within the word (2c); note that /s/ in Hungarian is spelt as "sz" and /z/ is spelt as "z."

(1)

a. /tb/ → [db]: *hát-ba* 'back.ILL';
két barát 'two friends'
/ʃb/ → [ʒb]: *has-ba* 'stomach.ILL';
hús bevezetése 'introduction of meat'

b. /bt/ → [pt]: *láb-tól* 'foot.ABL';
láb tünetei 'symptoms of foot'
/zt/ → [st]: *víz-től* 'water.ABL';
víz tárolása 'storing of water'

c. /skb/ → [zgb]: *maszk-ban* 'mask.INESS'

(2)

a. *szár* 'stem' vs. *zár* 'lock'
b. *mész* 'limestone' vs. *méz* 'honey'
c. *másznak* 'climb.PL.3.PRES' vs. *máznak* 'gloss.DAT'

Unlike in many surrounding languages (e.g., German, Slovak), word-final obstruents do not devoice in Hungarian (3). Sonorant consonants do not participate in RVA: obstruents maintain the voicing contrast before sonorants both within the word (4a) and across a word boundary (4b).

(3)
láb-ak [b] 'foot.PL' ∼ *láb* [b] 'foot'
láp-ok [p] 'marshland.PL' ∼ *láp* [p] 'marshland'
méz-ek [z] 'honey.PL' ∼ *méz* [z] 'honey'
mesz-ek [s] 'limestone.PL' ∼ *mész* [s] 'limestone'

(4)
a. *plakát* [pl] 'poster', *blöki* [bl] 'doggy', *sróf* [ʃr] 'screw', *zrí* [zr] 'fuss'
kész-nek [sn] 'ready.DAT', *kéz-nek* [zn] 'hand.DAT'
b. /tm/→ [tm] (*[dm]): *két mag* 'two seeds'
/sl/→ [sl] (*[zl]): *kész leves* 'ready soup'

According to the traditional generative literature, RVA in Hungarian is categorical, exceptionless, and completely neutralizing (Vago, 1980; Siptár and Törkenczy, 2000), which means that voiceless and devoiced or contextually voiced and underlyingly voiced segments cannot be distinguished on the basis of their phonetic and phonological behavior. More recent acoustic phonetic studies, however, suggest that neutralization might be incomplete with residual traces of the underlying voice feature of the obstruents (e.g., Jansen, 2004; Bárkányi and G. Kiss, 2015).

2.1.2 /s/-voicing in Spanish

Spanish, belongs to the same broader typological group as Hungarian, as it is also a true voice language, where stop phonemes can be either voiced or voiceless, although voiced stops are often realized as voiced approximants (unlike in Hungarian), and fricatives and affricates do not display such a symmetry (they are voiceless, except the palatal fricative).

Even though Spanish has RVA, because of the phonotactic restrictions of the language, the segment undergoing assimilation is mostly /s/. Spanish /s/-voicing presents a special case within RVA languages since there is no alveolar voiced fricative phoneme in the language. The Central-Northern Peninsular variety has two voiceless sibilant fricatives, an interdental /θ/ and an apico-alveolar /s/. All the other varieties have only one sibilant fricative /s/, which has a wide range of dialectal and individual realizations from apical to laminal, interdental, etc. (Quilis, 1993). Spanish clearly shows a preference for open syllables—coda obstruents are fragile in the language and there is high variability in their realizations (e.g., Hualde, 2005)—, therefore /s/-voicing in Spanish only occurs in dialects where syllable-final /s/ rarely undergoes aspiration and deletion. In these varieties when /s/ is followed by a voiced consonant—a voiced obstruent (5a) or a sonorant (5b), including glides (5c), within the same word or across a word-boundary—, /s/ becomes partially or fully voiced (Hualde, 2005). Importantly, in Hungarian there is no presonorant voicing (PSV) as in (5b).

(5)
a. *esbelto* [zβ] 'slim', *es bueno* [zβ] 'it's good'
b. *isla* [zl] 'island', *es largo* [zl] 'it's long'
c. *deshielo* [zj] 'thaw', *los hielos* [zj] 'the ices'

TABLE 1 Summary of the three laryngeal systems.

	Hungarian	English	Spanish
Type of laryngeal contrast	**True voice**	Aspirating	**True voice**
Laryngeal contrast within the obstruent inventory	**Symmetrical**	**Symmetrical**	**Symmetrical** (but limited to stops only)
RVA	**Yes**	No	**Yes**
PSV	**No**	**No**	Yes

Similarities highlighted in bold.

Most phonologists who studied /s/-voicing in Spanish found high degrees of individual variation (Schmidt and Willis, 2011), and claim that the process is gradual (e.g., Campos-Astorkiza, 2015), or that gradient data is the result of categorical but optional assimilation (Bárkányi, 2014). Note that although RVA is very common in true voice languages, PSV is much less frequent than preobstruent voicing, prosodic restrictions also seem to apply (Bárkányi and G. Kiss, 2015), and the phenomenon is viewed by some researchers as a result of extended passive voicing as opposed to the spreading of voicing, as in RVA (Jansen, 2004; Strycharczuk, 2012).

2.1.3 The voicing pattern of English

English, just like Hungarian, displays a symmetrical laryngeal obstruent system, but unlike Hungarian and Spanish, English is an aspirating language (Lisker and Abramson, 1964), that is, the contrast of stops is based on aspiration rather than voicing. "Voiced" stops, or as generally referred to in the phonological literature, lenis stops (in initial position) are produced with zero or short-lag VOT, thus phonetically they are typically voiceless and unaspirated, while voiceless, or fortis, stops are produced prevocalically with a relatively long-lag VOT (i.e., aspirated). In contrast to true voice languages, in English no systematic laryngeal spreading, i.e., RVA, is attested (Jansen, 2004; Szigetvári, 2020; see (6a)). Similarly to Hungarian, English does not have presonorant voicing either (6b).

(6)
a. *matchbox* [tʃb] (*[dʒb]); *anecdote* [kd] (*[gd]);
baseball [sb] (*[zb]); *bonus deal* [sd] (*[zd])
b. *disloyal* [sl] (*[zl]); *mismatch* [sm] (*[zm]);
business model [sm] (*[zm])

We summarize the relevant features of the three laryngeal systems in Table 1.

2.2 Voicing in multilingual studies

Most studies that deal with the acquisition of laryngeal features in L3 focus on the phonetic realization of voiceless stops, usually by measuring VOT in the speech of multilingual learners. While these studies tested different groups of trilingual speakers (heritage speakers, beginner L3 learners, advanced L3 learners) and employed different methodologies (reading, picture naming;

monolingual sessions, bilingual sessions, etc.) to find out whether learners created separate phonetic categories in their languages, no prevalent conclusions emerged, although realizations similar to L1 were more likely. Wunder (2011) based on the production of voiceless stops /p t k/ in the speech of eight L1 German, L2 English and L3 Spanish speakers found mainly L1 effects on the L3, but more importantly, in half of the cases, tokens displayed values in between the two languages. Llama and Cardoso (2018) and Llama and López-Morelos (2016) also found that L1 plays a more decisive role in L3 pronunciation, but these authors claim that language proficiency and language dominance are also significant factors. Amengual (2021) only examined the acoustic realization of /k/ in bilingual (L1 English–L2 Japanese and L1 Japanese–L2 English) and trilingual (L1 Spanish, L2 English and L3 Japanese) groups. This study also found that VOT values are closer to the L1 values than the target realizations of each language, but also demonstrated that participants produced language-specific VOT patterns which were influenced by language mode and cognate status (see Section 2.3). Very few studies examine perception, or link perception and production. A notable exception is Liu and Lin (2021), who claim that there is no direct link between the perception and production of voicing. In a study with 39 L1 Mandarin Chinese, L2 English and L3 Japanese or Russian learners, where participants had to carry out a reading task and a phoneme identification task targeting voiced and voiceless stops, Liu and Lin (2021) found that voiced and voiceless stops did not behave in the same way. There was a positive correlation between the perception and production of L3 voiceless stops in the initial stages of acquisition, but no correlation was found between the perception and production of L3 voiced stops. This means that, in perception, phonetic similarity led to confusion as predicted by SLM-r, but pre-voicing in stops—which was a novel phonetic feature for these L3 learners—was easily perceived. Participants, on the other hand, had difficulty producing voicing lead. Note that in these cases learners had to map an existing phonemic contrast (voiced vs. voiceless stops) in their L1 to a phonetically different voiced–voiceless contrast.

A different scenario is when learners have to acquire or block a phonetically very different allophone. Cabrelli and Pichan (2021) in a study focusing on the production of intervocalic voiced stops in the speech of early and late bilingual (L1 English–L2 Spanish and L1 Spanish–L2 English) L3 Brazilian Portuguese and Italian learners found an overall trend toward transfer of Spanish-like [+continuant] segments into the typologically similar Romance L3. The authors conclude that transfer was determined by global similarity between L3 and the source language (Spanish) despite this being non-facilitative. Note, however, that almost half of the realizations were either produced with a stop or partial stop closure, which is not fully compatible with the Typological Primacy Model (Rothman, 2015).

Studies on "feature changing" phonological processes where no new phonetic category is created like word-final devoicing (or the lack of it) also seem to indicate that L1 is hard to overcome. Kopečková et al. (2022) in a delayed repetition task with beginner L3 learners (L1 German, L2 English, L3 Polish and L1 Polish, L2 English, L3 German) found that more than half of the realizations showed L1 influence (the realizations were basically identical in all three languages), while the other half was some other sound substitution. The authors, however, did not measure the amount and proportion of voicing in the final obstruent; rather, they classified obstruents into three categories (voiced, voiceless and partially voiced). Furthermore, the research design did not control for RVA which in several instances could block word-final devoicing. Thus, these results should be interpreted with caution. Wrembel et al. (2020) arrived at a similar conclusion in a perception study with 13 L1 Polish, L2 English, L3 German teenagers. The authors found no significant development in either L2 English or L3 German in the perception accuracy and processing speed of word-final obstruent (de)voicing.

Similarly to word-final devoicing, RVA does not create a new segment either, i.e., learners do not face any articulatory difficulty, but they have to implement or block a postlexical phonological process that applies across the board. As far as we are aware, no studies deal with RVA in L3. Darcy et al. (2007) with French and American English speakers found that L1 English speakers compensated less, i.e., showed lower detection rates for items undergoing RVA (a phonological process absent in their language) than L1 French students. The authors observed that participants compensated more for devoicing, that is, they recognized a voiced phoneme that was realized as voiceless better than a voiceless phoneme realized as voiced, which could be a result of partial word-final devoicing occurring in English (Keating, 1984).

2.3 Cognate status effect and voicing

The facilitation effects of cognates have been extensively studied in psycholinguistics (see Amengual, 2012 for an overview). Research has consistently shown that reaction times are faster for cognates compared to non-cognates, they exhibit quicker and more accurate lexical access, display greater repetition priming effects, and are easier to learn. The detection of the cognate status of words leads to the formation of lexical connections (Ecke and Hall, 2021), which affects the morphosyntactic specification, meaning, as well as speech production of the new lexical item.

Previous research on bilingual speech indicates that the similarity of lexical items—considerable phonological and semantic overlap—might impact on the acoustic realization of segments within them (e.g., Mora and Nadeu, 2009; Amengual, 2016). Studies examining the possible cognate effects in the production of VOT give mixed results. While Flege and Munro (1994) found that English cognates in Spanish were pronounced with longer VOT values than non-cognates, Flege et al. (1998) did not replicate the same results. Amengual (2012), on the other hand, did find a significant effect of cognate status in the speech of bilinguals of different levels of competence. His participants produced a phonetic shift toward the non-target language, they produced /t/ with longer (more English-like) VOT values in the Spanish production of cognates compared to non-cognate words. The author explains this in the framework of the exemplar model of lexical representation (Bybee, 2001; Pierrehumbert, 2001) according to which, due to bilingual lexical connections, cognates facilitate phonetic interference in the bilingual mental lexicon. Amengual (2021) extended the study to trilingual L1 Spanish, L2 English and L3 Japanese learners and observed that although speakers produced different VOT values in the three languages, when the session was bi- or trilingual, speakers transferred non-target-like phonetic characteristics more than in monolingual sessions. Also, learners in their least dominant language (Japanese) produced lengthened, more English-like, VOT values which the author sees as a transient CLI.

2.4 Hypotheses

The current study aims to address how multilingual speakers handle the conflicting cross-linguistic influences on RVA and PSV in their speech productions. In order to determine the source of CLI, cognates and non-cognates are compared, and to examine the dynamic aspect of these assimilations, data from sandhi contexts (across a word-boundary) are compared to within-the-word realizations.

The following hypotheses are tested in the study:

Hypothesis 1: Inhibitory cognates are realized with voicing properties less similar to those in the target language than non-cognates.
Hypothesis 2: Facilitative cognates are more likely to be realized with target-like voicing properties than non-cognates.
Hypothesis 3: When cognates are contradictory, e.g., L1 is facilitative but L2 is inhibitory, or L1 is inhibitory but L2 is facilitative, it is the L1 pattern that dominates.
Hypothesis 4: Sonorants do not trigger voicing assimilation in sandhi in either Spanish (non-target-like) or English (target-like).
Hypothesis 5: Obstruents trigger RVA in sandhi contexts in both Spanish (target-like) and English (non-target-like).

3 Materials and methods

The production experiment aimed to investigate the proportion of voicing in the alveolar fricative in regressive voicing assimilation contexts, including presonorant voicing, in the speech of Hungarian learners of English and Spanish.

3.1 Participants

Fourteen young adult subjects (five male, nine female) participated in the experiment. Their ages ranged between 19 and 25 years (average 21.6). They were rewarded a voucher of 5,000 HUF for their participation in the experiment. All the subjects were students majoring in Spanish language and literature at Eötvös Loránd University, Budapest. They were all native speakers of Hungarian who started learning English and Spanish past 11 years of age. Their proficiency in both languages was at least B2 of the Common European Framework of Reference for Languages as they all successfully passed both English and Spanish proficiency exams administered by the University as part of their studies, but none of them spent more than 3 months in an English-speaking or Spanish-speaking country. This means that they were all proficient sequential trilingual speakers acquiring their L2 and L3 in a non-immersion context. Although they started learning English before Spanish, as both these interlanguages underwent development at the same time and there is no clear dominance difference between English and Spanish for these participants, in this paper we refer to both English and Spanish as L2 or L3. Eleven subjects speak another Romance language, but they consider themselves less proficient in this additional language than in English and Spanish. All of them claim to speak the Peninsular (Northern-Central Peninsular) variety of Spanish; 4 identify with American English, 6 with British English, and 4 claim to speak a mixed variety. None of the participants reported any speaking or hearing disorder.

3.2 Materials

In the Spanish part of the production experiment the target segment was /s/, while in the English part it was /s/ in sandhi context and both /s/ and /z/ within the word, in the following positions (see the Supplementary material for a complete list of test sentences and the cognate coding of test words, Supplementary Tables 9, 10):

Sandhi

- Word-finally before a voiced stop /b d g/ across word-boundary: e.g., SP *coche**s d**uros* 'tough cars;' ENG *bonu**s d**eal.*
- Word-finally before a sonorant consonant /m n l/ across word-boundary: e.g., SP *casa**s m**odernas* 'modern houses;' ENG *busine**ss m**odel.*

Word-internally, in triple cognate words

- In facilitative cognates before a voiced stop /b d g/, e.g., SP *le**sb**iana*, ENG *le**sb**ian.*
- In facilitative cognates before a sonorant consonant /m n l/, e.g., SP *pla**sm**a*, ENG *pla**sm**a.*
- In inhibitory cognates before a voiced stop /b d g/, e.g., ENG *ba**seb**all.*
- In inhibitory cognates before a sonorant consonant /m n l/ e.g., SP *e**sn**obismo*, ENG *Ya**sm**in.*
- In L1 facilitative and L2 inhibitory before a voiced stop /b d g/, e.g., SP *béi**sb**ol.*
- In L1 facilitative and L2 inhibitory before a sonorant consonant /l m n/, e.g., SP *Ya**sm**in*; ENG ***sn**ob.*
- In L2 facilitative and L1 inhibitory before a sonorant consonant /l m n/, e.g., SP *Bo**sn**ia*; ENG *Bo**sn**ia.*

Word-internally, in double cognate words English–Spanish

- In facilitative cognates before a voiced stop /b d g/, e.g., SP *Ra**sg**ora*, ENG *A**sb**ora.*
- In facilitative cognates before a sonorant consonant /m n l/, e.g., SP *fanta**sm**a*, ENG *phanta**sm**al.*
- In inhibitory cognates before a voiced stop /b d g/, e.g., SP ***desd**én*, ENG *di**sd**ain.*

- In inhibitory cognates before a sonorant consonant /m n l/ e.g., SP *des**l**eal*, ENG *dis**l**oyal.*

Word-internally, in non-cognate words

- Before a voiced stop /b d g/, e.g., SP ***esbelto*** 'slim', ENG /s/ *cros**sb**ar*, /z/ *hus**b**and.*
- Before a sonorant consonant /m n l/, e.g., SP ***asno*** 'donkey', ENG /s/ *Chris**tm**as*, /z/ *rose**m**ary.*

The selection of cognates presented a number of challenges. It is not easy to determine the degree of similarity (be it orthographic, phonological or semantic) that lexical items have to show in order to induce a cognateness effect. It is also important that the words are relatively frequent in all the languages so that learners are familiar with them. Furthermore, some facilitative or inhibitory combinations were logically impossible. For instance, the /s/+voiced stop sequence in both Hungarian and Spanish trigger RVA thus L1 is always facilitative in L2/L3 Spanish and always inhibitory in L2/L3 English (furthermore in English, these sequences are mostly heteromorphemic). Other clusters should be possible, but we could not find lexical items like facilitative dual cognates for /s/+voiced stop. In these cases, we used an invented proper name in a carrier sentence that suggested that the test word was a loan from the other language. In order to get a full picture of cognate status effect across these three languages, all three dual combinations (HU–ENG, HU–SP, and ENG–SP) should have been tested. However, with Hungarian not being an Indo-European language, there are not enough lexical items with the required segment sequences that are cognates with only one of the other two languages (note that in the sandhi context cognateness was not controlled for, the stimuli were all non-cognates).

Stimuli were embedded into 10–13-syllable-long neutral declarative carrier sentences. They occurred in the first half of the sentence, but were not sentence initial.

3.3 Method of the production experiments

As language mode might have an effect on the acoustic realization of sounds (Amengual, 2021) and language mode was not a variable tested in the present research, English and Spanish sessions were kept separate. Half of the participants started with the English session and the other half with the Spanish session. Before each session they had to read a few sentences in Hungarian to adjust microphone settings and to make sure learners did not have any speech disorder. As sessions had to be recorded on the same day, after the first session participants had a lunch break and came back for the second session which again started with the sentences in Hungarian. As a reviewer pointed out this might have had a risk of L1 priming, a limitation that must be kept in mind when interpreting the data. Sessions took place in the soundproof booth of the Hungarian Research Center for Linguistics.

Sentences and fillers (which formed part of another experiment) were read from a monitor screen in a randomized order, which was generated by SpeechRecorder (Draxler and Jänsch, 2004). Each test sentence was read four times. This meant 44 sentences for the English data and 39 sentences for the Spanish one by four repetitions by 14 speakers.

3.4 Measurements

The acoustic analysis was carried out in Praat (version 6.2.23; Boersma and Weenink, 2022). The spectrograms were segmented manually by the authors and a research assistant, and the following measurements were carried out on the basis of the boundaries inserted:

- Duration of the target consonant /s/ or /z/.
- Absolute length of the voiced interval.
- Ratio of the voiced part compared to the total length of the consonant.

In addition to voicing (vocal fold vibration) in the strict sense, a number of other systematically occurring phonetic-acoustic correlates of voicing contrast are attested in the literature. Voiced obstruents are generally shorter than their voiceless counterpart, and variation is also attested in the surrounding sounds. While voiceless consonants are longer, the preceding vowel is typically shorter (Wells, 1982). In the current research, the cognate status of test words had to be taken into account, so we could not control for the quality of the vowel preceding the fricative; therefore, only the proportion of voicing compared to the fricative interval was measured.

The duration of the fricative was determined on the basis of the frication noise. Voicing was measured based on the visual inspection of the spectrograms and oscillograms, and a low-pass filter with a cut-off frequency of 500 Hz was used to securely determine the exact portion of the voicing oscillation.

3.5 Perception experiment

In order to explore to what extent the production results are mirrored in perception, a perception experiment was designed. As most L2 perception studies aim to determine whether learners acquired a certain contrast, they often apply forced-choice tests where participants have to decide whether they hear phoneme *A* or phoneme *B*. However, we did not find it adequate in the present study. Firstly, because listeners might compensate for RVA (Kuzla et al., 2010; Bárkányi and G. Kiss, 2023). Secondly, because they might be biased against a segment that does not form part of the phoneme inventory, that is, they might be reluctant to choose [z] for Spanish; or they might be biased by the orthographic form of test words. As speakers are often unaware of the application of postlexical processes even in their L1, we wanted to leave it open that learners would like to respond "I don't know," or that they simply cannot perceive the processes under scrutiny at all or as a linguistically relevant feature. For this reason, we decided to test the perception of RVA and PSV in a more holistic way. A short (approximately one-minute long) story was recorded in both L2/L3 by two phonetically trained bilingual female speakers with native-like proficiency in both languages (Hungarian–English and

Hungarian–Spanish, respectively). Then, the same short story was recorded, but this time in the English text RVA was applied as would be in Hungarian, and in the Spanish text no pre-sonorant voicing was employed to mirror the L1 laryngeal patterns of listeners (another recording was made of the same story as a distractor that formed part of another experiment). Thus, there were three slightly different texts and participants listened to each text three times in a random order, so they had to listen to nine texts all together.

The experiment was carried out in Praat MFC with the same participants as in the production experiments, at the end of the production sessions. The screen was blank while participants listened to a text. Texts were separated by 1.5 s of silence and a 5-second-long bell. After the text finished, instructions appeared on the screen and participants had to rank on a scale from 1 to 5 how native-like the speaker sounded (with 1 not at all native-like and 5 completely native-like).

3.6 Statistical analysis

The statistical analysis (including the generation of the various plots and data tables) was carried out in R (R Core Team, 2022) using various *tidyverse* packages (Wickham et al., 2019). Linear mixed effects models were used to model the production data, using the packages *lm4* (Bates et al., 2015), *lmerTest* (Kuznetsova et al., 2017), and *broom.mixed* (Bolker and Robinson, 2022). The model function used the default settings (e.g., the Satterthwaite approximation was used to calculate the degrees of freedom for the *t*-distributions). The outcome variable was always the percent of voicing during the fricative constriction. The predictor variables were different depending on the phonological environment of the fricative. The effect of the *cognate status* was analyzed in the *word-internal* data, separately for the English and Spanish words, and separately for the two triggers (sonorants and voiced stops; for details see below). This is because the cognate groups were necessarily different by language and/or phonological environment. The contrast coding of the cognate status factor used the default "dummy" coding (i.e., each cognate group was compared to the non-cognate words). It was possible to analyse the effect of the *trigger environment* (presonorant vs. before voiced stop), the *language* (English vs. Spanish words) as well as their interaction in the *sandhi* environment, i.e., when /s/ was word-final and the trigger was at the beginning of a following word. In the sandhi environment, planned orthogonal contrast codings were used for the fixed-effect predictors, so that their main effects (the estimated marginal means) can be calculated and interpreted more easily, in addition to their interaction effect. The random-effect structure of the models contained subject and item (i.e., the words used in the experiments). Which exact model was used in which analysis will be detailed in the results section below. The best-fitting model was selected after carrying out model comparisons employing likelihood ratio tests (using maximum likelihood). A model was retained if the chi-square test was significant. The same procedure was used to test the utility of the random effects, except that in this case restricted maximum likelihood was used in the likelihood ratio tests. Pairwise comparisons using Bonferroni adjustment to the *p*-values were carried out using the *emmeans* package (Lenth, 2023), while R-squared effect sizes were calculated with the help of the *r2* function of the *performance* package (Lüdecke et al., 2021). We are going to abbreviate conditional R-squared as "R2c", and marginal R-squared as "R2m."

The perception experiment fitted cumulative link mixed models to the data using the package *ordinal* (Christensen, 2022). The outcome variable was the rating of nativeness by the participants, which was on an ordinal scale (1, 2, 3, 4, 5). The fixed predictor variable was the type of the recording they listened to (native-like vs. non-native-like). The models included subject as the random effect. The link function used was probit, as it is considered to be more suitable than the logit link function in models that contain random effects (Hahn and Soyer, 2005). The best-fitting model was selected using the same principles and procedure as in the case of the linear mixed effects models used in the analysis of the production data.

4 Results

4.1 Production experiments

4.1.1 Cognate status effect before sonorants

4.1.1.1 English /s/ before sonorants

Figure 1 shows the mean voicing percentage of presonorant /s/ in four cognate groups, the descriptive statistics can be found in Table 2. The negative sign "–" refers to inhibitory cognate status, while "+" refers to facilitative cognate status. Thus, for example, "SP–HU+" refers to English words that have cognates both in Spanish and Hungarian, but in Spanish the /s/ is realized with voicing, so it can potentially voice English /s/. However, as we can see, /s/ had little voicing across the groups, the SP–HU– group showed the highest average voicing at 22.8%.

The best-fitting linear mixed effects model was one with varying by-subject intercepts and slopes, and varying by-item intercepts. According to this model, none of the three cognate groups were significantly different from the non-cognate group. Pairwise comparisons resulted in none of the groups being significantly different from each other, either. Details of the model coefficients and the effect sizes are in Supplementary Table 1. We note that the word that was responsible for the slight increase of voicing in the SP–HU– group was *Yasmin*, with a mean voicing of 35.1% (SD = 23.8%).

4.1.1.2 Presonorant /z/ in English words

The mean voicing of /z/ in the cognate groups is displayed in Figure 2, the descriptive statistics are shown in Table 2. We can see that there is considerable voicing in /z/ in all groups, except in SP+HU–, i.e., when the word has a cognate in Hungarian pronounced as voiceless [s].

The word *Bosnia* contributed to the cognate effect in the SP+HU– the most: the fricative in this word was produced by the participants with a mean of only 13.9% (SD = 11.7%), there were no tokens above 54% of voicing at all.

The best-fitting model for this data was the one with varying by-subject intercepts and slopes, and varying by-item intercepts. According to this model, the difference between the non-cognate

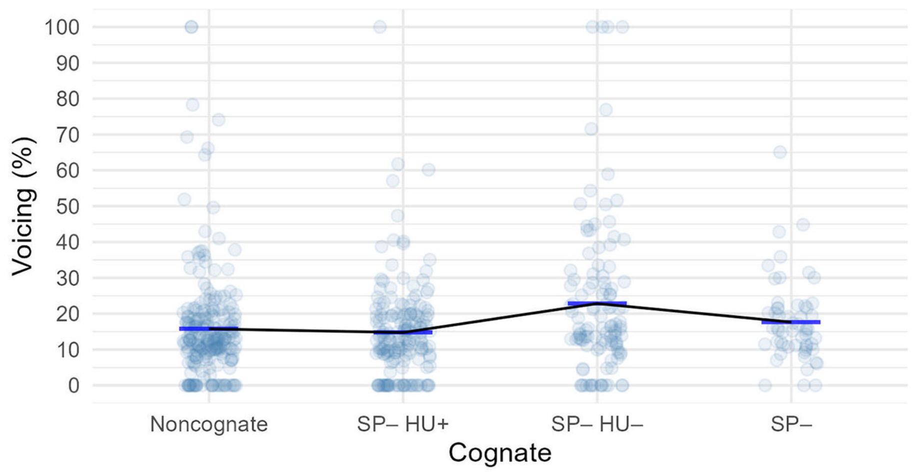


FIGURE 1
Voicing of /s/ before sonorants in the English words. Lines represent the means.

TABLE 2 Descriptive statistics for the voicing of presonorant /s/ and /z/ in English words in the cognate groups.

Sound	Cognate	*N*	Mean	SD	Median	Min	Max
/s/	Non-cognate	224	15.75	15.12	12.83	0	100
	SP–HU+	168	14.73	13.34	12.71	0	100
	SP–HU–	112	22.81	21.58	15.97	0	100
	SP–	56	17.63	11.53	15.56	0	65.06
/z/	Non-cognate	224	52.57	31.13	42.71	0	100
	SP+HU+	112	41.44	31.33	31.24	0	100
	SP+HU–	112	22.31	21.52	16.40	0	100
	SP+	112	49.95	34.05	35.82	6.62	100

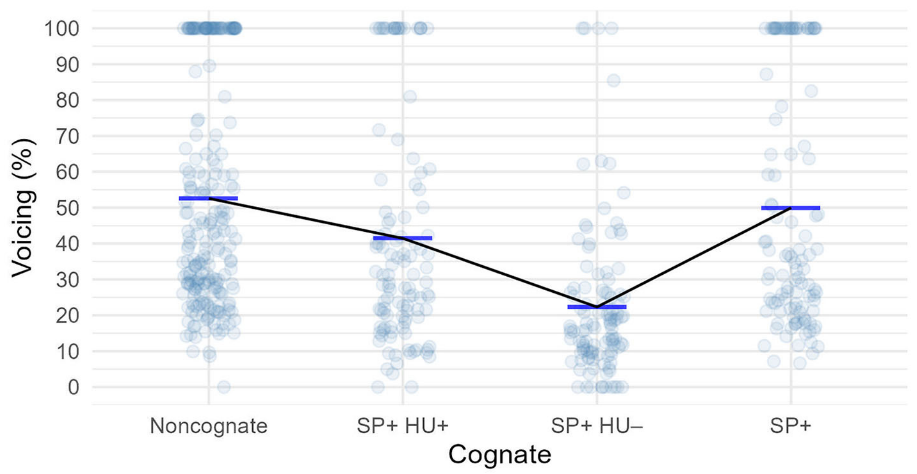


FIGURE 2
Voicing of /z/ before sonorants in the English words. Lines represent the means.

words and the SP+HU– cognates was statistically significant (Supplementary Table 2). Pairwise comparisons did not uncover further significant group differences.

4.1.1.3 Spanish words with presonorant /s/

The fricative remained relatively voiceless across all groups in the Spanish words (Figure 3; mean voicing ranged between 12.1

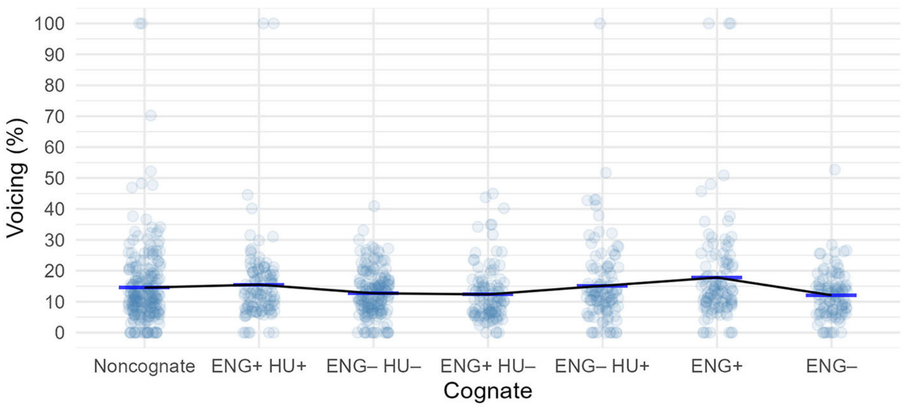


FIGURE 3
Voicing of /s/ before sonorants in the Spanish words. Lines represent the means.

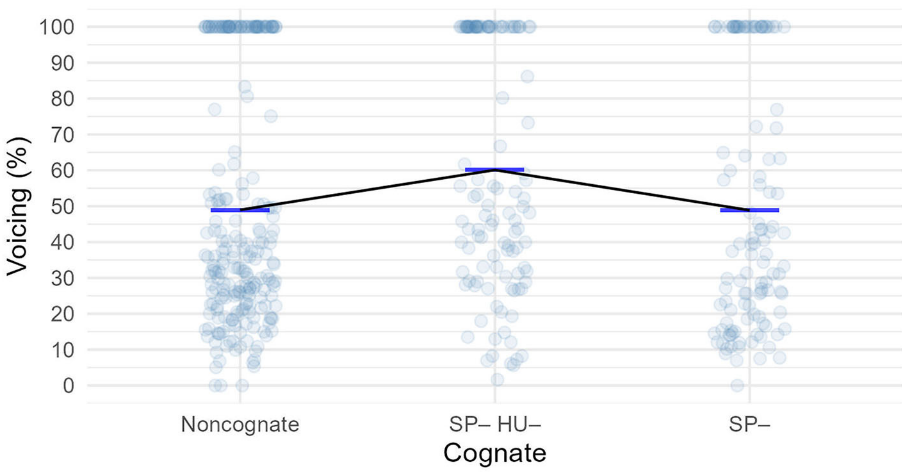


FIGURE 4
Voicing of /s/ before voiced stops in the English words. "SP–HU–" = "*baseball*"; lines represent the means.

and 17.8%). Three groups showed a small amount of increase in the mean voicing: ENG+HU+, ENG–HU+, and ENG+, which had the highest mean at 17.8 (still relatively little voicing though).

The best-fitting linear mixed effects model for the Spanish presonorant data was the one with varying by-subject intercepts and varying by-item intercepts. According to this model, there was no statistically significant difference between the non-cognate words and any of the cognate groups, i.e., we cannot observe any cognate status effect (Supplementary Table 3). Pairwise comparisons did not uncover any significant group differences either.

4.1.2 Cognate status effect before voiced stops

4.1.2.1 English words containing /s/ plus voiced stops

As we can see (Figure 4, Table 3), the "doubly" inhibitory cognate group (SP–HU–) showed the most average voicing (60.1%), but the fricative contained a fair amount of voicing in the other groups, too (close to 50% on average). We note that this group only included one word, *baseball*.

TABLE 3 Descriptive statistics for the voicing of /s/ before voiced stops in English words in the cognate groups.

Cognate	*N*	Mean	SD	Median	Min	Max
Non-cognate	224	48.96	33.67	35.77	0.00	100
SP–HU–	112	60.07	33.68	51.19	1.62	100
SP–	112	48.86	34.68	37.02	0.00	100

This data was modeled with a linear mixed effects model that contained by-subject and by-item varying intercepts. According to this model, neither of the cognate groups had a significantly different amount of voicing compared to the non-cognate words (Supplementary Table 4). Neither did the pairwise comparisons show a significant difference between the groups.

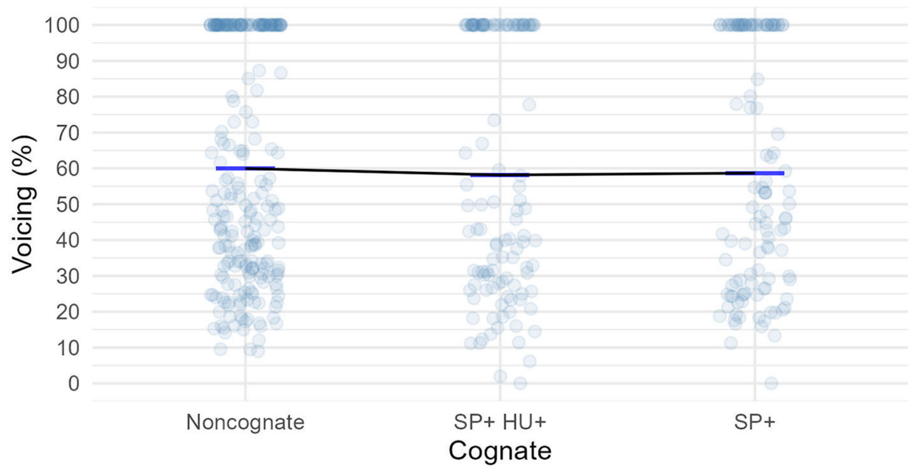


FIGURE 5
Voicing of /z/ before voiced stops in the English words. Lines represent the means.

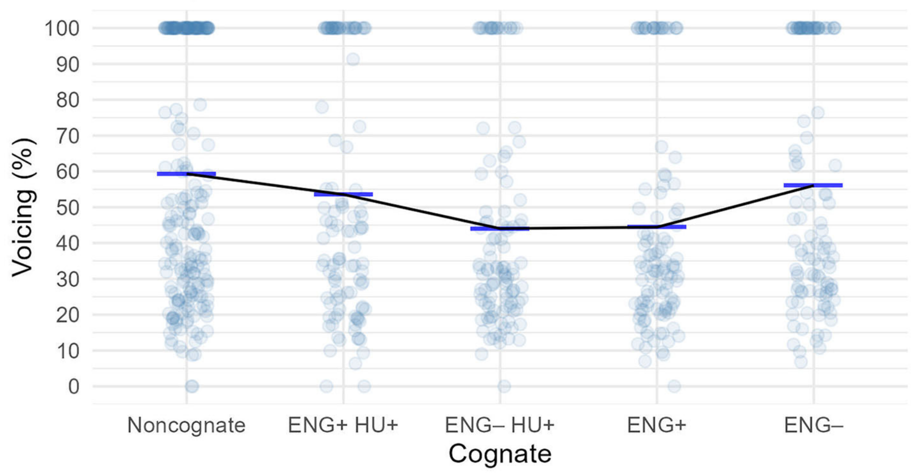


FIGURE 6
Voicing of /s/ before stops in the Spanish words. Lines represent the means.

4.1.2.2 /z/ before voiced stops in English words

As Figure 5 shows, the fricative contained a fair amount of voicing across all three groups (around 60%) in this case.

The best model fitted to the data was the one with by-subject intercepts and slopes, and by-item varying items. As expected, the model did not uncover any significant differences between the cognate groups and the non-cognate group, or between the cognate groups (Supplementary Table 5).

4.1.2.3 /s/ before voiced stops in Spanish

Just like for English /s/ and /z/, /s/ in the Spanish words was articulated with a considerable amount of voicing (between 44 and 60% on average; see Figure 6, Table 4).

The best-fitting model for the Spanish data was the one in which the intercepts were allowed to vary for both subjects and items, but not the slopes. According to this model, voicing in none of the cognate groups was significantly different from that in the non-cognate group. Pairwise comparisons did not uncover significant difference between any of the groups (Supplementary Table 6).

TABLE 4 Descriptive statistics for the voicing of /s/ before voiced stops in Spanish words in the cognate groups.

Cognate	*N*	Mean	SD	Median	Min	Max
Non-cognate	224	59.30	34.06	50.12	0.00	100
ENG+HU+	112	53.55	34.17	44.10	0.00	100
ENG–HU+	111	44.04	29.67	32.69	0.00	100
ENG+	112	44.43	30.89	33.12	0.00	100
ENG–	112	56.04	33.60	41.24	6.83	100

4.1.3 PSV and RVA across a word boundary

Now we turn to the results of the production data that involved the voicing of word-final /s/ followed by a word that began with a sonorant consonant or a voiced stop, i.e., PSV and RVA across a word boundary, in English and Spanish words. The mean voicing percentages can be seen in Figure 7 (the descriptive statistics are tabulated in Table 5). We can see that /s/ had little voicing before sonorants on average (in the Spanish words mean voicing was somewhat higher but the difference was not significant according to pairwise comparisons), whereas before voiced stops it was far more voiced, especially in the Spanish words in which /s/ contained 81% voicing on average. These results indicate the general absence of PSV and a strong RVA effect in both English and Spanish words.

The best-fitting linear mixed effects model contained trigger (sonorant vs. voiced stop), language (English vs. Spanish), as well as their interaction as fixed-effect predictors, and by-subject varying intercepts and correlated slopes for both trigger and language (but not their interaction), and by-item varying intercepts as random effects. According to this model, the main effect of *trigger* (i.e., the difference between the estimated marginal mean voicing of final /s/ before sonorants and voiced stops) was statistically significant: voicing of /s/ word-finally was significantly greater before the voiced stops compared to before the sonorants. The main effect of *language* (i.e., the difference between the estimated marginal mean voicing in English and Spanish words) was also statistically significant: /s/ had significantly more overall voicing in Spanish than in English. And finally, the *interaction* term was also statistically significant: the difference between PSV and RVA of final /s/ was significantly greater in Spanish words than in English words. Supplementary Table 7 provides a summary of the model.

4.2 Word-internal vs. word-final /s/

We also compared the voicing of word-internal and word-final /s/ in the English and Spanish words in the two environments (before sonorants and before voiced stops, see Figure 8, Table 6). Since the cognate status was not controlled for in the word-final position, the word-internal group only contained the non-cognate words in this comparison. As we can see, before sonorants, the mean voicing of /s/ is rather similar in the two environments (word-internal: 15.2%, SD = 14.2%; word-final: 18.6%, SD = 11.1%). The mean voicing of /s/ before voiced stops is, however, much higher in both environments (word-internal: 54%, SD = 34.2%; word-final: 66.7%, SD = 34.4%). In addition to this, we can again observe an interaction effect of the language of the words before voiced stops: /s/ had much more voicing word-finally than word-internally in the Spanish words compared to the English words.

We fitted two linear mixed effects models separately for the two trigger sounds (sonorants vs. voiced stops). The best model in both cases was the one that included environment (word-internal vs. word-final), language (English vs. Spanish), as well as their interaction as fixed-effect predictors, and by-subject varying intercepts and correlated slopes for both environment and language, and their interaction. When the trigger sound was a sonorant, the main effects of environment and language were not statistically significant; however, there was a significant interaction effect: the difference between the mean voicing of /s/ word-internally vs. word-finally was greater in the Spanish words. When the trigger was a voiced consonant, the main effects of environment and language, as well as their interaction, were statistically significant: i.e., overall, there was more voicing in /s/ word-finally, and there was more voicing in /s/ in Spanish overall; in addition, the difference between the mean voicing of /s/ word-internally vs. word-finally was greater in the Spanish words again. Supplementary Table 8 exhibits a summary of the two models.

4.3 Perception experiment

We begin with the English results. The left part of Figure 9 displays the ratings of the participants of the two texts. "Tom" was the native-like recording, while "TomVoi" was the one where Hungarian-like RVA was applied (1 = not at all native-like, 5 = completely native-like). As we can see, the ratings were lower for the non-native-like recording; for example, no participant ranked it with the highest score 5.

The best-fitting cumulative link mixed effects model was the one which included by-subject varying intercepts (but not slopes). According to this model, the non-native text significantly decreased the ratings, i.e., lower ratings were more likely ($b = -1.2$, SE = 0.27, $z = -4.46$, $p < 0.001$; R2m = 0.54, R2c = 0.17). These results indicate that the participants reliably differentiated between the two recordings, and rated the non-native one (with RVA applied) much lower.

The ratings of the Spanish recordings can be seen on the right of Figure 9. "Leyenda" was the native-like recording while "LeyendaSOVO" was the one in which no sonorant voicing was applied in the relevant phonological environments, hence this was the non-native-like recording.

The best-fitting model was the one which included by-subject varying intercepts (but not slopes). According to this model, while the non-native recording increased ratings, this increase was not statistically significant ($b = 0.46$, SE = 0.24, $z = 1.9$, $p = 0.056$; R2m = 0.41, R2c = 0.03). This result indicates that listeners did not reliably differentiate between the native recording (with PSV) and the non-native recording (without PSV).

5 Discussion

This study examined to what extent sequential multilingual speakers produce and perceive postlexical laryngeal processes in their L2/L3. The study also explored if cognates enhance CLI and if so, what properties determine whether it is the L1 or the L2 that has a more significant impact. The implications of the results presented in the previous section are described as follows.

5.1 Cognate status effect and voicing assimilation

The results of the production task revealed a somewhat complex picture of cognate status effect in relation to presonorant

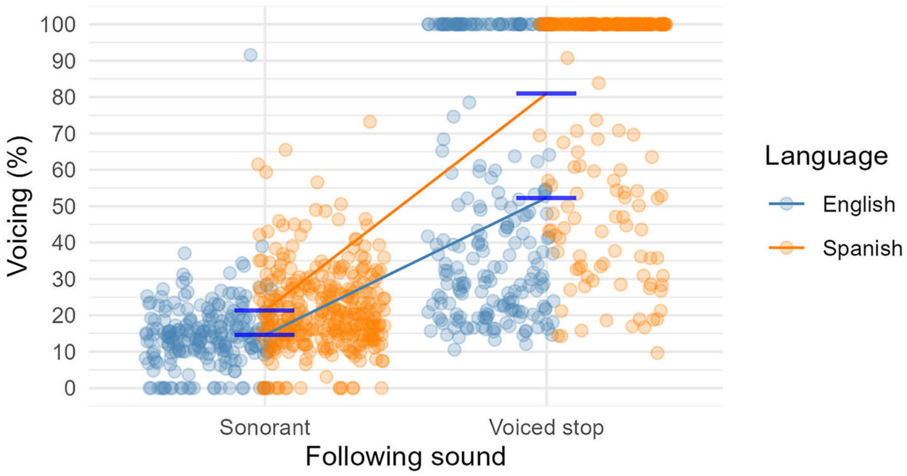


FIGURE 7
Voicing of word-final /s/. Lines represent the means.

TABLE 5 Descriptive statistics for the voicing of word-final /s/ before stops and voiced stops in English and Spanish words.

Trigger	Language	*N*	Mean	SD	Median	Min	Max
Sonorant	English	220	14.58	9.41	14.17	0.00	91.57
Sonorant	Spanish	336	21.31	11.31	19.49	0.00	73.22
Voiced stop	English	219	52.32	33.41	39.62	10.55	100.00
Voiced stop	Spanish	220	80.99	28.97	100.00	9.65	100.00

voicing and regressive voicing assimilation. In the presonorant voicing context, English /s/—in a non-target-like manner—displays increased voicing in cognates where L1 is inhibitory (Hungarian has a /z/ before the sonorant), thus supporting Hypothesis 1, although statistical analysis did not yield a significant difference here between the different lexical groups. This result calls for further caution as the increased voicing might be due to methodological reasons. The word that was responsible for it was *Yasmin*, which in some varieties of English is pronounced with a voiced fricative. The other inhibitory cognate, *Iceland* was probably not perceived as similar enough to produce an impact since vowel quality was too different (a diphthong in English while /i/ in both Hungarian and Spanish). On the other hand, in the realization of presonorant English /z/, where presonorant voicing applies vacuously, a cognate status effect was observed. Cognates with an inhibitory L1 influence were realized with significantly less voicing than the other lexical groups. Thus, English production data seem to support Hypothesis 1. It also means that the facilitative effect of L2 Spanish in these triple cognates could not counterbalance the inhibitory effect of L1 Hungarian, thus supporting Hypothesis 3. The reverse scenario (L1 facilitation and L2 inhibition) could not be tested in the latter phonological context since Spanish is always facilitative here because of PSV.

On the contrary, the Spanish production data do not show any cognate effects. There is a steady absence of pre-sonorant voicing, thus (partly) refuting Hypothesis 1. The results also reveal that facilitative cognates do not differ from non-cognate realizations, thus Hypothesis 2 (i.e., that facilitative cognates are acoustically more target-like than non-cognates) for presonorant voicing must be rejected. The reason for this could be that non-cognates already show a target-like realization which reached a ceiling with no possibility for further improvement. This, however, is not borne out for PSV in Spanish. Spanish presonorant /s/ in non-cognates is as voiceless (i.e., non-target-like) as English presonorant /s/ in non-cognates (which is target-like).

The question arises why our participants behaved differently in this respect in their two non-native languages. We hypothesize that the answer lies in phonemic encoding during the acquisition of these lexical items. While we think that both English and Spanish voiced realizations are acoustically more similar and should be identified with or mapped to Hungarian [z], and English and Spanish voiceless realizations are more similar to and should be mapped to Hungarian [s], the acquisition of a phoneme inventory is closely linked to the acquisition of a lexicon that includes minimal pairs (Darcy et al., 2017). As our participants are proficient speakers, they are likely to have acquired a stable phoneme inventory for both L2/L3 and have formed only one alveolar fricative category for Spanish, which is voiceless because Spanish does not display a /s/–/z/ contrast. In their Spanish speech, they implement only this voiceless segment across the board.

Turning to regressive voicing assimilation between adjacent obstruents, in neither of the two languages did participants treat cognates significantly differently from non-cognates, although some tendencies could be observed. In English, /s/ was on average

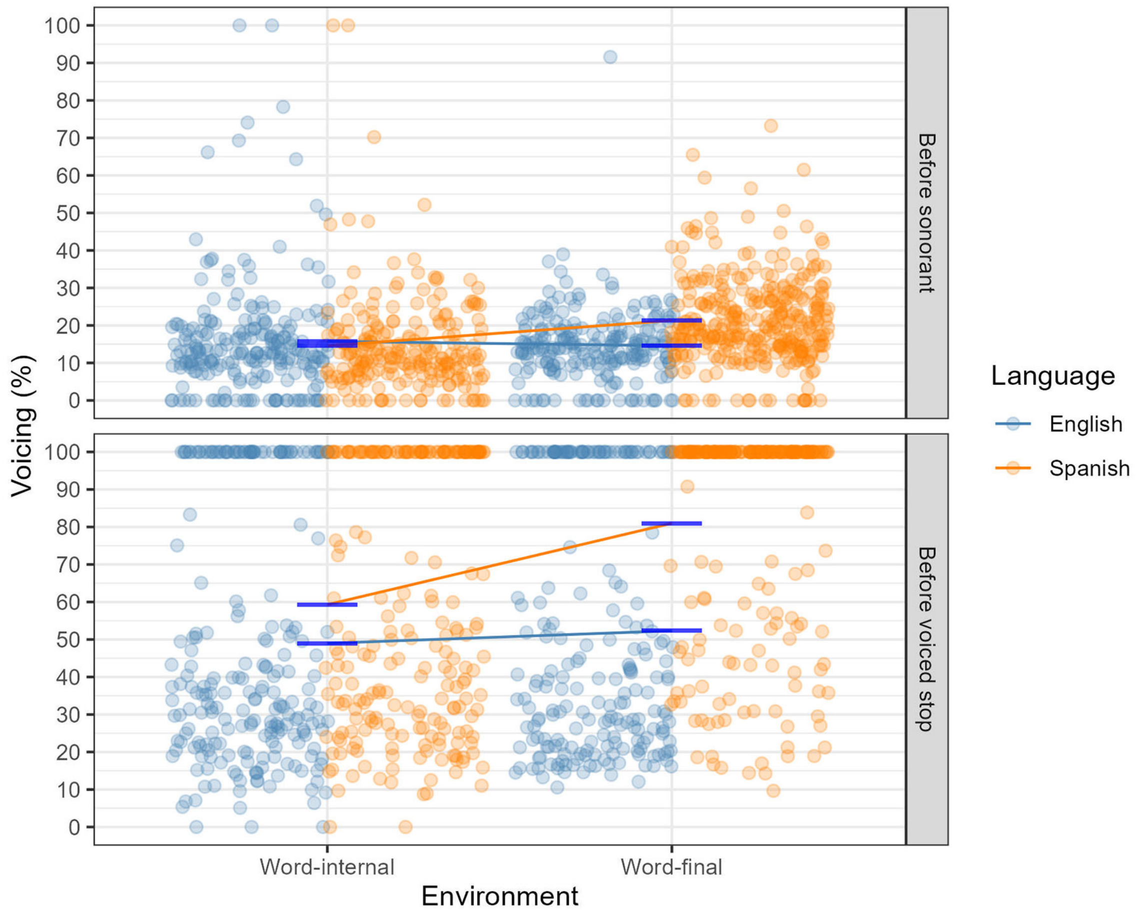


FIGURE 8
Voicing of word-internal vs. word-final /s/. Word-internal group only contains non-cognate words; lines represent the means.

TABLE 6 Descriptive statistics for the voicing of word-internal vs. word-final /s/ before stops and voiced stops in English and Spanish words.

Trigger	Environment	Language	N	Mean	SD	Median	Min	Max
Sonorant	Word-internal	English	224	15.75	15.12	12.83	0.00	100.00
Sonorant	Word-internal	Spanish	224	14.57	13.15	11.81	0.00	100.00
Sonorant	Word-final	English	220	14.58	9.41	14.17	0.00	91.57
Sonorant	Word-final	Spanish	336	21.31	11.31	19.49	0.00	73.22
Voiced stop	Word-internal	English	224	48.96	33.67	35.77	0.00	100.00
Voiced stop	Word-internal	Spanish	224	59.30	34.06	50.12	0.00	100.00
Voiced stop	Word-final	English	219	52.32	33.41	39.62	10.55	100.00
Voiced stop	Word-final	Spanish	220	80.99	28.97	100.00	9.65	100.00

11% more voiced before voiced stops in triple cognates than in non-cognates or English–Spanish cognates, again pointing in the direction of L1 having a larger impact on the phonetic realization of cognates than L2 (thus supporting Hypothesis 3) but only in the case of inhibitory cognates, thus, supporting Hypothesis 1; Hypothesis 2 must be rejected for RVA, too. In the Spanish data dispersion was slightly greater (15%), and no clear trend could be observed. Note that L1 is always facilitative in this context, just like in English words with /z/+voiced obstruent sequences. It is important to bear in mind that /s/ was produced with a fair amount of voicing in all these contexts which is non-target-like for English /s/ and target-like for English /z/ and Spanish. Thus, our results indicate that any potential lexical effects tend to be overridden by RVA. The voicing proportion measured in the present data is in line with research on RVA in Hungarian. In Bárkányi and G. Kiss (2019), the proportion of voicing in /s/ before voiced stops was on average 65.4% and before voiceless stops it was 15.1%. It has also been demonstrated that around 30% of voicing during the fricative is enough to induce voiced categorization, that is, an alveolar fricative with 30% of voicing proportion is more likely to

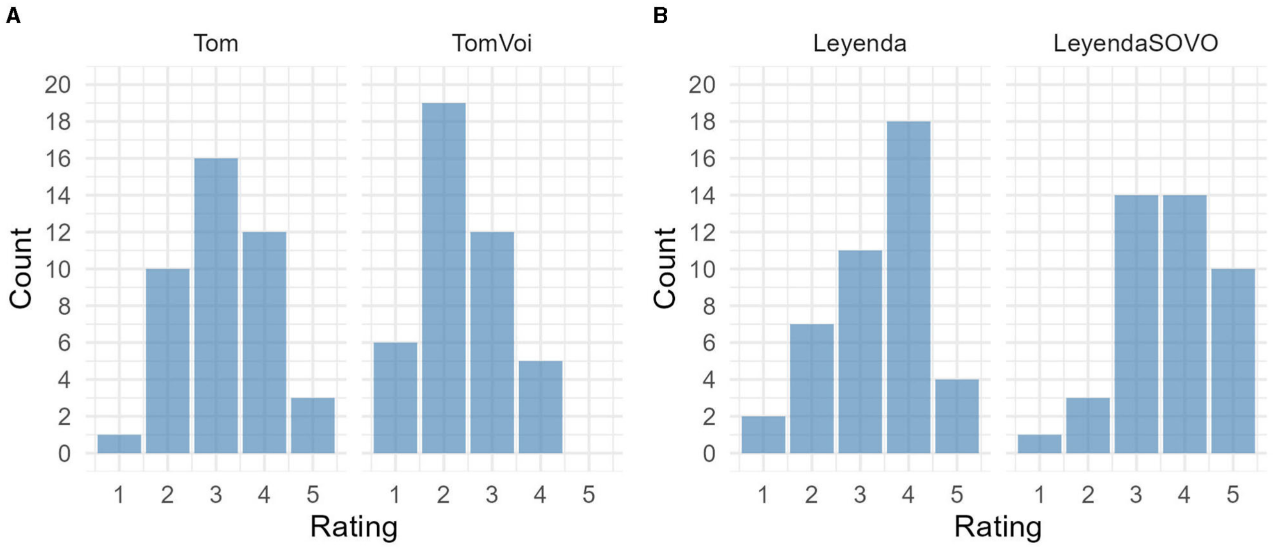


FIGURE 9
(A) Rating of the English texts. **(B)** Rating of the Spanish texts.

be categorized as /z/ by speakers of Hungarian (Bárkányi and G. Kiss, 2023).

5.2 Voicing assimilation across the board

The experiments were specifically designed to explore whether voicing assimilation as a dynamic process has been learned and unlearned. In order to test this, it is crucial to examine RVA and PSV across a word-boundary with the target segment being at the end of one word and the trigger in the next word. The patterns we observed are similar to those within the word. PSV does not seem to be applied in either of the two non-native languages. /s/ contains little coarticulatory voicing, which is expected and target-like in English, but had PSV been acquired, more voicing would be expected in Spanish. This experimental data supports Hypothesis 4 (i.e., sonorants do not trigger voicing assimilation in sandhi). It is interesting to note that although Spanish word-final /s/ was fairly voiceless (only 21.3% on average), it was voiced significantly more than word-internal /s/ or English /s/ in the same sandhi position. The reason for this might be that despite the fact that participants store the Spanish lexical items with a voiceless fricative and generally produce it as such at word level, two learners seem to have acquired PSV to some extent. They consistently produced tokens with 40–60% voicing in sandhi contexts. Overall, however, we can claim that there is little evidence for PSV of either within the word or across a word-boundary in the Spanish interlanguage of these multilingual advanced leaners. There are several factors that could have contributed to this. An appealing explanation is the considerable similarity of the Hungarian and Spanish laryngeal systems, namely, that both are true voice languages, both display RVA between adjacent obstruents, thus they are treated by our participants as having identical laryngeal systems. In addition, the fact that sonorant voicing in Spanish is variable might not serve as sufficient and salient input for learners to be "discovered."

Unlike PSV, the non-target-like application of RVA in English is perceived by the participants of this study; however, their productions do not mirror it: /s/ before a voiced stop was predominantly voiced (around 50%). This means that participants failed to block RVA in their English interlanguage, which supports Hypothesis 5. It is interesting to note that Spanish word-final /s/ resulted in significantly more voiced realizations than English /s/ in the same position (the word-internal fricative was also slightly more voiced in Spanish than in English). This might suggest that participants do aim to block RVA in English, but they are not very successful. This point is in need of further research as we cannot explain why the sandhi context in Spanish triggered so much more voicing than the same context within the word.

5.3 The link between perception and production

The results of the perception experiments also support the hypothesis that PSV in the Spanish interlanguage of the participants remained unnoticed (Figure 9). This does not necessarily mean that learners cannot hear voicing itself, but even if they do, they perceive it as random noise rather than a language specific phonological process. These data do not provide evidence in favor of any of the scenarios described in the Introduction (Section 1), we can only state that overall, PSV was not acquired in any domain, which is somewhat surprising since the participants of the present study are advanced learners. We consider it as an indication that the acquisition of dynamic phonological processes is different from the acquisition of contrastive segments, static inventories. We found intra- and interspeaker variation in the data (as mentioned in 5.2), we leave the exploration of individual learning patterns for future research.

The presence of RVA in the English interlanguage of the participants shows that learners do hear the non-target-like application of the process, but fail to block its implementation in their productions. This could be compatible with the predictions made by SLM—assuming this model can be extended to phonological processes, too—namely, that only accurate perception can be transferred to production. Our participants could be at the stage of accurate perception which has not yet been transferred to production. As learners in this case did not have to acquire any non-native segments, the possibility of a motor-articulatory difficulty blocking correct production can be discarded, as would be the case, for instance, when acquiring a trilled rhotic by speakers whose L1 does not have one. However, as our participants are highly proficient speakers of English, it is quite unlikely that they will supress RVA at a later stage. Thus, these results are also compatible with the claim that there is no direct correlation between perception and production, as reported by Liu and Lin (2021). Wrembel et al. (2020) found that Polish speakers did not show significant development in the acquisition of voicing and devoicing in English and German in the perception domain. This again, indicates that the acquisition of dynamic processes might have a different pattern and a ceiling effect might be reached earlier than in the case of learning contrastive segments.

5.4 Theoretical implications and future research

The findings of the present study can probably be best accounted for by Slabakova's (2017) Scalpel Model (SM). This model has been proposed for L3 morphosyntactic acquisition, just like the Linguistic Proximity Model, and also sees linguistic proximity as a decisive factor in determining transfer, but explicitly claims that additional (cognitive and experiential) factors can have a significant impact on CLI. Such factors include structural linguistic complexity, construction frequency, misleading input, negative evidence and prevalent language activation. We can discard negative evidence for now as learners are rarely corrected for the "erroneous" application of postlexical processes. As far as language activation is concerned, L1 is the language predominantly used by these learners, but there is no clear usage difference between their L2/L3. When extending SM for L3 phonology, we can add articulatory and perceptual complexity next to structural linguistic complexity. As mentioned in 5.3, articulatory complexity is not likely to play a role here, while perceptual salience (or the lack of it) might be important. Its impact on production needs further research: whether allophonic alternations that create salient novel segments that are easy to articulate are more likely to be implemented in production. The question of insufficient or "misleading" input due to the variable nature of PSV in Spanish has been dealt with in 5.2, and is a plausible reason for the lack of acquisition, which might be supported by sequence frequency, although frequency alone is not likely to have an impact on the acquisition of PSV and RVA. Based on CORPES (XXI), the average of the normalized frequency (per million words) of /s/+sonorant (/m, n, l/) and /s/+voiced stops (/b, d, g/) is as follows. While the sequence of /s/+sonorant is slightly more frequent than that of /s/+voiced stop within the word (1,346.5 as opposed to 1,041.4), across the word boundary, there are twice as many /s/+voiced stop occurrences than /s/+sonorant ones (9,906.7 vs. 4,897.8). The role of occurrence frequency in postlexical processes awaits further research.

Participants failed to learn PSV, a typologically uncommon process. However, they also failed to unlearn RVA, a typologically common process, a default process for true voice languages, but absent from aspirating languages. This might indicate that they simply transferred their L1 laryngeal system into their subsequent languages. Already Eckman (1977) pointed out that German learners could not suppress laryngeal neutralization in English (but English learners had less difficulty in learning the laryngeal properties of German), which the author explains with typological markedness. It would be worth testing whether learners whose native language displays both PSV and RVA are able to block the former more readily than the latter. The primacy of L1 was corroborated by the inhibitory transfer in triple cognates too. Therefore, one important finding of this study is that in postlexical phonological processes, L1 plays the primary role, and L2 does not appear to exert a strong enough influence. It is to be clarified by further research in what ways the acquisition of phonemic contrasts and static phonological features differ from the acquisition of dynamic phonological processes, and whether there is a parallelism between these and the acquisition of lexical items and grammatical knowledge.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the studies involving humans because, it is not mandatory at Eötvös Loránd University. All participants received an information sheet about the purpose, data collection and anonyimization in relation to the present experiments. They all gave their written informed consent. The studies were conducted in accordance with the local legislation and institutional requirements.

Author contributions

ZsB: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Supervision, Writing—original draft, Writing—review & editing. ZGK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The present research has been supported by the National Research, Development and Innovation Grant: K142498.

Acknowledgments

We would like to thank Firas Shbeeb for his help with the experiments and acoustic analysis, Anna Hamp for her help in the preparation of materials, Péter Siptár for his comments on earlier versions of this paper, and the editor and the two reviewers for their very valuable comments that made this paper much better.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2023.1304666/full#supplementary-material

References

Amengual, M. (2012). Interlingual influence in bilingual speech: cognate status effect in a continuum of bilingualism. *Bilingualism* 15, 517–530. doi: 10.1017/S1366728911000460

Amengual, M. (2016). Cross-linguistic influence in the bilingual mental lexicon: evidence of cognate effects in the phonetic production and processing of a vowel contrast. *Front. Psychol.* 7, 617. doi: 10.3389/fpsyg.2016.00617

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *J. Acoust. Soc. Am.* 149, 1271–1284. doi: 10.1121/10.0003559

Archibald, J. (2022). Phonological parsing via an integrated I-language: the emergence of property-by-property transfer effects in L3 phonology. *Linguist. Approach. Bilingual.* 13, 743–766. doi: 10.1075/lab.21017.arc

Baker, W., and Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: the role of individual differences. *Int. Rev. Appl. Linguist.* 44, 231–250. doi: 10.1515/IRAL.2006.010

Bardel, C., and Sánchez, L. (2017). "The L2 status factor hypothesis revisited: the role of metalinguistic knowledge, working memory, attention and noticing in third language learning," in *L3 Syntactic Transfer: Models, New Developments and Implications*, eds T. Angelovska and A. Hahn (Amsterdam: John Benjamins Publishing Company), 85–102.

Bárkányi, Zs. (2014). "Reflexiones sobre la asimilaci6n de sonoridad dela lsl l.Introducci6n," in *GPS 60°: Köszöntő kötet Giampaolo Salvi 60. születésnapjára. Studi di linguistica neolatina per i 60 anni di Giampaolo SalviGPS 60*, ed Z. Fábián (Budapest: ELTE), 25–36.

Bárkányi, Zs., and G. Kiss, Z. (2015). "Why do sonorants not voice in Hungarian? And why do they voice in Slovak?," in *Approaches to Hungarian: Volume 14: Papers from the 2013 Piliscsaba Conference*, eds K. É. Kiss, B. Surányi, and É. Dékány (Amsterdam: John Benjamins Publishing Company), 65–94.

Bárkányi, Zs., and G. Kiss, Z. (2019). A fonetikai korrelátumok szerepe a zöngekontraszt fenntartásában. Beszédprodukciós és észleléses eredmények [The role of phonetic correlates in voicing contrast. Results from speech production and perception]. *Általános Nyelvészeti Tanulmányok* 31, 57–102.

Bárkányi, Zs., and G. Kiss, Z. (2023). Production and perception of voicing contrast in assimilation contexts in Hungarian. *J. Uralic Linguist.* 2, 5–49. doi: 10.1075/jul.00013.bar

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beckman, J., Jessen, M., and Ringen, C. (2013). Empirical evidence for laryngeal features: aspirating vs. true voice languages. *J. Linguist.* 49, 259–284. doi: 10.1017/S0022226712000424

Benrabah, M. (1991). "Learning English segments with two languages," in *Actes du XIIème Congrès International des Sciences Phonétiques* (Aix-en-Provence: Université de Provence, Service des Publications), 334–337.

Berkes, É., and Flynn, S. (2012). "Multilingualism: new perspectives on syntactic development," in *The Handbook of Bilingualism and Multilingualism* (Hoboken, NJ: John Wiley & Sons, Ltd), 137–167.

Best, C. T. (1994). "The emergence of native-language phonological influences in infants: a perceptual assimilation model," in *The Development of Speech Perception: Theoretical and Methodological Issues in Cross-Language Speech Research*, ed H. C. Nusbaum (Timonium, MD: York Press), 171–204.

Best, C. T., and Tyler, M. D. (2007). "Commonalities and complementarities: nonnative and second-language speech perception," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds O. S. Bohn and M.J. Munro (Amsterdam: John Benjamins Publishing Company), 13–34.

Boersma, and Weenink, D. (2022). *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.2.23.* Retrieved from https://www.praat.org (May 2, 2023).

Bolker, B., and Robinson, D. (2022). *Broom.mixed: Tidying Methods for Mixed Models.* Retrieved from https://CRAN.R-project.org/package=broom.mixed (May 2, 2023).

Bybee, J. (2001) *Phonology and Language Use*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics).

Cabrelli, J., and Pichan, C. (2021). Initial phonological transfer in L3 Brazilian Portuguese and Italian. *Linguist. Approach. Bilingual.* 11, 131–167. doi: 10.1075/lab.18048.cab

Cal, Z., and Sypiańska, J. (2020). The interaction of L2 and L3 levels of proficiency in third language acquisition. *Poznan Stud. Contemp. Linguist.* 56, 577–603. doi: 10.1515/psicl-2020-0019

Campos-Astorkiza, R. (2015). "Segmental and prosodic conditionings on gradient voicing assimilation in Spanish," in *Hispanic Linguistics at the Crossroads: Theoretical Linguistics, Language Acquisition and Language Contact. Proceedings of the Hispanic Linguistics Symposium 2013*, eds R. Klassen, J. M. Liceras, and E. Valenzuela (Amsterdam: John Benjamins Publishing Company), 127–144.

Christensen, R. H. B. (2022). *Ordinal: Regression Models for Ordinal Data. R Package Version 2022.11-16.* Retrieved from https://CRAN.R-project.org/package=ordinal (May 2, 2023).

CORPES XXI. *Real Academia Española: Corpus del Español del Siglo XXI*. Available online at: https://www.rae.es/banco-de-datos/corpes-xxi (accessed November 2, 2023).

Darcy, I., Peperkamp, S., and Dupoux, E. (2007). Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners. *Pap. Lab. Phonol.* 9, 9.

Darcy, I., Tremblay, A., and Simonet, M. (2017). Editorial: phonology in the bilingual and bidialectal lexicon. *Front. Psychol.* 8, 2017. doi: 10.3389/fpsyg.2017.00507

Derwing, T., and Munro, M. J. (2015) *Pronunciation Fundamentals*. Amsterdam: John Benjamins Publishing Company.

Draxler, C., and Jänsch, K. (2004). *SpeechRecorder: a Universal Platform Independent Multi-Channel Audio Recording Software.* Retrieved from https://www.bas.uni-muenchen.de/Bas/software/speechrecorder/ (May 2, 2023).

Ecke, P., and Hall, C. J. (2021). The Parasitic Model: lexical acquisition and its impact on morphosyntactic transfer. *Linguist. Approach. Bilingual.* 11, 45–49. doi: 10.1075/lab.20088.eck

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Lang. Learn.* 27, 315–330. doi: 10.1111/j.1467-1770.1977.tb00124.x

Escudero (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*.

Flege, J., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)", in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83. doi: 10.1017/9781108886901.002

Flege, J. E. (1995). "Second-language speech learning: theory, findings, and problems," in *Speech Perception and Linguistic Experience: Issue in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 157–232.

Flege, J. E., Frida, E. M., Walley, A. C., and Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Stud. Sec. Lang. Acquisit.* 20, 155–187. doi: 10.1017/S0272263198002034

Flege, J. E., and Munro, M. J. (1994). The word unit in second language speech production and perception. *Stud. Sec. Lang. Acquisit.* 16, 381–411. doi: 10.1017/S0272263100013437

Hahn, E. D., and Soyer, R. (2005). Probit and logit models: differences in the multivariate realm. *J. Royal Stat. Soc.* 1–12.

Hermas, A. (2015). The categorization of the relative complementizer phrase in third-language English: a feature re-assembly account. *Int. J. Bilingual.* 19, 587–607. doi: 10.1177/1367006914527019

Hualde, J. I. (2005). *The Sounds of Spanish*. New York, NY: Cambridge University Press.

Jansen, W. (2004) *Laryngeal Contrast and Phonetic Voicing: A Laboratory Phonology Approach to English, Hungarian, and Dutch*. (Doctoral dissertation), University of Groningen, Groningen, Netherlands.

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286–319.

Kopečková, R., Gut, U., Wrembel, M., and Blas, A. (2022). Phonological cross-linguistic influence at the initial stages of L3 acquisition. *Sec. Lang. Res.* 2022, 1–25. doi: 10.1177/02676583221123994

Kuzla, C., Ernestus, M., and Mitterer, H. (2010). "Compensation for assimilatory devoicing and prosodic structure in German fricative perception," in *Laboratory Phonology 10*, eds C. Fougeron, B. Kühnert, M. D'Imperio, and N. Vallée, (Berlin, New York: De Gruyter Mouton), 731–758. doi: 10.1515/9783110224917.5.731

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). lmerTest Package: tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lenth, R. V. (2023). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. Retrieved from https://CRAN.R-project.org/package=emmeans (May 2, 2023).

Lisker, L., and Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384–422. doi: 10.1080/00437956.1964.11659830

Liu, J., and Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *SAGE Open* 11, 10. doi: 10.1177/2158244020985510

Llama, R., and Cardoso, W. (2018). Revisiting (non-)native influence in VOT production: insights from advanced L3 Spanish. *Languages* 3, 30. doi: 10.3390/languages3030030

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

Llama, R., and López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *Int. J. Multilingual.* 13, 444–58. doi: 10.1080/14790718.2016.1217602

Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., and Makowski, D. (2021). An R package for assessment, comparison and testing of statistical models. *J. Open Sour. Softw.* 6, 3139. doi: 10.21105/joss.03139

Mora, J. C., and Nadeu, M. (2009). Experience effects on the categorization of a native vowel contrast in highly proficient Catalan-Spanish bilinguals. *J. Acoust. Soc. Am.* 125, 2775. doi: 10.1121/1.4784755

Pierrehumbert, J. B. (2001). "Exemplar dynamics: word frequency, lenition and contrast," in *Frequency and the Emergence of Linguistic Structure*, eds J. Bybee and P. Hopper (Amsterdam: John Benjamins Publishing Company), 45, 137–157.

Quilis, A. (1993) *Tratado de fonología y fonética españolas*. Madrid: Gredos.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. Vienna: Foundation for Statistical Computing. Retrieved from: https://www.R-project.org/ (May 2, 2023).

Rothman, J. (2015). Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: timing of acquisition and proficiency considered. *Bilingualism* 18, 179–190. doi: 10.1017/S136672891300059X

Schmidt, L., and Willis, E. (2011). "Systematic investigation of voicing assimilation of Spanish /s/ in Mexico City," in *Selected Proceedings of the 5th Conference on Laboratory Approaches to Romance Phonology*, ed S. M. Alvord (Somerville, MA: Cascadilla Proceedings Project), 1–20.

Siptár, P., and Törkenczy, M. (2000). *The Phonology of Hungarian*. Oxford: Oxford University Press.

Slabakova, R. (2017). The scalpel model of third language acquisition. *Int. J. Bilingual.* 21, 651–665. doi: 10.1177/1367006916655413

Strycharczuk, P. (2012). *Phonetics–Phonology Interactions in Pre-sonorant Voicing*. (Doctoral dissertation), University of Manchester, Manchester, United Kingdom.

Szigetvári, P. (2020). Emancipating lenes: a reanalysis of English obstruent clusters. *Acta Linguist. Acad.* 67, 39–53. doi: 10.1556/2062.2020.00004

Vago, R. M. (1980) *The Sound Pattern of Hungarian*. Washington, DC: Georgetown University Press.

van Leussen, J. W., and Escudero (2015). Learning to perceive and recognize a second language: the L2LP model revised. *Front. Psychol.* 6, 1000. doi: 10.3389/fpsyg.2015.01000

Wang, D., and Nance, C. (2023). Third language phonological acquisition: understanding sound structure in a multilingual world. *Lang. Linguist. Compass* 17, e12497. doi: 10.1111/lnc3.12497

Wells, J. C. (1982) *Accents of English*. Cambridge: Cambridge University Press.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: the Linguistic Proximity Model. *Int. J. Bilingual.* 21, 666–682. doi: 10.1177/1367006916648859

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., et al. (2019). Welcome to the tidyverse. *J. Open Sour. Softw.* 4, 1686. doi: 10.21105/joss.01686

Wrembel, M. (2021). Transfer vs. dynamic cross-linguistic interactions: exploring alternative avenues in L3 research. *Linguist. Approach. Bilingual.* 11, 109–115. doi: 10.1075/lab.20092.wre

Wrembel, M., Gut, U., Kopečková, R., and Balas, A. (2020). Cross-linguistic interactions in third language acquisition: evidence from multi-feature analysis of speech perception. *Languages* 5, 52. doi: 10.3390/languages5040052

Wrembel, M., Gut, U., Kopečková, R., and Balas, A. (2022). The relationship between the perception and production of L2 and L3 rhotics in young multilinguals; an exploratory cross-linguistic study. *Int. J. Multilingual.* 2022, 1–20. doi: 10.1080/14790718.2022.2036158

Wrembel, M., Marecka, M., and Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *Int. J. Multilingual.* 16, 513–533. doi: 10.1080/14790718.2019.1583233

Wunder, E. M. (2011). "Crosslinguistic influence in multilingual language acquisition: phonology in third or additional language acquisition," in *New Trends in Crosslinguistic Influence and Multilingualism Research*, eds G. D. Angelis and J. M. Dewaele (Bristol: Multilingual Matters), 105–128. doi: 10.21832/9781847694430

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 09 February 2024
DOI 10.3389/flang.2024.1295265

Check for updates

OPEN ACCESS

EDITED BY
Jennifer Cabrelli,
University of Illinois Chicago, United States

REVIEWED BY
Shannon Barrios,
The University of Utah, United States
John Matthews,
Chuo University, Japan

*CORRESPONDENCE
Fernanda Barrientos
fernanda.barrientos-contreras@uni-konstanz.de

RECEIVED 15 September 2023
ACCEPTED 17 January 2024
PUBLISHED 09 February 2024

CITATION
Barrientos F (2024) Out with the old, in with the new: contrasts involving new features with acoustically salient cues are more likely to be acquired than those that redeploy L1 features. *Front. Lang. Sci.* 3:1295265.
doi: 10.3389/flang.2024.1295265

Out with the old, in with the new: contrasts involving new features with acoustically salient cues are more likely to be acquired than those that redeploy L1 features

Fernanda Barrientos*

Department of Linguistics, University of Konstanz, Konstanz, Germany

Feature-based approaches to second language (L2) phonology conceptualize the acquisition of new segments as operations that involve either the addition of new phonological features, or the rebundling of existent ones. While the deficit hypothesis assumes that only features that are fully specified in the L1 can be redeployed to the L2 in order to create new segments, it has been shown that features which are completely absent in the L1 can also be learned. This article investigates whether a learning scenario in which features are only partially available (that is, they are present in the L1, but are redundant with other features) is less challenging than learning an entirely new feature, even when the new feature has acoustically salient cues. Since Spanish has a much smaller vowel system /i e a o u/, L2 learners of German with Spanish as L1 need to learn a system with front rounded vowels as well as tense/lax contrasts. We tested L1 Spanish speakers' perception of the German contrasts /i/ ~ /ɪ/ (e.g., *Miete/mitte*, where [+/− tense] is acquired) and /u/ ~ /y/ (e.g., *Spulen/spülen*, where L1 feature [+/−round] redeploys to a front vowel). The results showed that experienced L2 learners are more successful when discriminating between sounds in a feature acquisition scenario than in redeployment; however, neither of the non-native contrasts was easier to perceive than the other in the identification task. The differences in performance between tasks and in the acoustic saliency of the cues by contrast (F2 vs. duration and F1) suggests that L2 phonological acquisition is likely to take place at a surface level and favors learning through attunement to auditorily salient acoustic cues over internal rearrangement of abstract features, regardless of their presence in the L1.

KEYWORDS

speech perception, feature-based models, L2 phonology, underspecification, vowel contrast, perceptual cues, feature redeployment

1 Introduction

In L2 phonology, whether transfer and acquisition are surface-based or phonological phenomena (Major, 2008, p. 68) is an ongoing discussion; nevertheless, research following an acoustic-based approach vastly outnumbers feature-based accounts. The main claims of the Speech Learning Model (SLM) and its revised version (SLM-r) (Flege, 1995; Flege and Bohn, 2021), which propose an acoustic-based approach to the acquisition of L2 segments, or the Perceptual Assimilation Model (PAM/PAM-L2) (Best, 1995; Best and Tyler, 2007), with a direct-realist perspective, have been extensively tested and confirmed by a large number of empirical studies.

On the other hand, more abstract models have also been proposed, where topics such as the role of universals and feature prominence are discussed, e.g., the Ontogeny-Phylogeny Model (OPM) (Major, 2001) or the Feature Competition Model (FCM) (Hancin-Bhatt, 1994), respectively. Among such phonology-based accounts, of particular interest for this research is the Feature-Based Model (FBM) (Brown, 1998, 2000). The model builds upon FCM and is motivated by Feature Geometry (Clements, 1985; Clements and Hume, 1995), a phonological framework wherein segments are no longer defined as "bundles" of unorganized features (à la SPE) but rather as a hierarchy, with tiers to which features are associated. Further developments of Feature Geometry include Underspecification Theory (Archangeli, 1988; Avery and Rice, 1989), according to which abstract segmental representations only contain the features needed for contrast with regards to the rest of the segments in the inventory.

The key aspect of Brown's FBM is that the L1 constrains acquisition whenever the L2 phonemic inventory has features that the L1 does not have, or rather, when such features are underspecified in the learner's L1. In such cases, the L2 segment simply cannot be fully acquired due to the fact that the relevant features are underspecified in the L1 and thus are unavailable for transfer. Brown's interpretation of the inability to learn new features is that the L1 hinders the L2 learner's access to Universal Grammar. As a corollary to the above, cases in which the L1 has the same features as the L2 but in different natural classes (e.g., when length is available in the L1 only in vowels, but the L2 has it in consonants) or when the acoustic cue associated to a particular contrast is relevant in the L1 but to a lesser extent, should result in easier acquisition, a prediction that has been borne out, at least in the case of length contrasts (e.g., McAllister et al., 2002; Pajak and Levy, 2014).

However, other studies have suggested that L1 speakers/listeners of languages without a tense/lax contrast are indeed able to distinguish between tense and lax segments in an L2, e.g., speakers of L1 Catalan perceiving English /i/ ∼ /ɪ/ and /e/ ∼ /ɛ/ (also "long/short," since the contrast is usually realized via differences in duration) when performing certain perceptual tasks, at least to a certain extent, regardless of their amount of experience in the L2 (Cebrian, 2006); this ability to "turn off" the attunement to the L1 acoustic cues in favor of new ones was also observed by Bohn (1995) in L1 speakers of Spanish learning English /i/ ∼ /ɪ/, in a phenomenon known as the "Desensitization Hypothesis."[1] Furthermore, it has also been found that the presence or absence of the relevant phonological features in the L1 does not seem to play a role when explaining difficulty in the acquisition of L2 contrasts (Barrios et al., 2016). In sum, the evidence regarding the acquisition of tense/lax contrasts disproves a strong interpretation of Brown's theory, although the feature's status in the grammar may differ across experiments: while Brown's experiment looks into the effect of underspecified features below the coronal node that may nevertheless allow for allophonic forms that match contrasting L2 sounds, the [tense] feature is completely absent in both Spanish and Catalan.

This research examines the effect of underspecification in the acquisition of new L2 sounds; that is, whether there is a difference in the perception of nonnative sounds when the feature responsible for a given contrast in the L2 is completely absent in the L1's grammar, vs. when the feature in question is present but also redundant with other features (i.e. feature +A occurs always in segments with feature +B) in the L1. In a null hypothesis, both underspecification and absence of a feature in the grammar would have the same effect on L2 learning: that is, that new L2 contrasts involving either absent or redundant features should pose the same degree of difficulty in acquisition. Conversely, the alternative hypothesis would state that either type of feature (underspecified or absent) would be easier to acquire than the other. On the one hand, it could be hypothesized that a present feature is better than nothing at all; on the other hand, if the absent feature has an acoustically salient cue it may be easier for L2 learners to perceive and produce. Furthermore, this study also analyzes the role of the acoustic cues involved in contrasts, that is, whether a new feature with acoustically salient cues may have an advantage over an existing one with cues that are perceptually less salient, due to either intrinsic auditory non-saliency or because the cue is already in use for the perception of another phonological contrast. By examining the perception of both types of L2 contrasts through perception tasks that tap different levels of representation, we aim to find potential differences which could in turn shed light on the phonological status of L2 representations.

2 Theoretical background: features in L2 phonology

2.1 Feature-based experiments

A considerable part of what we know about L2 phonological acquisition is largely based upon empirical studies that investigate the acquisition of new segments in the L2 from an acoustic-based approach; in this regard, the SLM(r) has been extensively cited as a model for acquisition that focuses on acoustic similarity to explain difficulties in L2 acquisition. Thus, SLM (and also PAM-L2) predicts that difficulties in L2 acquisition arise whenever the input is perceived by the L2 learner as similar enough to an L1 category. On the other hand, Brown (1998) points out that these models do not attempt to explain the impact of the native grammar on the perception of L2 sounds (139–140). More importantly, Brown looks into a more abstract level of representation, whereby the crucial factor for predicting difficulty is the internal configuration of the segment (that is, the segment's abstract phonological features as well as their dependencies within the segment's geometry).

The evidence offered by Brown (1998) consisted of two different perceptual tasks (AX discrimination and picture identification) involving the /r/ ∼ /l/ contrast, carried out by experienced L2 listeners whose L1 was either Mandarin or Japanese (neither of these have these sounds in contrastive distribution).

1 The results of this study were statistically reanalyzed in Flege et al. (1997) where it was concluded that the differences in cue weighting between native and L1 Spanish nonnative speakers of English were not significant. However, and as discussed in Escudero and Boersma (2004), the statistical power of the test performed cannot be high since relatively high differences yielded non-significant results.

The results showed that Mandarin speakers performed better than the Japanese participants in both tasks. Brown suggests that this advantage of Mandarin listeners over Japanese listeners is given by the transfer of the feature [COR] and terminal node [anterior], present in Mandarin's sibilants, to English liquids /r/ ∼ /l/. On the other hand, Japanese does not contrast any of their coronal phonemes, since the [COR] feature is not further specified.[2] However, in order for this account to work, Brown suggests that English /r/ is a phoneme whose place node is specified as [COR], whereas English /l/ has an empty place node. This is an alternative to the usual assumption that the relevant feature in English for the /r/ ∼ /l/ distinction is [lateral], along the lines of previous research (e.g., Spencer, 1984). This led to Brown's proposal that fully specified L1 features not only can be transferred to the L2 in order to convey phonemic contrasts between sounds that are not present in the L1, but also that the perception of new L2 contrasts depends *exclusively* on this transfer operation; that is, if the features in question are not present, then acquisition is not possible. Surface phenomena such as the acoustic similarity of L2 sounds in relation to L1 sounds, or the type of acoustic cue involved in the perception of new contrasts, do not seem to play a role in the transfer of phonological features to the L2. Follow-up experiments (Brown, 2000) have shown again that features that are contrastive in the L1 can be redeployed to the L2, but not those that are redundant in the L1.

The hypothesis that it is the segment's feature geometry, and not the acoustic similarities of the input with regard to L1 sounds, that predicts different degrees of difficulty in L2 acquisition was also explored by LaCharité and Prévost (1999) who examined the perception of English /θ ∼ t/, /ŋ∼ n/ in coda position, and /h/ ∼ Ø by native speakers of French. The tasks were all discrimination (AX with minimal pairs, ISI = 500 ms, and ABX, same minimal pairs, ISI = 1000 ms). The results showed an effect of contrast only in the task with the shorter ISI; the ABX task did not yield any significant differences. The conclusions suggest that creating a new articulator node, e.g., Pharyngeal for /h/, may be more difficult than adding a terminal feature to an existing node (such as [distributed] for /θ/). However, it is worth noting that the tasks in the study point to different results because they tap different types of knowledge, with the AX task being more likely to tap phonetic/acoustic perception than the ABX one; likewise, discrimination tasks with longer ISI are more likely to tap phonological knowledge (Werker and Tees, 1984; Werker and Logan, 1985).

The perception of Polish post-alveolar sibilants /ɕ, ʑ/ by naïve Croatian listeners was tested by Ćavar and Hamann (2011). While these sounds do not have a phonemic status in Croatian, they do exist as allophones of different phonemes (of /ʂ/ and /ʐ/ respectively). The participants carried out an identification task whose stimuli consisted of the target consonant with a prefixed vowel. While the Polish group performed at 99.7% accuracy, the Croatian group performed at 96%. Further language groups (Slovenian and German) showed poorer performance. The results suggest that Croatian speakers were able to transfer [continuant, voice, back] from their L1, unlike German and Slovenian speakers, since they lack this exact configuration of features within one sound. However, it is still possible that the Croatian listeners were able to map these allophonic representations onto their own phonemic categories, which are contrastive via different features.

All in all, experiments taking a feature-based approach may be leading to different results not just because of differences in the methodology, but also because the role of the features in the L1 grammar differs across experiments. However, generalizations regarding the validity of feature-based approaches are difficult to establish given the lack of empirical research on the acquisition of abstract phonological features in L2.

2.2 Experiments involving duration

Current approaches to phonological length propose that phonological length is not a property of the segment itself and therefore not a feature such as [long] proposed in SPE. Instead, and at least in the case of vowel length, a long vowel can be conceptualized as the result of two higher-level, prosodic representations linked to one lower-level representation (segment) (Odden, 2011). Therefore, and assuming that English /i/ ∼ /ɪ/ is not a length-only contrast but a tense/lax one, the results observed in these experiments are not informative regarding feature-based approaches.

Experiments involving length-only contrasts have shown that length, when realized only via duration, is not difficult to perceive by nonnative speakers, regardless of their level of experience. The discrimination of consonant length by speakers of Korean, Cantonese, Vietnamese and Mandarin, tested on nonce words modeled after Polish phonology, yields higher d-prime scores for speakers of the first three languages, all of them with a *vowel* length contrast; Mandarin speakers, who do not have a length contrast at all in their L1, showed the lowest d-prime scores (Pajak and Levy, 2014). The perception of vowel length in experienced speakers of L2 Swedish with different native languages was tested by McAllister et al. (2002), where listeners had to listen to a real word and a nonce word, and determine which one was the real word based on vowel duration. The results showed that native speakers of L1 Estonian (with a three-way length distinction) performed best, followed by speakers of L1 English, and in last place, L1 Spanish. Length of residence (LOR) in Sweden did not play a role in the results of the L1 Spanish group, but there was a significant correlation between accuracy and LOR in the L1 English group; this suggests that speakers of languages that make use of a certain acoustic cue (if only partially) for establishing phonemic contrasts in the L1, may have an advantage over languages that do not use this cue for phonemic contrast.

2 Even though Brown states that "Japanese, on the other hand, does not contrast any phonemes within the coronal place of articulation" (p. 153) and presents the Japanese vowel inventory without postalveolar consonants, one reviewer pointed out that Japanese does have alveolar/postalveolar minimal pairs, such as [sakai] "border" vs. [ɕakai] "society." The best account for this apparent contradiction is that (ɕ) is a merge of two underlying segments: /s/ and a palatal /j/ (Labrune, 2012, p. 67).

2.3 Experiments involving the tense-lax contrast

The [tense] feature has been traditionally used for the characterization of the difference between tense/lax segments such as /i/ ∼ /ɪ/ in English (Giegerich, 1992) and German (Wiese, 2000). While the tense/lax contrast may be seen in phonological terms as simply another type of feature contrast, this is not necessarily the case. From a phonetic perspective, the contrast between tense and lax vowels has been shown to be based both on spectral and durational differences, whereby tense vowels are associated to longer duration as well as peripheral formant frequency values, in English (Hillenbrand et al., 1995; Leung et al., 2016) and German (Delattre and Hohenberg, 1968; Jessen et al., 1995). Since the inherent differences in duration in the tense/lax distinction are not encoded in the segmental representation as features, this leads to the question whether [tense] should instead be conceptualized as a suprasegmental only or replaced by the [ATR] feature, among other options (see Durand, 2005 for a detailed discussion).[3]

Experiments in the acquisition of vowel contrasts where the [tense] feature is involved show that L2 learners are able to attune their perception to the duration cue, even when this is not present in their L1. Bohn (1995) tested the perception of English vowels /i/ ∼ /ɪ/ in L1 speakers of Spanish and Mandarin, who made use of the duration cue (not used in Spanish or Mandarin for phonemic contrasts) more than the spectral differences. Bohn called the phenomenon of "shutting down" the perception of an L1 cue in favor of a completely new one "the desensitization hypothesis," which seems to take place in L2 listeners regardless of their language experience. Cebrian (2006) examined the perception of the tense-lax distinction in English vowel pairs /i/ ∼ /ɪ/ and /e/ ∼ /ɛ/ by native speakers of Catalan, whose native vowel system has /e/ ∼ /ɛ/ but the realization differs only in terms of the spectral values. The results showed that L2 listeners over-rely on duration regardless of their level of experience in the L2. On the other hand, it also seems to be the case that L2 experience has an influence in the perception of acoustic cues: Escudero and Boersma (2004) looked into the perception of English /i/ ∼ /ɪ/ by L1 Spanish speakers who were living in either southern England or Scotland and found out that L2 experience correlated to the listeners' ability to match the acoustic cue weightings shown by the native speakers of the linguistic community in which they were immersed.

3 Phonological features, acoustic cues, and perceptual tasks

3.1 Acoustic cues = phonological features?

While all the experiments mentioned above touch upon the idea that non-native speakers can(not) learn L2 segments, either by transferring existing features or by learning completely new ones, they are different in crucial ways. These experiments differ not only in the object to be measured (cues or features), but also in the assumptions made by the authors about them (for instance, whether they make the connection between cue and feature, or not). For instance, it is worth noting that some of these experiments (e.g., Bohn, 1995; Escudero and Boersma, 2004) do not straightforwardly refer to the idea of "feature" as an abstract phonological notion, since what the learners are being tested on is whether they are able to make a different use of the available acoustic cues in the input in order to acquire a new contrast. While this may count as a very concrete way to operationalize the abstract notion of feature into a specific acoustic cue, this is not *exactly* the same as testing for the acquisition of an abstract feature. Thus, if we assume that F1 values relate to [high] and [low], that F2 values relate to [back], and that both F1/F2 and duration relate somehow to [tense], then it could be hypothesized that L2 speakers who are learning how to process the acoustic cues in a target-like manner would in turn acquire an underspecified feature in the L1 that is relevant for a contrast in the L2.

However, while learning native-like cue weighting in L2 *may* lead to the creation of a new abstract feature, it does not necessarily entail the acquisition of the feature in question; that is, it might be a necessary condition, but not a sufficient one. For instance, it can be claimed that (a) learning new weights for the available cues could lead to a remapping of the cues onto different *existing* representations; that is, L2 sounds that were initially mapped onto the same L1 segment are now mapped onto two different ones. In the PAM/PAM-L2 framework, this would mean that learners would go from single-category assimilation to two-category assimilation. And (b), the learning of new cue-weightings could also lead to an in-between state where the difference is only noticeable at a pre-lexical level, but does not permeate to a more abstract, lexical representation.

Furthermore, do the specifics of the acoustic cue matter? In this regard, three points must be taken into consideration. Firstly, it is important that the acoustic cue associated to a contrast is relatively easy to perceive by the L2 listener. Archibald (2009) argues that the robustness of the acoustic cue is crucial, regardless of whether the learning operation is feature reassembly or feature acquisition. Experiments on the production of the /n/ ∼ /ɲ/ contrast in Spanish by L1 English/L2 Spanish speakers suggest that both the weakness of the relevant acoustic cue as well as the functional load of the contrast are possible causes for a lack of L2 convergence (Stefanich and Cabrelli, 2021). By taking only this into account, a great deal of the variation in the results of different tests could be explained.

Secondly, it might be tempting to assert that abstract features can be redeployed to the L2, without checking for the nature of the cue. However, (and this is something that needs to be stressed!), the presence of a feature in the L1 does not mean that both the L1 and the L2 will use the same cue for it. Consider for instance how [voice] is present in the plosives of both English and Spanish, but they are instantiated by different cues (voicing in Spanish vs. VOT in English); the effect of such differences in acoustic cues that implement the same contrast can be seen even in experienced bilinguals (Flege and Eefting, 1987). In a perhaps less extreme case, and if we assume that [round] is

3 ATR is identified as a pharyngeal feature and has been more widely used nowadays as it may account for an array of phonological phenomena in different languages (Vaux, 1996, 1999). However, the exact place for this feature in the geometry is immaterial for the purposes of this research; here we will use [tense], the traditional terminology used in L2 phonology.

present and contrastive in English,[4] perceiving front rounded vowels should not be that difficult for L1 English speakers; however, English speakers' difficulty in perceiving and producing the /u/ ∼ /y/ contrast in French has been widely studied (e.g., Flege and Hillenbrand, 1984; Gottfried, 1984; Flege, 1987) and increased experience does not seem to improve accuracy (Levy and Strange, 2008). Regarding German, studies in perceptual assimilation show that English speakers perceive /y/ mostly as /u/ (Polka, 1995) even after years of L2 experience (Mayr and Escudero, 2010). These studies suggest that even though the relationship between the [round] and the [+back, -low] in English is not the same as the one in Spanish, the F2 values of [+high] vowels that are not assigned to /i/ are nevertheless automatically parsed as [+back] *and therefore also* as [+round], which could be the case for Spanish as well.

Finally, one third issue is whether it is possible that a cue that is used in the L1 to parse a certain contrast [+/–back] may be reused in order to convey another contrast such as [+/–round]. It is worth noting that the relevant acoustic cue for the /u/ ∼ /y/ contrast is F2 (Delattre et al., 1952; Fant, 1971), which is also a cue used for vowel backness; therefore, for listeners of languages with no front rounded/back unrounded vowels in their inventory, /y/ may be perceived simply along the backness dimension (Lisker and Rossi, 1992), thus being categorized as exemplars of either /i/ or /u/. Then, the problem may not be the saliency of the cue, but the fact that L2 listeners make use of this cue in their L1 to break down the perceptual space into their L1 categories along a dimension that is already taken. Thus, and regardless of whether these L1 categories are far apart from each other (such as /i/ and /u/ in Spanish), the listener's L1 perception has already assigned boundary values across the cue dimension (i.e. F2 for backness). In this regard, listeners acquiring an L2 contrast that can be perceived through a new cue where no boundaries for L1 phonemes have been assigned would have less difficulty than when they learn a contrast where the relevant cue is already in use.

3.2 Perceptual tasks and their relationship to levels of representation in perception

The idea that perceptual tasks yield different results due to the level of representation to which they are aimed has been tested empirically in several works, for instance, in perception of L2 input which is phonotactically illegal in L1 (Freeman et al., 2022) and in L1 pre-lexical perception (Gerrits and Schouten, 2004). A model that considers different modes of perception which are prompted by different tasks is the Automatic Selective Perception model (Strange, 2011), which proposes a *phonological* mode and a *phonetic* mode of perception. Listeners engage in phonological mode mostly when using the L1, and it recovers only the necessary information for word recognition. This is achieved by means of over-learned, automatic selective perception routines that aim to make perception efficient; tasks that trigger phonological mode involve lexical decisions or grammaticality judgments. On the other hand, the phonetic mode is characterized by a more selective focus on the acoustic cues, which involves a higher cognitive load; tasks that elicit information under phonetic mode are those where the attentional focus is on the acoustic input, for instance, discrimination or category goodness (Strange, 2011, p. 460). Furthermore, the ASP model distinguishes between auditory and perceptual saliency. While auditory saliency is largely language-independent and refers to physiological responses to acoustic stimuli, perceptual saliency refers to the strength of the reaction to stimuli, which is moduled by linguistic experience (Strange, 2011, p. 458). Ultimately, it can be argued that the saliency of a given cue cannot be detached from the linguistic system that hosts it; thus, the only potential tool for quantifying saliency in behavioral experiments is to either look into acoustic values and calculate raw differences, or to manipulate the cues in a given stimuli.

One further point made in the ASP model is the difference between automatic and attentional processing, where automatic processing takes place when processing L1 sounds, or in L2 processing that has been automatized enough to resemble that of the L1. Automatic processing is expected to yield shorter reaction times, while attentional processing requires more time. When taking L1 reaction times as a baseline, especially during a task that requires accessing lexical representations, it is possible to estimate how well-learned a given L2 perceptual routine is.

In sum, the question whether learning new L2 segments is licensed by the presence of a given redeployable feature in the L1, or if new features are also learnable, is a rather complex one. The complexity lies not just on the lack of homogeneity in the methodology, but also on the intricate relationship between acoustic cues, phonological features, and perceptual tasks. Thus, the usually stated view that feature-based approaches have little predictive power in L2 acquisition should be reconsidered on the basis of the points made above: is feature redeployment independent of acoustic cues, and therefore, is redeployment fully abstract or phonetic-driven? And if the latter, does the nature, saliency, and L1 use of the cue play a role in the potential redeployment of a feature? The present study attempts to shed light on these questions.

4 This study

The current study tests whether adult Spanish learners of German can combine [-back] with [+round] in order to acquire /y/ as a segment that contrasts with /u/ (e.g., *Blüten/bluten;* 'blossom'/'to bleed') via [back], and whether acquiring the contrast between /u/ and /y/ is easier than learning the /i/ ∼ /ɪ/ contrast, where the addition of a new feature (namely, [tense]) is required. The Spanish vowel system /i e a o u/ does not have front rounded vowels, but it does have front unrounded vowels /i e/ and back rounded vowels /o u/. This vowel system always presents [+round] with [+back] and never with [-back], for which acquiring /y/ requires restructuring the possibilities that the L1 gives in terms

4 Granted, this is a difficult assumption, since back, non-high vowels vary greatly across varieties of English. However, some of the many feature matrices that have been proposed for the English vowel system (e.g., Giegerich, 1992, p.110) show that for instance /ɑ/ and /ɔ/ are contrastive only by (round). Likewise, the existence of a mid, back unrounded vowel such as /ʌ/ shows that English can at least decouple the [+round] feature from [+back, −low], which is not the case of Spanish.

of which features are allowed to combine into one segment. The learning process referring to the reassembly of features to create a new segment has been referred to as feature redeployment (Archibald, 2005). However, it is worth noting that this term refers to features that bear contrast on their own; in Spanish, [+round] is predictable from the features [+back, −low], for which it does not count as a fully contrastive feature in the L1. Since a potential scenario wherein [+round] in Spanish rebundles with [+high, −back] is not covered by the original definition of redeployment, we will extend this term to include L1 features that are present but do not bear contrast.

On the other hand, when L1 Spanish learners of German acquire the /i/ ∼ /ɪ/ contrast (e.g., *Miete/Mitte;* 'rent'/'middle') the learning task is of a different nature. The only difference between /ɪ/ and /i/ in German is given by the feature [tense], which is a feature that Spanish does not have. However, Spanish does use duration (one of the acoustic cues involved in production and perception of the tense feature) as one out of several (and equally relevant) cues for stress (Ortega-Llebaria, 2006), but not for signaling phonemic contrasts; thus, L1 Spanish learners of German could in theory transfer this cue for the perception of new phonemic contrasts. Nevertheless, and as stated in section 3.1., the use of a given cue in the L1 does not entail the existence (or the automatic acquisition) of an abstract feature such as [tense], although it can be assumed that this is a preliminary step in order to acquire the abstract feature. We will call this learning task feature acquisition.

Furthermore, the differences in acoustic cues in terms of their saliency could be privileging one scenario over the other; that is, whether learning the contrast between /u/ and /y/ is easier than /i/ ∼ /ɪ/, or vice versa. As mentioned in section 2, experiments suggest that L1 Spanish learners of English are more sensitive to duration, as it seems to be for them an auditorily salient cue; thus, this could mean that they would have less difficulty acquiring a new feature with an unused, salient cue, than redeploying an existing one with a cue that the L1 uses for the perception of the front/back contrast.

Thus, the following hypotheses can be considered:

1) *Feature redeployment is as difficult as feature acquisition:* In this case, the presence of the feature in question in the L1 [i.e., (round)] does not facilitate acquisition of a contrast involving this feature in L2; furthermore, a contrast involving a new feature is also difficult to learn. The results in Barrios et al. (2016) seem to support this hypothesis, which would also be supported by this study if the perception of the non-native contrasts /i/ ∼ /ɪ/, and /u/ ∼ /y/ is significantly worse than their baseline L1-like performance (/u/ ∼ /i/).
2) *Feature redeployment is easier than feature acquisition:* Brown's (1998, 2000) main claim is that features that are present/active and fully specified in the L1 should be able to rebundle in order to create a new segmental representation. On the other hand, Brown's hypothesis states that underspecified features will not be acquired. If we restrict ourselves to the definition of redeployment as the transfer of fully contrastive features only, then neither contrast should be easier to acquire than the other. However, by extending the term to cover the transfer of redundant features, we would see better performance with the /u/ ∼ /y/ contrast (eventually, comparable to /i/ ∼ /u/), assuming that redeployment is licensed by the presence of the feature in the L1, and not by its ability to bear contrast on its own in the L1 grammar.
3) *Feature acquisition is easier than feature redeployment:* In this scenario, the blank slate benefits learning, and/or it might be aided by the acoustic saliency of the cues involved in a given contrast. Such results would be in line with Bohn (1995) and the Desensitization Hypothesis. Better performance with the /i/ ∼ /ɪ/ contrast (and comparable to /u/ ∼ /i/) would support this hypothesis.
4) *Feature redeployment is as easy as feature acquisition:* the most optimistic scenario is where native-like performance can be seen in both cases. While at least there is evidence that a fair amount of learning does take place (e.g., Escudero and Boersma, 2004), a scenario with ceiling effects for feature acquisition has, to the best of our knowledge, not been attested. Here, both nonnative contrasts /i/ ∼/ɪ/ and /u/ ∼ /y/ would be comparable to /u/ ∼ /i/.

In order to test these hypotheses, the perceptual tasks need to tap different levels of representation. Since the /i/ ∼ /ɪ/ contrast is aided by a new, acoustically salient cue, we set out to probe whether cue saliency facilitates perception at a more superficial level (that is, a phonetic mode of perception, without resorting to long-term representations). In this sense, an AX discrimination task would shed light on whether listeners are able to weight the acoustic cues involved in the contrasts in a target-like manner. However, potentially increased sensitivity to a cue does not necessarily entail the acquisition of a contrast at a phonemic level; therefore, a task that prompts the listener to make use of abstract representations, such as a categorization task, is needed. In this regard, a picture identification task is ideal for this purpose as it excludes potential orthographic effects in the responses as well as it probes the existence of L2 phonemic representations encoded at the lexical level.

Furthermore, reaction times (RT) are also measured, as it will provide a better understanding of the underlying cognitive process in the L2 learner, especially in terms of the automaticity of their perceptual routines. Differences in RTs by contrast, and particularly between native and nonnative contrasts, should indicate whether there is a substantial difference when processing the stimuli.

5 Methods

The experiment was carried out online. Most of the participants were recruited by word of mouth, while others were contacted via social networks. The experiment was set up on an online platform created specifically for this purpose in HTML and JavaScript using JsPsych (De Leeuw, 2015) in order to measure reaction times in a precise manner.[5] Participants were asked to use their keyboard instead of the mouse when giving responses (i.e., to

5 As of 2020, JavaScript-based online experiments have been deemed less accurate than lab-based studies, but still within acceptable ranges in terms of variability (De Leeuw and Motz, 2016; Pronk et al., 2020; Anwyl-Irvine et al., 2021), where jsPsych shows a mean precision between 3 and 7 ms across operating systems and browsers (Bridges et al., 2020).

press Q for the option showing on the left side of the screen, and P for the one to the right). All instructions were given in German.

5.1 Participants

Participants ($N = 19$, mean age: 35.9 years) identified themselves as native speakers of Spanish, with a self-reported B2 CEFR level of German. Most of them learned German after age 18 (two participants started at age 9 and 5) and were living in Germany for at least a year at the time they were tested (one participant did not report, and another one had spent only 2 months). They also reported having used a variety of learning methods, of which the most frequent were language classes, both at their home country before their arrival and in Germany. Two of them reported two native languages: one Spanish and Catalan, and another Spanish and English; these participants were excluded.

5.2 Materials

Nineteen German words containing the vowels /y/, /i/, /u/, and /ɪ/ were recorded by a phonetically trained female native speaker of German. The selection of a small number of words is due mostly to the online nature of the experiment, where participants are more likely to get distracted or quit the experiment before finishing, and considering that repetition was needed in order to run statistical analyses. Furthermore, the picture identification task required that the words could be easy to represent in an image, which also increases the likelihood of being understood by B2-level speakers.

In order to avoid differences in intonational curves and obtain similar durations, the words were embedded in the carrier sentence *Ich sage* ____ *noch mal* ('I say ____ again'). Each word was recorded three times, and for each word one token was selected (ideally without clicks, vocalized consonants, or creaky voice). All tokens had the same intonation curve due to the embedding in the carrier sentence. For the AX task, trials with different words were manipulated with Praat so that the both words also had the same onset, coda, and schwa lengths while keeping the original pitch; this was to ensure that the only difference was the vowel itself and nothing else in the stimuli. While onsets were rather consistent in terms of duration, words ending with vocalized consonants or sonorants had different durations. In pairs where the differences in duration was small enough to not distort the pitch and both vowels were tense, the sound files were scaled to the average duration. In pairs with the /i/∼ /ɪ/ contrast and/or with considerably different durations, the final sound in the longer word was shortened, avoiding abrupt changes in intensity and respecting zero-crossings. The list of words can be seen in Table 1, arranged by minimal pairs per contrast (see Table 1). In order to rule out a potential assimilation of /y/ to /i/ (*contra* the evidence from L1 English learners of French and German), a fourth minimal pair /i/ ∼ /y/ was added.

5.3 Procedure

After the consent and language background questionnaire, L1 Spanish learners of German were tested for perception of the /u/ ∼ /y/, /i/ ∼ /ɪ/, /i/ ∼ /y/, and /u/ ∼ /i/ contrasts with an AX (same-different) discrimination task with an ISI of 1500 ms (as in Barrios et al., 2016), followed by a two-way, forced-choice picture identification task. The online setup included a break screen between the tasks, which recommended the participant to take a 5-min break. The total experiment time was around 10 min.

5.3.1 AX discrimination task

The words were taken apart from the carrier sentence with Praat (Boersma and Weenink, 2024) and presented to the participants with an ISI of 1500 ms. Participants heard two German words consisting of a minimal pair from Table 1 (e.g., *Blüten* and *bluten);* duplicates of the same word were also included. There was a total of 48 randomized trials. In trials where the words were different, the minimal pair was presented twice (once with each word in first place; that is, WordA – WordB, and then WordB - WordA). There was also one instance of WordA – WordA and one of WordB – WordB. During each trial, the screen showed the question: *Sind die Wörter gleich oder unterschiedlich?* ("Are the words the same or different?") The buttons provided were *Gleich* ("Same") on the left and *Unterschiedlich* ("Different") on the right. Participants had a maximum time of 5 s to provide a response.

5.3.2 Two-way forced-choice picture identification task

Participants carried out a picture identification task where only two alternatives were given. The same words in the previous task were here presented in isolation, while showing the participants two pictures (e.g., if the aural stimulus was the word *Blüten*, the pictures shown corresponded to *Blüten* and *bluten*) with the question *Welches Wort haben Sie gehört?* ("Which word did you hear?"). Participants had a maximum time of 5 s to make a decision. Each stimulus was shown three times; with 18 trials per contrast, there was a total of 72 randomized trials.

6 Results

6.1 AX discrimination task

Sensitivity (*d'*) yields a single score per participant by vowel contrast, and is calculated by taking into account the z-scores of the total counts of hits minus false alarms (MacMillan and Creelman, 1991). In order to compute *d'*, all trials containing e.g., the /u/ ∼ /i/ contrast (WordA-WordB, WordB-WordA, WordA-WordA and WordB-WordB for all three minimal pairs) were tabulated in terms of the participant's hits (answered "different" when the words were different), false alarms (answered "different" when the words were the same), misses (answered "same" when the words were different), and correct rejections (responded "same" when the words were the same). These *d'* scores were calculated with *R* (v.4.3.1; R Core Team, 2023) and the package *SensR* (Christensen

TABLE 1 List of German words used in the experiment.

/u/ ∼ /i/	/i/ ∼ /ɪ/	/u/ ∼ /y/	/i/ ∼ /y/
Tour – Tier (“tour” – “animal”)	*Miete – Mitte* (“rent” – “middle”)	*Tour – Tür* (“tour” – “door”)	*Biene – Bühne* (“bee” – “stage”)
Spulen – spielen (“spools” – “to play”)	*Schief – Schiff* (“crooked, not straight” – “ship”)	*bluten – Blüten* (“to bleed – blossom”)	*spielen - spülen* (“to play” – “to rinse”)
Stuhl – Stiel (“chair” – “stick”)	*Stiel – still* (“stick” – “quiet”)	*Spulen – spülen* (“spools” – “to rinse”)	*Kiel – kühl* (“Kiel” – “cool”)

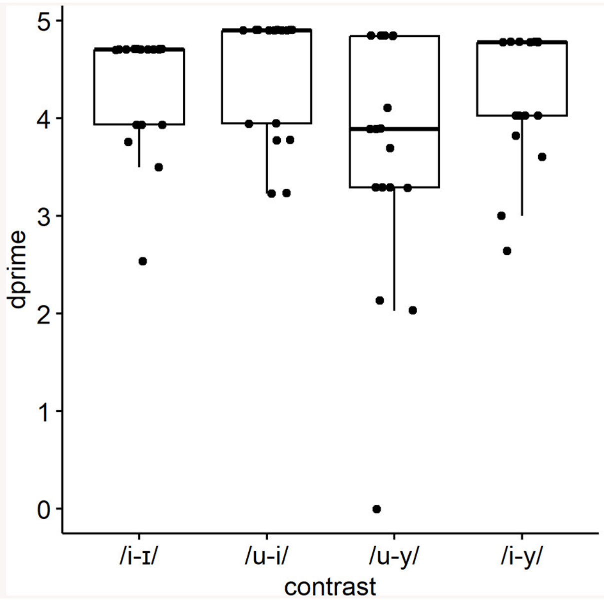


FIGURE 1
Sensitivity (d-prime) by contrast, AX discrimination task. The dots represent the d’ value obtained by each participant for each contrast.

and Brockhoff, 2017). In order to correct infinity values, a log-linear approach was used. Since the *d*’ data violated normality assumptions, a series of non-parametric tests using the *R* package *rstatix* (Kassambara, 2023) were carried out. A Friedman test showed a significant effect of contrast on sensitivity expressed as d’ values (χ^2 $F(3) = 13.4$, $p < 0.01$). A Kendall’s W test for effect size (W = 0.26) indicates a small effect. A *post-hoc*, two-sided Wilcoxon signed-rank test with a Bonferroni adjustment comparing performances in all three non-native contrasts against the native /u/ ∼ /i/ showed that the only significant difference is that between /u/ ∼ /i/ and /u/ ∼ /y/ (T = 136, $p < 0.05$). Figure 1 shows the distribution of *d*’ scores by contrast, where the differences in distribution by contrast can be seen: overall, the participants’ d’ was lower for the /u/ ∼ /y/ contrast.

Regarding reaction times (RT), trials with null responses were discarded. Then, two analyses were carried out. First, differences in RT for all responses across stimuli pairs grouped by the vowels present in the stimuli (e.g., /u/ ∼ /i/, /i/ ∼ /i/, /u/ ∼ /y/) were probed with a mixed effects model using the R package *lme4* (Bates et al., 2014), with stimuli pair as a fixed effect, and subject and item as random intercepts. The model shows a significant effect of stimuli pair ($\chi^2 = 21.7$, df = 7, $p < 0.001$). Figure 2 shows the distribution of RT values for all stimuli pairs.

A comparison across all pairs (same and different) with least-squared means was calculated using the *emmeans* package (Lenth, 2022). The reaction times for stimuli pair /u/ ∼ /y/ showed to be significantly higher than all other pairs, except /i/ ∼ /ɪ/, /ɪ/ ∼ /ɪ/, and /y/ ∼ /y/. Table 2 shows the statistical values for each significant comparison.

The values show that participants are considerably slower when discriminating between /y/ and /u/ than all same stimuli pairs, with the exception of the /ɪ/ ∼ /ɪ/ and /y/ ∼ /y/ pairs. On the other hand, participants were equally slow at reacting to the different pair /i/ ∼ /ɪ/, which suggests that the non-native /I/ triggers longer reaction times, even though the decisions made throughout the task with pairs involving this sound were mostly correct. It is also worth noting that trials with the /i/ ∼ /y/ pair yielded faster reaction times than those with /u/ ∼ /y/, which suggests that the participants perceive these sounds as clearly different and thus do not perceptually assimilate /y/ to their native category /i/.

One further mixed-effects model but only with the trials with different stimuli, with stimulus pair as predictor, and item and subject as random intercepts was fitted, and again there was a significant effect of stimulus pair ($\chi^2 = 14.39$, df = 3, $p < 0.01$). The least-squares mean analysis showed the same significant differences for /u/ ∼ /y/ and /u/ ∼ /i/ ($\beta = 354.67$, SE = 107, $t = 3.32$, $p < 0.05$) as well as for /u/ ∼ /y/ and /i/ ∼ /y/ ($\beta = 345.60$, SE = 107, $t = 3.24$, $p < 0.05$).

6.2 Picture identification task

Regarding the picture identification task, we carried out a mixed-effects model with correct/incorrect as the dependent variable, vowel contrast as predictor, and subject and item as random intercepts ($\chi^2 = 66.93$, df = 3, $p < 0.001$). The participants showed very high performance when choosing between /u/ and /i/ (95% correct); similar results were obtained for the non-native contrast /i/ ∼ /y/ (99% correct). On the other hand, non-native contrasts /i/ ∼ /ɪ/ and /u/ ∼ /y/ yielded significantly lower correct counts: 76% and 73%, respectively. Figure 3 shows the categorizations by contrast.

A *post-hoc* comparison between all the contrasts with least-square means (with Tukey adjustment for *p*-values) showed significant differences across all contrasts, except between /i/∼ /ɪ/ and /u/ ∼ /y/, as well as / u/ ∼ /i/ and /i/ ∼ /y/. Table 3 presents the results of pairwise comparisons.

This task’s RTs yielded different patterns by contrast, where the contrasts /u/ ∼ /i/ and /i/ ∼ /y/ displayed lower reaction times (Figure 4). However, a mixed-effects model including all

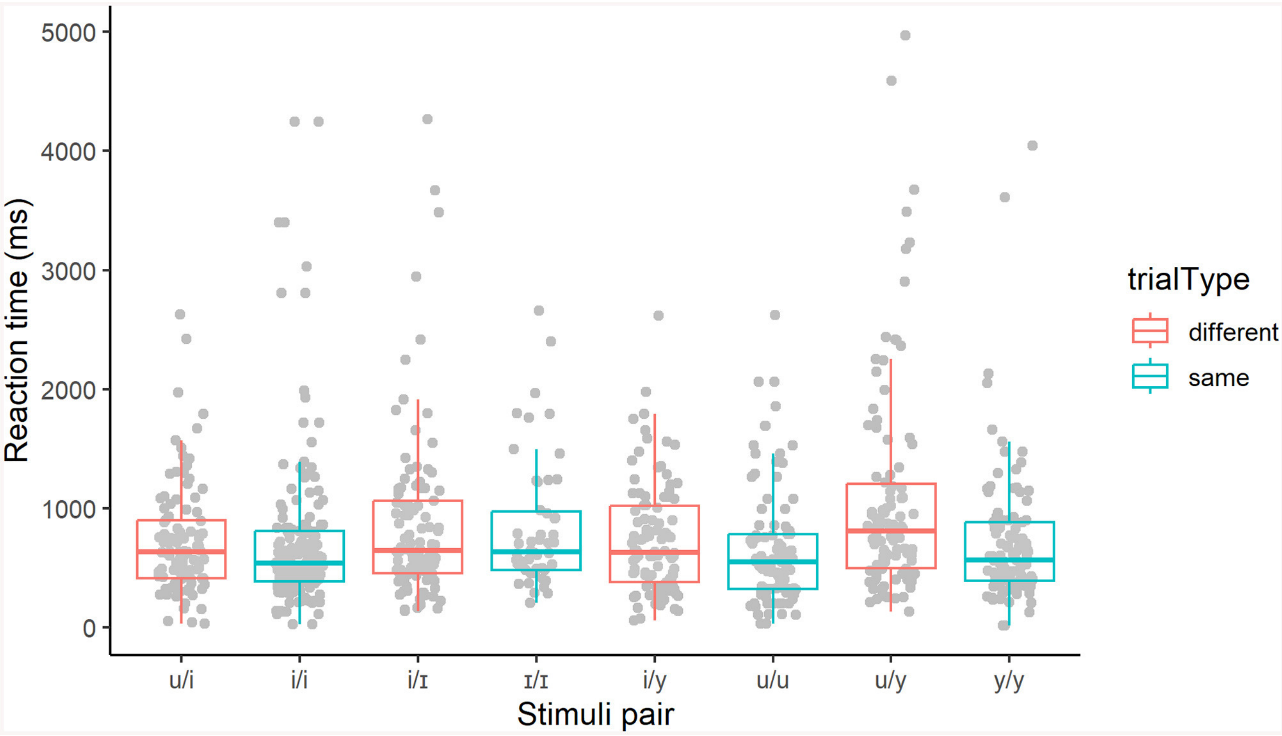


FIGURE 2
AX discrimination task: Distribution of reaction times (RT) by stimuli pair (all responses).

TABLE 2 Statistics for pairwise comparisons, reaction times.

Contrast comparison	Estimate	SE	t.ratio	*p*-value
/u/ ∼ /y/ and /u/ ∼ /i/	353.38	104.7	3.38	<0.05
/u/ ∼ /y/ and /i/ ∼ /y/	340.72	98.7	3.45	<0.05
/u/ ∼ /y/ and /i/ ∼ /i/	345.60	104.6	3.31	<0.05
/u/ ∼ /y/ and /u/ ∼ /u/	426.75	111.2	3.83	<0.01

responses except time-outs, with vowel contrast as fixed effect and subject and item as random intercepts showed that these differences in reaction times across vowel contrasts are not significant ($p = 0.11$).

7 Discussion

The present study has attempted to determine whether the /u/ ∼ /y/ contrast found in German is more likely to be acquired by L1 Spanish learners of German than the /i/ ∼ /ɪ/ contrast, with four hypotheses in sight. Hypothesis 1 points to a no-learning scenario, with the native-like /u/ ∼ /i/ contrast unmatched in perception. Hypothesis 2 follows feature-based approaches, wherein accurate perception of new L2 sounds can be facilitated by the presence of the same feature in in the L1. Hypothesis 3 favors the acquisition of the /i/ ∼ /ɪ/ contrast, as duration is an auditorily salient cue that offers learners a blank slate where new categories can be created. Finally, Hypothesis 4 predicts a full-learning scenario where all nonnative contrasts are perceived in a native-like manner. To this end, two perceptual tasks were carried out: one tapping phonetic knowledge (AX discrimination), and another tapping phonological knowledge (picture identification).

First, the results suggest that /y/ is unequivocally assimilated to /u/ (or rather, perceived as certainly not /i/), in accordance with studies on the perception of /y/ by L1 English learners of German and French. This can be seen in the results of the discrimination task, where d-primes and reaction times for /i/ ∼ /y/ were not significantly different than those for the native pair /u/ ∼ /i/. Furthermore, the picture identification task yielded excellent accuracy for /i/ ∼ /y/. This suggests that, in the presence of a [+/–round] contrast such as /i/ ∼ /y/, perception is driven by the acoustic cue F2 in such a way that a value between the usual ones for categories /i/ and /u/ is perceived as [+back], and not as [-back].

The fact that the same group of listeners showed lower accuracy for both non-native contrasts in the picture identification task (where more abstract representations are in use by the listener) but performed better with a contrast involving a new feature (plus an additional cue) in the AX discrimination task (which focuses on prelexical perception) shows that the acquisition of the /u/ ∼ /y/ contrast by native speakers of Spanish is not aided by the presence of the relevant phonological features [round, back, high] in the L1. Even if we interpret the more-than-above-chance performance in both tasks as a signal of facilitation, the even better performance with /i/ ∼ /ɪ/ in the AX discrimination task shows that the active presence of the features in question in the L1 are not necessarily an advantage in relation to absent features such as [tense]. However, the following points should be taken into consideration.

7.1 Differences by task

The results of the AX discrimination task suggest that the /u/ ∼ /y/ contrast is more difficult to acquire than /i/ ∼ /ɪ/, with the latter being closer to their performance in the native contrast /u/ ∼ /i/; based on this task alone, it may be concluded that L2 speakers are more likely to acquire a contrast when the feature in question is *not* present in the L1, therefore supporting Hypothesis 3. On the other hand, the results for the picture identification task show that both non-native contrasts are significantly more difficult than the native one; that is, this task would support the idea that acquiring non-native contrasts is difficult regardless of whether the feature in question can be transferred or not (Hypothesis 1). Furthermore, the tasks yielded different patterns for reaction times: while the AX discrimination task was able to yield significant differences, the picture identification task did not.

The best approach for explaining the seemingly contradictory results would be to assume that discrimination is chiefly a phonetic task, regardless of the relatively long ISI and minimal pairs used. While shorter ISI might facilitate the retrieval of acoustic details stored in the short-term memory, the possibility that the participants were still able to remember the acoustic detail of the previous word after 1.5 s is also possible, albeit to a lesser extent. Furthermore, the stimuli were recorded by one speaker, which does not require the listener to filter out interspeaker variation. Hence, learners may be less likely to rely on long-term representations when providing answers to the task. On the other hand, the picture identification task would not depend on acoustic detail and the listeners would need to rely on long-term representations; thus, changes in the perception of non-native contrasts due to L2 experience is more likely to be attested when

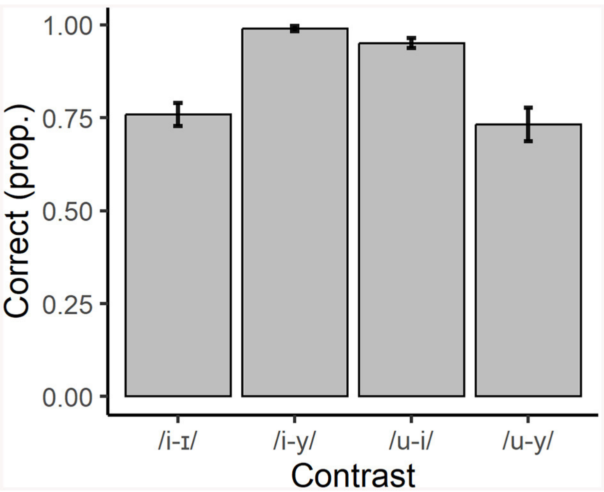


FIGURE 3
Mean proportion of correct responses by contrast in the picture identification task. Error bars show a 95% confidence interval.

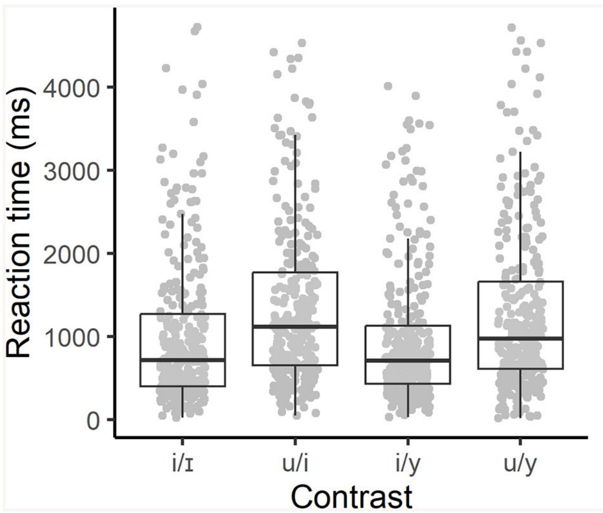


FIGURE 4
Reaction times by contrast (all responses), picture identification task. Dots show RTs per trial.

TABLE 3 Pairwise comparisons, identification task.

Contrast	Estimate (log odds)	SE	z.ratio	*p*-value
/u/ ∼ /i/ and /i/ ∼ /ɪ/	1.93	0.36	5.38	<0.0001
/u/ ∼ /i/ and /i/ ∼ /y/	−1.67	0.66	−2.53	0.056
/u/ ∼ /i/ and /u/ ∼ /y/	2.08	0.36	5.82	<0.0001
/i/ ∼ /ɪ/ and /i/ ∼ /y/	−3.59	0.62	−5.77	<0.0001
/i/ ∼ /ɪ/ and /u/ ∼ /y/	0.15	0.28	0.55	0.94
/i/ ∼ /y/ and /u/ ∼ /y/	3.73	0.62	6.01	<0.0001

tapping surface-like knowledge, and less so when looking into higher levels of representation.

The facilitatory effect of the AX discrimination task, with no inter-speaker variation, can be seen in the reaction times: not requiring a long-term representation to respond to the stimuli may have removed any potential differences due to lack of knowledge of a given stimulus. The significantly slower reaction times for /u/ ∼ /y/ can be interpreted exclusively as a function of the difficulty of perceiving the difference between sounds, even when having the opportunity to direct their attention to the cue in question.

From a formal phonological perspective, the claim in Hancin-Bhatt (1994) about feature prominence, that is, how relevant the feature is in the grammar of the L1, may also account for the poorer performance for /u/ ∼ /y/ in the AX discrimination task. In Spanish, the feature [+round] is redundant with [+back, -low], for which its prominence is not high. However, it is worth noting that this same contrast in L1 English learners of French is also difficult to perceive and produce, even when English does have [+back, -round], although only in low vowels. Here one could assume that since English has segments with the [+back, -round] and the [+back, +round] feature bundling, it could be easier for L1 English learners of L2 French to redeploy [+/− round] to [-low] vowels. It seems that this is not the case, and that in the end L1 speakers of both English and Spanish behave the same when it comes to the perception of /y/. This suggests that perhaps feature prominence needs to be coupled with the salience and functional load of the acoustic cues involved in the contrast.

One further point to consider is that the frequency effect from the words chosen for the stimuli could have affected the results for the picture identification task. In this regard, an item analysis shows a higher number of incorrect responses in the discrimination task for the stimuli pair *bluten/Blüten* (47%, while all other items show an incorrect rate between 0% and 16%), and a 33% of incorrect responses in the picture identification task. A post-experiment analysis of word frequency with German corpora (Leipzig Corpora Collection, 2022) was carried out, which showed that the frequency class of *bluten* and *Blüten* is 15 and 12, respectively (the higher the number, the less frequent the word); however, the analogous pair *Spulen/spülen* (frequency class 17 and 16, respectively) yielded higher correct response ratios (95% in discrimination and 79% in identification). It is worth noting that even though frequency in corpora may provide a range of probability as to whether an L2 learner has come across a given word, it may not be the best way to estimate the learner's familiarity with a certain lexical item, as L2 learners usually learn they vocabulary through modified input (e.g., textbooks, student materials) that does not contain the same lexical items and grammatical structures as the texts contained in corpora (newspapers, websites, etc. written and read by L1 speakers).

Furthermore, the reaction times for the /u/ ∼ /y/ contrast in the picture identification task were not significantly slower than those obtained for the other contrasts, which suggests that the *bluten/Blüten* pair may have been more difficult to perceive not due to lexical frequency, but perhaps to the phonological context: together with *Miete/Mitte*, in this pair the vowel is preceded by a sonorant and followed by a plosive.

7.2 Differences by contrast

The differences in performance by contrast observed in the AX discrimination task suggest that learners could be more sensitive to auditorily salient acoustic cues when the attention is fully focused on perceiving differences between sounds, and not on lexical access. Table 4 shows the differences in the acoustic cues present in the stimuli. Again, it is worth noting that these differences in acoustic cues do not play a significant role when the task prompts listeners to rely on long-term representations.

The values in Table 5 show how the acoustic cues differ by contrast. Two large differences in F1 and duration can be seen in /i/ and /ɪ/ (F1 and duration); on the other hand, one large difference between /u/ and /y/ is observed (F2). F3 values show relatively small differences across contrasts, especially in Bark. Regarding the perceptual assimilation of /y/, these values show, and especially in the Bark scale, that /y/ is much more distant from /u/ than /i/ in terms of F2; yet, the listeners seem to favor a perceptually and auditorily distant vowel over the closer one. Duration has also been deemed a more perceptually salient cue than frequency (Bohn, 1995), which seems to be relevant in a task where the focus is not on abstract representations but in picking up acoustic cues; likewise, most of the studies showing successful learning of new perceptual cue weightings and/or acquisition of new features focus on duration or quantity.

However, it is worth stressing that (a) duration seems to be salient to L2 listeners even when this cue is not relevant for L1 perception; and (b) the level of perceptual saliency may be influenced not only by linguistic experience, but also by its role in the L1 regarding perception of other feature contrasts. Judging at least by the results obtained here, F2 seems to be a perceptually salient cue that is weighted by L1 Spanish speakers in such a way that non-peripheral F2 values are categorized as /u/ regardless of its acoustic proximity to /i/; on the other hand, F1 and duration seem to be more acoustically salient as the differences are larger. Furthermore, the studies in section 2 show how duration is a cue that Spanish speakers are reported to rely on, despite being irrelevant in their L1 for the perception of phonemic contrasts. It seems then, that acoustic cues to L1 vowel contrasts are more difficult to be recalibrated in order to perceive a category whose values lie between two L1 vowels; on the other hand, duration provides a blank slate that does not need to be shared with other L1 categories.

7.3 The role of the cue feature redeployment

While the discussion above seems to point to the conclusion that Brown's FBM has no predictive power, several aspects need to be taken into account. Apart from the limitation derived from the role of [round] in Spanish, this study showed a significant difference between pre-lexical and lexical levels of perception, which are conceptualized in terms of phonetic and phonological mode by Strange (2011) ASL model. In this regard, it is likely that tasks prompting a phonetic mode of perception show better results, as the listener's attentional focus is on the acoustic cues. But

TABLE 4 Mean acoustic values of vowels in stimuli and standard deviation (SD), in Hertz and Bark.

Vowel	F1 Hz (SD)	F1 Bark (SD)	F2 Hz (SD)	F2 Bark (SD)	F3 Hz (SD)	F3 Bark (SD)	Duration, ms (SD)
i	244 (23)	2.58 (0.23)	2428 (653)	13.94 (2.31)	3559 (356)	16.78 (0.73)	150 (50)
ɪ	502 (49)	4.97 (0.42)	1901 (179)	12.54 (0.64)	3042 (293)	15.71 (0.68)	77 (2)
y	290 (39)	3.03 (0.39)	1726 (147)	11.91 (0.56)	2615 (153)	14.69 (0.39)	141 (48)
u	352 (72)	3.61 (0.67)	719 (224)	6.58 (1.52)	2956 (98)	15.53 (0.23)	156 (18)

TABLE 5 Differences in acoustic cues by contrast.

Contrast	ΔF1 (Hz)	ΔF1 (Bark)	ΔF2 (Hz)	ΔF2 (Bark)	ΔF3 (Hz)	ΔF3 (Bark)	Δ Duration (ms)
i ∼ ɪ	258	2.39	527	1.4	517	1.07	85
u ∼ i	108	0.45	1709	7.36	603	1.25	32
u ∼ y	62	0.59	1007	5.33	341	0.84	10
i ∼ y	46	1.04	702	2.03	944	2.09	42

even though this type of task yields better performance, there is a limitation (and this is the second point): is the cue in question salient enough regardless of its relevance in the L1? What is the role of this cue in the L1? And how many cues are available for the learner?

A further point, also related to acoustic cues, is the relationship between cues and features. Is attunement to a certain cue a *conditio sine qua non* for the acquisition of a phonological feature? Our research has shown that even though duration is not relevant in the L1 for establishing phonemic contrasts, the listeners in this study are able to use this cue in the L2, but it still seems to be insufficient when the cognitive load in the task increases. However, the moderate-to-high performance in both tasks shows that learning does take place, so Hypothesis 1 (i.e. that both reactivating and rebundling features are equally difficult learning tasks) is valid only insofar as we compare this to L1 performance. One thing is clear: learning does take place, though the performance cannot be compared to native-language contrast perception.

8 Conclusion

Nonnative speakers with a reasonably high level of L2 experience are relatively more successful when discriminating between sounds where a new feature is acquired, than when the acquisition of a phoneme requires existing features (however underspecified) to be redeployed; however, neither contrast is easier than the other one when it comes to identification. These results support conflicting hypotheses: while the discrimination task supports a theory of acquisition along the lines of Bohn (1995) Desensitization Hypothesis, the identification task seems to replicate the results in Barrios et al. (2016). That is, this work found no supporting evidence of redeployment as posited by the FBM; however, further studies where the feature to be redeployed is fully contrastive would shed more light on the issue.

The evidence shown here suggests that the differences in performance between tasks and in the saliency of the acoustic cues involved (duration vs. F2) may be due to L2 phonological acquisition taking place at a surface level, which could explain why the L2 speakers' sensitivity to acoustic cues is higher when the cue is acoustically salient (assuming that the larger the difference in values for the cue, the more auditorily salient the cue is), regardless of whether the learning task is redeployment or acquisition of a new feature. This may in turn suggest that new, auditorily salient cues would aid the learning of completely new features over internal rearrangement of abstract features, albeit at a surface, pre-lexical level.

Finally, among the limitations in this study is the inability to completely control for the type of acoustic cues to be tested, either in terms of its saliency or informativity in the L1; furthermore, the familiarity of the words used in the experiment may have differed by participant, although the minimum proficiency requirement, plus the instructions being given in German, made it less likely for the participants to be completely unfamiliar with the lexical items. One further point to address is that we did not control for exposure to English, which has an analogous vowel contrast /i/∼/ɪ/, though the realizations are not quite the same. Nevertheless, we hope that the present study motivates further research looking into the role of abstract phonological features in L2 speech perception, where comparisons between L2 feature acquisition and redeployment can be made on the basis of equally salient acoustic cues, and contrasts whose features in the L1 grammar are fully contrastive.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://github.com/pandabar/features.

Ethics statement

Approval for experiments within the Department of Linguistics was granted by the Ethics Committee at the University of

Konstanz on 04.02.21 (IRB statement 05/2021). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FB: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

References

Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., and Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behav. Res. Methods* 53, 1407–1425. doi: 10.3758/s13428-020-01501-5

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology* 5, 183–207. doi: 10.1017/S0952675700002268

Archibald, J. (2005). Second language phonology as redeployment of L1 phonological knowledge. *Can. J. Ling.* 50, 285–314. doi: 10.1017/S0008413100003741

Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Second Lang. Res.* 25, 231–233. doi: 10.1177/0267658308100284

Avery, P., and Rice, K. (1989). Segment structure and coronal underspecification. *Phonology* 6, 179–200. doi: 10.1017/S0952675700001007

Barrios, S., Jiang, N., and Idsardi, W. J. (2016). Similarity in L2 phonology: evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Sec. Lang. Res.* 32, 367–395. doi: 10.1177/0267658316630784

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

Best, C., and Tyler, M. (2007). "Nonnative and second-language speech perception: commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds B. Ocke-Schwen and M. J. Munro (Amsterdam: John Benjamins), 13–34.

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.

Boersma, P., and Weenink, D. (2024). *Praat: Doing Phonetics by Computer [Computer Program]. 6.1.* Available online at: http://www.praat.org (accessed December 29, 2023).

Bohn, O. S. (1995). "Cross language speech production in adults: first language transfer doesn't tell it all," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 279–304.

Bridges, D., Pitiot, A., MacAskill, M. R., and Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ* 8, e9414. doi: 10.7717/peerj.9414

Brown, C. (2000). "The interrelation between speech perception and phonological acquisition from infant to adult," in *Second Language Acquisition and Linguistic Theory*, ed. J. Archibald (Wiley-Blackwell).

Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Lang. Res.* 14, 136–193. doi: 10.1191/026765898669508401

Ćavar, M., and Hamann, S. (2011). "Phonemes, features and allophones in L2 phonology," in Generative Investigations: Syntax, Morphology, and Phonology, eds P. Bański, B. Łukaszewicz, M. Opalińska, and J. Zaleska, 308–320.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *J. Phonetic.* 34, 372–387. doi: 10.1016/j.wocn.2005.08.003

Christensen, R. H. B. and Brockhoff, P. B. (2017). *sensR - An R-Package for Sensory Discrimination*. R package version 1.5-0. Available online at: http://www.cran.r-project.org/package=sensR/

Clements, G. N. (1985). The geometry of phonological features. *Phonology* 2, 225–252. doi: 10.1017/S0952675700000440

Clements, G. N., and Hume, E. (1995). "The internal organization of speech sounds," in *The Handbook of Phonological Theory*, ed. J. Goldsmith (Cambridge, MA: Blackwell), 245–306.

De Leeuw, D., and Motz, J. R. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behav. Res. Methods* 48, 1–12. doi: 10.3758/s13428-015-0567-2

De Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* 47, 1–12. doi: 10.3758/s13428-014-0458-y

Delattre, P., and Hohenberg, M. (1968). (1968). Duration as a cue to the tense/lax distinction in German unstressed vowels. *J. Int. Rev. Appl. Ling. Lang. Teach.* 6, 367–90. doi: 10.1515/iral.1968.6.1-4.367

Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8, 195–210. doi: 10.1080/00437956.1952.11659431

Durand, J. (2005). "Tense/Lax, the vowel system of English and phonological theory," in *Headhood, Elements, Specification and Contrastivity: Phonological Papers in Honor of John Anderson*, eds P. Carr, J. Durand, and C. Ewen (Amsterdam: John Benjamins), 77–98. doi: 10.1075/cilt.259.08dur

Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Stu. Sec. Lang. Acquisition* 26, 551–585. doi: 10.1017/S0272263104040021

Fant, G. (1971). *Acoustic Theory of Speech Production: With Calculations Based on X-ray Studies of Russian Articulations (No. 2).* London: Walter de Gruyter.

Flege, J. E. (1987). The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification. *J. Phonetic.* 15, 47–65. doi: 10.1016/S0095-4470(19)30537-6

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research* 92, 233–277.

Flege, J. E., and Bohn, O. S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83.

Flege, J. E., Bohn, O. S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonet.* 25, 437–470. doi: 10.1006/jpho.1997.0052

Flege, J. E., and Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *J. Phonetic.* 15, 67–83. doi: 10.1016/S0095-4470(19)30538-8

Flege, J. E., and Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The J. Acoustic. Soc. Am.* 76, 708–721. doi: 10.1121/1.391257

Freeman, M. R., Blumenfeld, H. K., Carlson, M. T., and Marian, V. (2022). First-language influence on second language speech perception depends on task demands. *Lang. Speech* 65, 28–51. doi: 10.1177/0023830920983368

Gerrits, E., and Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Percept. Psychophys.* 66, 363–376. doi: 10.3758/BF03194885

Giegerich, H. (1992). *English Phonology: An Introduction*. Cambridge: Cambridge University Press.

Gottfried, T. (1984). Effects of consonant context on the perception of French vowels. *J. Phonetic.* 12, 91–114. doi: 10.1016/S0095-4470(19)30858-7

Hancin-Bhatt, B. (1994). Segment transfer: a consequence of a dynamic system. *Sec. Lang. Res.* 10, 241–69. doi: 10.1177/026765839401000304

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The J. Acoustic. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872

Jessen, M., Marasek, K., Schneider, K., and Classen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. *Proc. Int. Congr. Phonetic Sci.* 13, 428–431. doi: 10.21437/ICSLP.1996-404

Kassambara, A. (2023). *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2. Available online at: https://CRAN.R-project.org/package=rstatix/ (accessed December 30, 2023).

Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press.

LaCharité, D., and Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. *Proc. BUCLD* 23, 373–385.

Leipzig Corpora Collection (2022). *German News Corpus Based on Material From 2022.* Available online at: https://corpora.uni-leipzig.de?corpusId=deu_news_2022 (accessed December 11, 2023).

Lenth, R. V. (2022). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. R Package Version 1.7.2. Available online at: https://CRAN.R-project.org/package=emmeans (accessed December 30, 2023).

Leung, K. K., Jongman, A., Wang, Y., and Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The J. Acoust. Soc. Am.* 140, 45–58. doi: 10.1121/1.4954737

Levy, E. S., and Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *J. Phonetic.* 36, 141–157. doi: 10.1016/j.wocn.2007.03.001

Lisker, L., and Rossi, M. (1992). Auditory and visual cueing of the [±rounded] feature of vowels. *Lang. Speech* 35, 391–417. doi: 10.1177/002383099203500402

MacMillan, N. A., and Creelman, D. (1991). *Detection Theory: A User's Guide*. Cambridge: CUP Archive.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. London: Routledge.

Major, R. C. (2008). "Transfer in second language phonology: a review," in *Phonology and Second Language Acquisition*, eds J. G. Hansen Edwards and M. L. Zampini (Amsterdam: John Benjamins Publishing Company).

Mayr, R., and Escudero, P. (2010). Explaining individual variation in L2 perception: rounded vowels in English learners of German. *Biling. Lang. Cognit.* 13, 279–297. doi: 10.1017/S1366728909990022

McAllister, R., Flege, J. E., and Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *J. Phonetic.* 30, 229–258. doi: 10.1006/jpho.2002.0174

Odden, D. (2011). "The representation of vowel length," in T*he Blackwell Companion to Phonology*, eds M. Oostendorp, C. J. Ewen, E. Hume, and K. Rice (New York, NY: John Wiley and Sons), 1–26. doi: 10.1002/9781444335262.wbctp0020

Ortega-Llebaria, M. (2006). Phonetic Cues to Stress and Accent in Spanish. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, edited by Manuel Díaz-Campos, 104–18. Somerville, MA: Cascadilla Proceedings Project.

Pajak, B., and Levy, R. (2014). The role of abstraction in non-native speech perception. *J. Phonetics* 46, 147–160. doi: 10.1016/j.wocn.2014.07.001

Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *J. Acoustic. Soc. Am.* 97, 1286–1296. doi: 10.1121/1.412170

Pronk, T., Wiers, R. W., Molenkamp, B., and Murre, J. (2020). Mental chronometry in the pocket? Timing accuracy of web applications on touchscreen and keyboard devices. *Behav. Res. Methods* 52, 1371–1382. doi: 10.3758/s13428-019-01321-2

R Core Team (2023). *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Spencer, A. (1984). Eliminating the feature [lateral]. *J. Ling.* 20, 23–43. doi: 10.1017/S002222670000983X

Stefanich, S., and Cabrelli, J. (2021). The effects of L1 English constraints on the acquisition of the L2 Spanish alveopalatal nasal. *Front. Psychol.* 12:640354. doi: 10.3389/fpsyg.2021.640354

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phonetics* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001

Vaux, B. (1996). The status of ATR in feature geometry. *Ling. Inquiry* 27, 175–182.

Vaux, B. (1999). A note on pharyngeal features. *Harvard Working Papers Ling.* 7, 39–63.

Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percep. Psychophys.* 37, 35–44. doi: 10.3758/BF03207136

Werker, J. F., and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1866–1878. doi: 10.1121/1.390988

Wiese, R. (2000). *The Phonology of German*. Oxford: Oxford University Press.

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 23 April 2024
DOI 10.3389/flang.2024.1327600

Check for updates

OPEN ACCESS

EDITED BY
Baris Kabak,
Julius Maximilian University of Würzburg, Germany

REVIEWED BY
Kakeru Yazawa,
University of Tsukuba, Japan
Brandon Prickett,
University of Massachusetts Amherst, United States

*CORRESPONDENCE
Anne-Michelle Tessier
✉ amtessier@ubc.ca

RECEIVED 25 October 2023
ACCEPTED 07 March 2024
PUBLISHED 23 April 2024

CITATION
Zhang S and Tessier A-M (2024) Modeling the consequences of an L1 grammar for L2 production: simulations, variation, and predictions. *Front. Lang. Sci.* 3:1327600.
doi: 10.3389/flang.2024.1327600

Modeling the consequences of an L1 grammar for L2 production: simulations, variation, and predictions

Sijia Zhang and Anne-Michelle Tessier*

Department of Linguistics, University of British Columbia, Vancouver, BC, Canada

Introduction: This paper presents a constraint-based grammar of Mandarin low vowel + nasal coda (loVN) sequences first as acquired by L1 learners, and then as transferred to L2 English.

Methods: We simulate phonological learning in Harmonic Grammar using a gradual, error-driven GLA learner, drawing on evidence from L1 Mandarin speakers' perceptual data to support our initial state assumptions. We then compare our simulation results with L2 English production (both anecdotal and ultrasound data), as well as evidence from Mandarin loanword phonology.

Results: Our results align with multiple patterns in the previous empirical literature, including an asymmetry among surface repairs for VN sequences, and we show how these emerge from our assumptions about both the L1 Mandarin grammar and the grammar's evaluation method (i.e., weighted constraints).

Discussion: We discuss the extent to which these results derive from our somewhat novel analysis of place contrasts in L1 Mandarin, and the variability in loVN outputs that we encode directly into the L1 grammar, which are then transferred to the L2 context. Ultimately we discuss how this type of modeling can make falsifiable predictions about phonological development, in both L1 and L2 contexts.

KEYWORDS

phonological acquisition, L2 phonology, phonological variation, Mandarin, harmonic grammar, phonological learnability, gradual learning algorithm

1 Introduction

Although several decades of research has used computational simulations to investigate L1 phonological learning, comparing simulated data to observed stages of child speech (e.g., Levelt et al., 1999; Curtin and Zuraw, 2002; Hayes, 2004; Jarosz, 2010; Becker and Tessier, 2011; interalia), comparable work in the L2 acquisition literature has been rather more sparse. Often, such studies in the second language domain have focused on more phonetic questions of category learning, cue weighting and the like (starting especially with Escudero and Boersma, 2004), also including grammatical accounts of perceptual L2 learning (van Leussen and Escudero, 2015, and references therein). In the first decade after the advent of Optimality Theory, some crucial insights were investigated as to how gradual constraint re-ranking could capture stages of L2 learning and e.g., the emergence of the unmarked (e.g., Broselow et al., 1998). However, as phonological research has embraced various other brands of constraint-based grammars, such as weighted constraints in Harmonic Grammar (HG), Maximum Entropy grammars, and the like, there have been relatively few L2 acquisition analyses using these tools. One particularly interesting avenue for research, starting even with the Stochastic OT of Boersma (1998) and Boersma and Hayes (2001), is how an L1 grammar with multiple possible surface optima (i.e., grammatical variation) might be used to learn an L2 grammar built of the same constraints, and what consequences that inherent variation might bring about. Adding constraint weightings to the analytical mix also introduces the possibility that lower-weighted constraints might in the course of learning "gang up" in groups on higher-weighted constraints, providing different acquisition stages than would be predicted by a ranked constraint grammar.

This paper is an attempt to look at all of these questions, beginning with a modest but fairly detailed account of one phonotactic pattern: sequences of low vs. followed by nasal codas in Northern (Beijing-area) Mandarin, as compared with North American English. Starting with an analysis of the L1 Mandarin phonotactics and its inherent variability, we implement some simple computational learning simulations of the pattern, using a Harmonic Grammar of weighted constraints and a classic GLA learner. We then use the same simulated learner to implement a Full Transfer approach to L2 acquisition (see also Schwartz and Sprouse, 1996; as in van Leussen and Escudero, 2015). We can then compare how the L1 Mandarin transferred grammar treats the range of loVN sequences in L2 English with existing empirical data, especially from an L2 ultrasound production study (Liu, 2016) but also drawing on perceptual data and loanword phonology.

The paper is structured as follows. Section 2 compares the phonotactic restrictions of loV-N sequences in Mandarin vs. English and introduces a set of constraints to capture the inventories of both languages. Section 3 first presents how the learner can acquire an L1 Mandarin grammar of loV-N: §3.1 establishes some assumptions about ranking biases and the learner's input data over time, and §3.2 demonstrates via learning simulations in an HG-GLA model how the learner can reach the target L1 grammar successfully. We then turn to the L2 acquisition of English loV-N in Section 4, beginning with a clarification of the scope of our L2 study (§4.1). We present two key error patterns in the early stages of L2 simulations (§4.2), and show that these patterns are indeed found in previously reported literature and anecdotal reports (§4.3). After further discussion of the nature of L2 learning in our simulations (§4.4–4.5), Section 5 compares our analysis of L1 Mandarin with a standard alternative in the literature (§5.1), and discusses some of the crucial theoretical aspects of our approach (§5.2). We end with a short general discussion of L2 phonological learning and modeling.

2 The phonotactics of loV-N sequences, in Beijing Mandarin and English

2.1 The shape of Mandarin rhymes: low vowels and coda nasals

Both English and Beijing Mandarin have three nasal consonants /m, n, ŋ/ and two low vowels which contrast for [+/–back] in their surface inventories. Standard North American English uses [æ, ɑ][1] while Mandarin uses [a, ɑ] (Duanmu, 2007). However,

TABLE 1 Restricted inventory of Beijing Mandarin loV-N, compared to English.

	Nasal coda	/m/	/n/	/ŋ/
Low vowel	/–back/	*am	**an** *shan* [ʂan^{55}] "mountain"	*aŋ
	/+back/	*ɑm	*ɑn	**ɑŋ** *shang* [ʂɑN^{55}] "wound"

Beijing Mandarin (at least) has two crucial phonological restrictions on lowV-nasalC that do not apply in English: (1) only [n] and [ŋ] are allowed in coda position, but not [m] (Duanmu, 2007); (2) low vowels must agree for [+/–back] with a following coronal or dorsal nasal coda (Duanmu, 2007; Luo et al., 2020). Thus, English has 6 possible loV-N combinations while Mandarin has only two, as shown in Table 1.

In addition, the surface realizations of the legal Mandarin loV-N sequences [an] and [ɑŋ] may in fact be quite variable. Many studies demonstrate that coda nasals can be lenited (produced with no full oral closure) or deleted entirely, along with nasalization and possible compensatory lengthening of the preceding vowel (see Wang, 1993; Chen, 2000; Fang, 2004; Duanmu, 2007 inter alia; Luo et al., 2020). For example, Chen (2000) conducted acoustic measurements on the production of Mandarin coda nasals /n, ŋ/ preceded by high, mid, and low vowels (e.g., *shan-ao* [ʂan^{55}.ɑʊ^{51}]"hollow of the hill"). It was found that more than half of the word tokens containing the loV-N rhyme were produced without an oral closure for the nasal coda at normal speech rates, but that this oral closure deletion was less frequently observed when the nasal followed a high or mid vowel (Chen, 2000, p. 53, 54).

Thus, there are arguably three possible surface outputs for each input loV-N:

(a) faithful: the vowel is followed by nasal produced with full closure;
(b) lenited coda: the vowel is nasalized and the nasal coda is perhaps "weakened" but not fully deleted (here transcribed phonologically as [ãn] and [ɑ̃ŋ];
(c) deleted coda: the vowel is nasalized (and possibly lengthened) and the nasal is deleted (here transcribed as [ã] and [ɑ̃])[2].

The three variants of the output form are not contrastive; all that remains contrastive is the distinction between [-bk] and [+bk] sequences.

Table 2 compares these three surface variants in Mandarin with the corresponding range of surface possibilities in English. Unlike in Mandarin, a nasal coda in an English loV-N sequence is crucial to meaning, since its place is not determinable from the vowel – thus, there are four rows in the English section of Table 2 compared to two rows in the Mandarin section. With respect to the vowel, we will assume that the phonological target output for

1 This is a quite broad transcription, and abstracts away from a lot of phonetic detail between and across dialects, especially with regard to the location of [æ], e.g., its precise height and backness, as well as its status of nasalization or diphthongization. For our purposes, it is only crucial that it be a phonologically low front vowel.

2 We do not discuss the possible compensatory lengthening of the vowel when nasal coda is deleted in this paper, although it is reported in Duanmu (2007).

TABLE 2 Surface forms/variants of loV-N in Beijing Mandarin vs. English.

	Input	Possible outputs		
		(a) VN	(b) Ṽ N	(c) Ṽ
Beijing Mandarin	/{+lo, -bk}{+nasal}/: /an/, /aŋ/, /ãn/ or /ãŋ/	[an]	[ãn]	[ã]
	/{+lo, +bk}{+nasal}/: /ɑn/, /ɑŋ/, /ɑ̃n/ or /ɑ̃ŋ/	[ɑŋ]	[ɑ̃n]	[ɑ̃]
English	/{+lo, -bk}{+nasal, -bk}/ :/æn/ or /æ̃n/	[æn]	–	–
	/{+lo, +bk}{+nasal, -bk}/: /ɑn/ or /ɑ̃n/	[ɑn]	–	–
	/{+lo, -bk}{+nasal, +bk}/: /æŋ/ or /æ̃ŋ/	[æŋ]	–	–
	/{+lo, +bk}{+nasal, +bk}/: /ɑŋ/ or /ɑ̃ŋ/	[ɑŋ]	–	–

these sequences in English is a simple, oral vowel – that is, that the anticipatory vowel nasalization that appears before a nasal consonant is sufficiently partial, automatic and non-contrastive that it is not part of the phonological representation (e.g., Cohn, 1993; see also Beddor et al., 2013).

2.2 A constraint set to capture Mandarin and English

Here we lay out a constraint set that can minimally capture both the Mandarin and English inventories of lowV-nasal sequences, with reference to cross-linguistic typologies. All of these constraints are fairly standard from the previous literature, so they will be introduced fairly briefly. Note that all the tableaux that illustrate constraint definitions in this section are included in Appendix A.

The markedness constraints that we will use are in (1) and (2) below. First, there are the two constraints in (1) which are violated by those structures that are banned outright in Mandarin compared to English. Tableau A1 provides a few example outputs to illustrate how RHYME HARMONY and NOCODA [m] work.

(1) RHYME HARMONY — Assign a violation mark to every sequence of two segments, a low vowel + nasal consonant, where one is [+back] and the other [–back] [adapted from Duanmu (2007)]

NOCODA[m] — Assign a violation mark to every labial nasal consonant associated with the Coda position of a syllable

Then, there are the constraints violated by some subsets of the surface variants for loV-N in Mandarin – some of which are also ultimately relevant to the English inventory. Tableau A2 presents how the constraints in (2) work in getting different surface outputs in Mandarin.

(2) *VN — Assign a violation mark for every oral vowel followed by a nasal consonant

*NASALV — Assign a violation mark for every nasalized vowel

NOCODA — Assign a violation mark to every segment associated with the Coda position of a syllable

The key types of unfaithfulness that our grammars will need to consider are changes in place, changes in nasalization, and segmental deletion. Thus, we begin with the four Faithfulness constraints in (3), using the framework of McCarthy and Prince (1995). Note that in this system, changes in [+/–back] violate one of two Ident constraints, whereas [+nasal] is protected by a Max constraint. Tableau A3 provides some examples showing how these faithfulness constraints work.

(3) IDENT [+/–BACK] -V — Assign a violation mark for every pair of input and output vowels in correspondence which disagree in specification for [+/–back]

IDENT [+/–BACK] -N — Assign a violation mark for every pair of input and output nasal consonants in correspondence which disagree in specification for [+/–back]

MAX — Assign a violation mark for every input segment without an output correspondent

MAX[NASAL] — Assign a violation mark for every input [nasal] feature without an output correspondent

As can be seen by comparing the third to fifth candidates in Tableau A3, we are interpreting the primitive [nasal] feature on both consonants and vowels – even adjacent ones – to be separate features, each protected by faithfulness. Thus, we assume that deleting a nasal consonant incurs a violation of MAX[NASAL] even if the preceding vowel is already nasalized. In Tableau A4, we see that an underlying nasal consonant which is deleted but triggers nasalization on the preceding vowel does *not* violate MAX[NASAL]; the underlying and surface forms both have one instantiation of [nasal], in correspondence with each other directly (just not associated with segments that are in correspondence)[3].

In addition, we include two DEP constraints into our constraint set: one that penalizes segmental insertion, and one that specifically penalizes inserting a [nasal] feature onto a vowel, as in (4). Epenthesis of either nasalization onto a vowel or a full nasal consonant after a nasal vowel can both be compelled by one or more

3 This approach to the representations of adjacent features is of course not the only or even more the common one, but it will allow us to make clear how constraints interact in the languages being learned.

of the markedness co-occurrence constraints above in Tableau A1. Tableau A5 demonstrates how the two DEP constraints work.

(4) DEP Assign a violation mark for every output segment without an input correspondent

DEP[NASAL] Assign a violation mark for every output [nasal] feature associated with a [+vocalic] segment without an input segment associated with an input [nasal] feature

3 The L1 grammar of Mandarin VN sequences

3.1 The initial state

3.1.1 Constraint weightings

Our learner's initial state consists of two weighting biases. The first is to weight all markedness constraints above faithfulness constraints, as is well established in decades of literature. This general bias captures both the fact that children's production grammars begin with highly unmarked outputs on the whole, and that any alternative starting point is more prone to subset/superset traps, in the sense of Angluin (1980) and Berwick (1985). In OT, discussion of this general bias begins with Smolensky (1996), see also extensive discussion in e.g., Gnanadesikan (2004), Hayes (2004), Tessier (2016).

The second bias deals with the relative ranking of faithfulness constraints, and adopts Steriade (2001)'s proposal of a *P-Map*, whereby faithfulness constraints that militate against more perceptually salient changes are higher ranked (or weighted) than those which ban less salient changes. In particular, we include a bias for weighting Ident-Vplace above Ident-Nplace, reflecting the result that changes to vowel place of articulation are relatively more salient. The most immediately relevant such results come from Zhang (2023), in which native speakers of Mandarin gave similarity ratings of loV-N pairs, and they perceived [æn]/[æŋ] as more similar than [æn]/[ɑn] or [æŋ]/[ɑŋ]. We discuss this study further in Section 4.5.

We combine these two biases into the initial state of a Harmonic Grammar below. The first – M>>F – is simply a starting point, which evidence can easily overturn. The second, however, is taken to be as universal a fixed weighting as possible, and therefore will be implemented in our simulations so as to persist as much as possible, regardless of errors and learning data.

3.1.2 Data in two stages

In Section 2.1, Table 2 provided six possible Mandarin outputs. In the earliest stage of phonotactic learning, without any firm knowledge of meanings associated with these surface forms, the learner's assumption is that all six such outputs are faithfully mapped. This "Identity Map" or purely phonotactic learning grammar, will be our first simulated learning task. Later, the L1 Mandarin learner must determine – via semantic word learning and

TABLE 3 (A) Stage 1: Purely phonotactic learning and (B) Stage 2: Revised learning.

Input	Output
(A)	
/an/	[an]
/ãn/	[ãn]
/ã/	[ã]
/ɑŋ/	[ɑŋ]
/ɑ̃ŋ/	[ɑ̃ŋ]
/ɑ̃/	[ɑ̃]
(B)	
/an/	[an]
	[ãn]
	[ã]
/ɑŋ/	[ɑŋ]
	[ɑ̃ŋ]
	[ɑ̃]

associated reasoning[4] – that not all of these strings are uniquely mapped, and that in fact the grammar must instead produce surface allophonic variation from input loV-N sequences. In our Mandarin learning simulations, we model this two-stage process by first letting the learner reach a stable grammar that produces the mappings in Table 3A, then feeding it the input-output mappings in Table 3B and learning again.

3.2 An HG-GLA learner

3.2.1 Weighted constraints in the GLA

Our phonotactic grammar is formalized as a weighted Harmonic Grammar, in which the candidate with the highest harmony value is the optimum (Legendre et al., 1990; Smolensky and Géraldine, 2006; Potts et al., 2010). A violation score is assigned to each constraint, which is a negative number corresponding to the number of violations of that constraint. The harmony value of a candidate is calculated as multiplying each constraint's violation score by its weight, and then summing up. In addition, a small amount of noise drawn from a Gaussian distribution ($SD = 2.0$) is added to the constraint weights on each iteration of Eval.

The learner we adopt is error-driven and gradual, using the HG version of the Gradual Learning Algorithm (Boersma and Hayes, 2001). On each trial, the HG-GLA learner feeds an input to the current grammar and maps it to its currently-optimal output candidate. If that optimal candidate is the target grammar's intended *winner* for that input, no learning occurs. If that optimum is not identical to the intended winner, however, then that *loser* form is used to create an error. Recalling that at first, our learner

4 For relevant work about this reasoning, see e.g., Pater et al. (2012), O'Hara (2017), and Nelson (2019); thanks to an anonymous reviewer for pointing these out.

assumes that the observed surface winners are identical to their input forms, some sample errors in the L1 Mandarin purely phonotactic learning stage are below shown in Table 4.

When an error is made (i.e., there exists a mismatch between the *winner* target output form and the *loser* output form chosen by the current grammar), the learner increases the weight of the constraints that prefer the winner, and decreases the weight of constraints that prefer the loser. For instance, given the error that occurs in Table 4(i), the learner will make an update by decreasing the weight of *VN and increasing the weights of MAX and MAX[NASAL]. The HG-GLA learner adjusts the weights on each learning trial until the target form matches with the optimum produced by the current grammar of the learner. Learning occurs only in this type of mismatch scenario based on positive evidence – that is to say, it only learns when it has been prompted by an observed target form (e.g., Hayes, 2004; Prince and Tesar, 2004). In the constraint-based literature on phonological learning, this approach is in contrast to learners which use Bayesian-style reasoning to consider *unobserved* surface forms and decrease their predicted likelihood in the grammar (as in e.g., Jarosz, 2006; Hayes and Wilson, 2008 and many others).

3.2.2 Basic learning parameters and biases

The learning simulations were implemented in Praat (Boersma and Weenink, 2016). In order to impose the bias Markedness >> Faithfulness in a Harmonic Grammar, we initialized the weight of all the markedness constraints at 100 with the plasticity of 1, including *VN, *NASALV, NOCODA, NOCODA[M], and RHYMEHARMONY. All faithfulness constraints but one (see next paragraph), were initialized with a weight of 1 and plasticity of 1, including IDENT[BK]-N, MAX, MAX[NASAL], DEP and DEP[NASAL].

We implemented an initial bias between IDENT[BK]-V and IDENT[BK]-N such that IDENT[BK]-V had a weight 10 higher than IDENT[BK]-N (IDENT[BK]-V = 11). In order to retain the relative weighting between the two Ident constraints, we set the plasticity of IDENT[BK]-N at 0.1, a much smaller value than that of IDENT[BK]-V and all other constraints (= 1). The difference in the plasticity between the two IDENT constraints allows IDENT[BK]-N move at a slower rate compared to IDENT[BK]-V, so that the bias IDENT[BK]-V >> IDENT[BK]-N can be persistently imposed throughout learning (see esp. Jesney and Tessier, 2011).

We used the update rule *symmetric all*, which is defined such that the weight of all constraints that are violated more in the target output than in the learner's output is lowered, and the weight of all constraints that are violated more in the learner's output than in the target output is raised (Boersma and Hayes, 2001). The learning proceeded at a constant plasticity at 1 (number of plasticities = 1, initial plasticity = 1), with an evaluation noise set at 2^5. The learning strategy was set to LinearOT (Keller, 2006), so that no constraint weights could drop below zero on evaluation, and any negative one-time disharmonies were treated as zero.

3.2.3 Simulating the purely phonotactic stage

Recall from Section 3.1.2 that at Stage 1, the learning data is fully-faithful – that is, all three surface variants of the loV-N outputs [an, ãn, ã] or [ɑŋ, ɑ̃ŋ, ɑ̃] are assumed to come from identical corresponding inputs. On each learning trial, the learner is fed one such mapping from the six possible pairs (e.g., /ãn/ → [ãn]). After 10000 such learning trials, the end state grammar typically look like the example in (5) below, which shows one set of precise constraint weights from a simulation. This grammar produces all six of the required fully-faithful mappings:

(5) RH, NOCODA[M] >>	MAX[NASAL] >>	*VN >>
100, 100	**82.6**	**72**
*NASALV >>	MAX, NOCODA >>	
62	**52, 48**	
IDENT[BK]-V >>	DEP, DEP[NASAL],	IDENT[BK]-N
11	**2, 1,**	**1**

Compared to the initial stage, the weights of RHYMEHARMONY (=100) and NOCODA[M] (=100) in (5) have not changed; no L1 Mandarin surface form violates either constraint, so there is no pressure for demotion or promotion. The two IDENT[BK] constraints IDENT[BK]-V (=11) and IDENT[BK]-N (=1) have also not changed, since there is no markedness pressure that could be better satisfied in any of the input learning data by changing [+/–back]. Thus, we remove these four constraints temporarily from our analysis, and explain in the remainder of this section how the remaining constraint weights are updated in this first stage of learning. Together, these constraints determine the optimal candidate among the four possible outputs – [an], [ã], [ãn], and [a] – given one of the three inputs (Here we illustrate for the [–back] pairs, but all of the below also applies to the [+back] pairs). Note that the tableaux that illustrate these crucial weightings are all included in Appendix B.

Tableau B1 shows how the /ã/ input is mapped faithfully; the winning candidate violates only *NASALV. In (6), we provide the two weighting conditions that explain how the winning candidate's violation of *NASALV can be optimal – the third candidate in Tableau B1 is harmonically bounded (see shading on candidate iii), so *any* weighting conditions will rule it out.

(6) For /ã/ to map faithfully to [ã]:
- w(*VN + NOCODA) > w(*NASALV) ruling out [an][6]
- **w(MAX[NASAL]) > w(*NASALV) ruling out [a]**

The second of these weighting conditions is in bold because it represents a simple trading relation between two constraints (akin to a ranking argument in classic OT).

The next input /an/'s fully-faithful output violates two markedness constraints: *VN, and NOCODA, shown in Tableau B2. For the faithful candidate to win, the weights of these two constraints must sum to less than the weighted violations of the other options, and these conditions are listed in (7).

5 For an argument that this noise should be assessed *after* evaluation, i.e. on the probabilities of candidates themselves as in MaxEnt grammars, rather than at the time of evaluation, see Kawahara (2020) and Hayes (2022). In our simulations, this distinction is not (to our knowledge) crucial.

TABLE 4 Sample errors in L1 Mandarin purely phonotactic learning.

/input/ (here: /winner/)	outputs: [winner] vs. [loser]	Rhyme Harmony	*NasalV	*VN	Max	Max [nasal]	No Coda
(i) /an/	an vs. a			L	W	W	L
(ii) /ãŋ/	ãŋ vs. ɑŋ		L	W		W	
(iii) /ã/	ã vs. ɑ		L			W	

(7) For /an/ to map faithfully to [an]:

- w(*NASALV + MAX)	> w(*VN + NOCODA)	ruling out [ã]
- w(*NASALV)	**> w(*VN)**	**ruling out [ãn]**[7]
- w(MAX[NASAL] + MAX)	> w(*VN + NOCODA)	ruling out [a]

Again, the bolded weighting condition is a straight competition between two constraints.

Finally, we consider the fully-faithful mapping of /ãn/, which violates both *NASALV and NOCODA [see Tableau B3 and (8) below].

(8) For /ãn/ to map faithfully to [ãn]:

- w(MAX[NASAL] + MAX)	> w(NOCODA)	ruling out [ã]
*- w(MAX[NASAL] + *VN)*	*> w(*NASALV)*	*ruling out* [an]
- w(2* MAX[NASAL] + MAX)	> w(*NASALV + NOCODA)	ruling out [a]

Note that one of these weighting conditions (shown in italics) is a specific case of a more general weighting condition already established [see (6)'s second line].

To summarize this section, this fully-faithful (or purely phonotactic) grammar is one in which deleting nasalization is not allowed, regardless of whether it is in a marked context or not, and in which vowel nasalization is also not added just to avoid oralV-nasalC sequences. These two statements summarize the two simple weightings above: w(MAX[NASAL]) > w(*NASALV) > w(*VN). Somewhat lower weighted but still relevant are MAX and NOCODA, in that order, so that segmental deletion to avoid codas is not possible – and also, as the more complex weightings above show, that deletion is in fact never optimal, due to the constellation of higher-weighted M and F constraints. At the bottom of this grammar are DEP and DEP[NASAL], which will play more of a role later.

It is important to notice that this grammar, while fully-faithful to all the inputs it has seen, does *not* accomplish this pattern by simply ranking all F >> M. The majority of the faith constraints are weighted at the bottom – only MAX[NASAL] or MAX have risen significantly above their initial state values. Because the learner is biased to start with M constraints weighted high, any errors which can be attributed to the competition between markedness pressures will result in the reordering of those constraints as well as the promotion of F. Thus, this learner's initial biases plus this particular set of somewhat antagonistic markedness constraints results in a grammar which maps these inputs faithfully.

3.2.4 Simulating the revised learning stage, and the role of variation

The second stage of Mandarin learning in our simulation occurs once the learner has discovered that the six surface outputs learned in stage 1 correspond to only two underlying contrasts. As discussed in Section 3.1.2, this learner now knows that output variants [an, ãn, ã] are all derived from a single [−back] input, and [ɑŋ, ãŋ, ã] are all surface realizations of a [+back] input. In the learning data given in 3.1.2, the underlying forms are /an/ and /ɑŋ/.

The Stage 2 learner begins with the end-state grammar from stage 1 from 3.2.3 above, and is now given its new, frequently unfaithful mappings as learning data, with each of the three variants given equal probability (0.33) as the correct output. Here too our learner is successful – after 10,000 trials, the learner is able to generate the three output variants with the frequency shown in the input-output pair distribution plot (Figure 1). One sample of such a grammar is given in Table 5; which illustrates how the weightings change between the two stages.

From the comparison above, it is fairly clear that the change at Stage 2 has come from evening out the weightings of pairs of markedness constraints – *VN and *NASALV are now very similar in weight, as are MAX and NOCODA – and most other weights have remained the same. This is shown in the two comparison tableaux B4 and B5. To choose the unfaithful candidate in Tableau B4 we need:

w(*VN + NOCODA) > w(*NASALV + MAX)

Comparing the Stage 2 grammar in Table 5 with this weighting condition, we see that in each bracket there is one constraint weighted around 67-68, and another around 50. Similarly, to choose the unfaithful candidate in Tableau B5 we need:

w(*VN) > w(*NASALV + DEP[NASAL])

Since DEP[NASAL] is still very low-weighted, this again makes the choice very variable.

Changing the number of learning trials (10000 ± 1000) results in slightly different proportions of the output variants (e.g., [an] or [ãn] as the most frequent output form rather than [ã] as shown in

6 Technically this should be w(*VN + NoCoda + Dep) > w(*NasalV), but Dep has little to do here because of its low weight.

7 Again, the second term should be w(*NasalV + Dep[nasal]), but Dep[nasal] has little to do here.

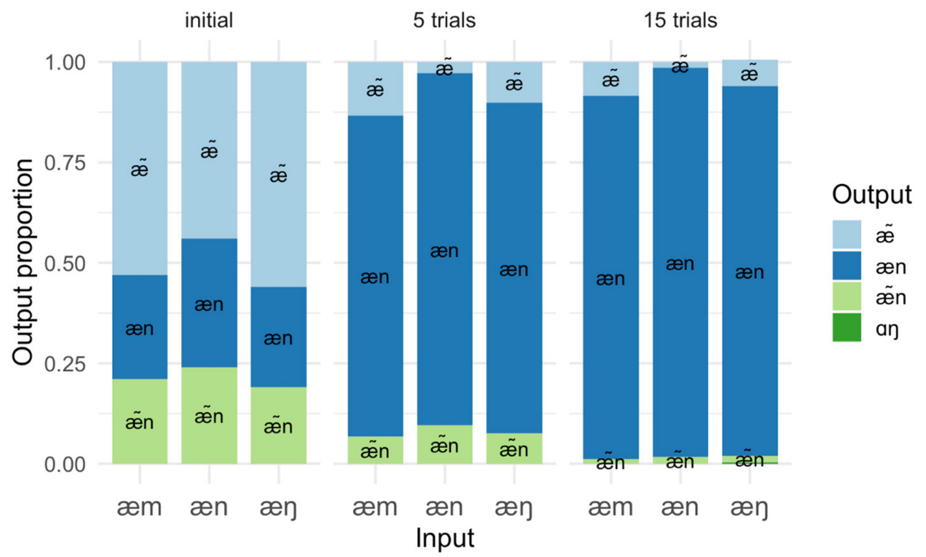


FIGURE 1
Output proportions generated in L1 Mandarin revised learning (Stage 2).

Figure 1) – but the overall results is roughly equivalent variation, which we will take to be a successful Mandarin end-state grammar.

To re-iterate and also foreshadow, this end state creates variability through similar weightings of a set of antagonistic markedness constraints, such as *VN and *NasalV, which make opposite demands when a vowel is followed by a nasal. In the next section, we will see how this variability impacts L2 acquisition, when additional loVN sequences are introduced.

4 The L2 acquisition of English loV-N sequences by L1 Mandarin speakers

4.1 Our goals in modeling L2 acquisition

At the outset, we wish to clarify the general purpose of our modeling L2 development and its context. Our goal is *not* to make specific predictions about individual grammars of L2 English learners, such as absolute rates of acquisition, ultimate attainment, and the like. As a reviewer rightly points out, if the HG-GLA learner we adopt is given the right constraint set and an informative set of input/output mappings to learn from, it will eventually learn a correct end state grammar, mimicking "perfect" L2 acquisition. The speed with which this is achieved will be a consequence of the plasticity parameter settings, as well as the relative frequency of different mappings the learner is fed, and we do not have any insights as to these aspects of development.

Instead our goal is to spell out the consequences of our assumptions about the L1 grammar and learning biases, as they make predictions about L2 learning. In particular, we will demonstrate that our proposals from Section 3 of how to capture inherent variation in L1 Mandarin loV-N sequences make two clear predictions when that grammar is applied to an L2 like English, and then discuss the extent to which these predictions align with observed data. For the sake of completeness, a later Section

TABLE 5 Sample end-state grammar of Stage 1 and Stage 2 learning.

	(a) End state Stage 1 grammar	(b) End state Stage 2 grammar
RhymeHarmony, NoCoda[m]	100	100
Max[nasal]	82	83.35
*VN	72	68.20
*NasalV	62	67.20
Max	52	50.58
NoCoda	48	50.41
Ident[bk]-V	11	11
Dep[nasal]	2	0.97
Dep	1	1
Ident[bk]-N	1	1

4.4 demonstrates that this simulated learner can indeed reach a target-like English end-state – but we do not intend to imply that any or all L2 learners of English from L1 Mandarin backgrounds necessarily acquire grammars that are identical to L1 English ones (see more in Section 4.4 below).

As mentioned in the introduction, we adopt a Full Copying model of L2 acquisition (also following Schwartz and Sprouse, 1996; in this context, see especially Escudero, 2005; van Leussen and Escudero, 2015). This means that the initial state of our L2 grammar is precisely the end state of the L1 grammar. We note, however, that we focus only on the acquisition of a *production* grammar; while van Leussen and Escudero (2015)'s model (the L2LP model revised) concerns the full copying of an L1 *perception* grammar to apply initially in perceiving L2 surface forms (see Escudero, 2005's Optimal Perception Hypothesis; see also Boersma, 2011 and other work on bidirectional learning of perception and production). In this work, we assume that the learner has progressed to the point of relatively accurate input perception at the point where we begin our L2 production learning scenarios. We return to our learner's perception/production assumptions in Section 4.5.

4.2 Two simulation predictions: transferring L1 Mandarin grammar to L2 English

The target English grammar – the idealized goal end state of L2 learning – maps all of the six output loV-N sequences as faithful to their input forms (Recall from Section 2 that we assume that no vowel nasalization at this phonological level of English – that is, whatever degree of anticipatory nasalization is produced on these vowels, it is consistent and non-contrastive, so it does not form part of these input/output mappings). So with this fully-faithful English grammar as its target, what does a Full Copying L1 Mandarin learner look like? There are two crucial properties of this L2 learner revealed by our simulations that we will focus on.

Table 6a shows the winning output candidates produced by the L1 end-state grammar described in the previous section (i.e., Table 5b), now given English inputs. From now on, we restrict ourselves to the [–back] vowels and their possible loV-N outputs – since nothing in our grammar distinguishes between front and back vowels, beyond whether or not they harmonize. Thus, the input form in bold /æn/ represents those which are also found in the L1 Mandarin input lexicon (i.e., it will also describe the treatment of /ɑŋ/ *ceteris paribus*), and in the initial L2 English grammar these existing input forms are mapped to the same three output options in the L1 system. The other four inputs in Table 6 are novel, and they all raise potential violations of the two undominated markedness constraints in the L1 Mandarin grammar: the first input violates NoCoda[m], /æm/, and the bottom row input contains place mismatch that violates RhymeHarmony, /æŋ/.

The first result of the simulation that we highlight is that all of the output candidates for the two types of novel English inputs are in some way unfaithful to the input nasal – that is, either it is deleted, or it is unfaithful to nasal place, while vowel place is kept faithful. Two such options are illustrated in the tableaux of Tables 7, 8. The reason that this initial state grammar satisfies Rhyme Harmony by repairing nasal place, rather than vowel place, is our built-in bias for Ident[Bk]-V to be weighted above, and with greater plasticity than, Ident[Bk]-N. In other words, this learner prefers to be unfaithful to nasal place rather than vowel place without evidence, despite their L1 experience having provided no overt alternations toward this choice of repair.

The second simulation result that we want to examine is a skew in the relative frequency of nasal deletion vs. the other two output options, when comparing the L1 legal input vs. the novel ones. To see this in Table 6, we have bolded the proportion of deletion candidates for each input. Here just focusing on the initial L2 grammar, in the first (6a) column, we can see that the /æn/ input that does not violate RhymeHarmony surfaces with deletion 44% of the time, whereas the other two inputs generate deletion candidates 53 and 56% of the time.

In the tableaus of Table 9, we see how this asymmetry between deletion and non-deletion candidates comes about. The Tableau in 9 (a) compares the winners [æ̃] and [æn] given an input which does *not* violate Rhyme Harmony, /æn/. Here the choice between these two possible outputs comes down to a fight bettesween similarly-weighted constraints, such as Max and NoCoda. In comparison, Tableau 9 (b) compares the same two possible output types, but given an input like /æŋ/ that violates undominated RhymeHarmony. Since the fully-faithful option is ruled out (iii), the resulting competition between (i) and (ii) involves the same closely-weighted constraint set, but *also* the crucial violation of Ident[Bk]-N.

The result will be that whatever the distribution of probabilities for the two potential outputs (i) and (ii) in (9a), the probability of deletion will be slightly higher in (9b). Specifically, deletion in (9bii) will be relatively more harmonic than (9aii), by virtue of the additional violation of Ident[Bk]-N accrued by (9bi). Given the precise weightings of our L1 end-state grammar – where Ident[Bk]-N is weighted just at 1 – this adds a few percentage points in favor of the deletion candidate in Tableau 9 (bi), producing the skew in Table 6a's italicized proportions. We emphasize that this result at the beginning of L2 production comes in large part from the L1 already being variable in its outputs for loVN sequences, so that these multiple output options transfer even to the L2 novel inputs which are L1-illegal. Since L2 learning is initialized as varying between possible output candidates, we can observe this skew toward deletion over nasal place substitution for mismatched VN.

4.2.1 Simulating L2 development beyond the initial state

Using the learning parameters described in Section 3 above, the L1 Mandarin initial state grammar can quickly be re-arranged to replicate some aspects of the L2 English system. Table 10 illustrates some typical constraint weightings during these learning stages: moving on from the initial state (a) to its next stages after a few learning trials (given our previous parameters, after five learning cycles in Table 10b and 15 learning cycles in Table 10c). Those constraints which are been overall promoted or demoted are indicated with an up or down arrow. While only small changes in constraint values have occurred after so few learning trials, the overall effect on output probability distributions is significant, as the later columns of Table 6 demonstrates.

The first overall change in distributions shown in Table 6 (b, c) is a sharp proportional increase in output candidates with the goal output *shape,* namely oral vowel followed by nasal consonant. After only 15 trials, this output shape accounts for more than 90% of each of the three output types above. It is not surprising that the first aspect of the L1 grammar to be overturned in L2 learning is precisely the delicate balance between three potential output winners – a very small nudge to those constraints (e.g., Max up and NoCoda down) is enough to focus the grammar on one of these three variants.

With respect to second feature of our simulation, the skew in deletion rates depending on input nasal place, we continue to see the preference for deletion when the input contains a [+/–back]

TABLE 6 Output proportions generated from L2 learning at the initial state (L1 end-state grammar), and after 5 and 15 trials.

V[-bk] target	Output candidate	(a) % at initial state	(b) % after 5 trials	(c) % after 15 trials
/æm/	[æ̃]	**53**	**13.3**	**8.4**
	[æ̃n]	21	6.8	1.2
	[æn]	26	79.9	90.4
/æn/	[æ̃]	**44**	**2.8**	**1.4**
	[æ̃n]	24	9.5	1.7
	[æn]	32	87.7	96.9
/æŋ/	[æ̃]	**56**	**10.2**	**6.5**
	[æ̃n]	19	7.6	1.7
	[æn]	25	82.2	92
	[ɑŋ]	0	0	0.3

TABLE 7 Errors for novel English input/æm/at L2 initial state.

/æm/	NoCoda [m] 100	*VN 68.2	*NasalV 67.2	NoCoda 50.4	Max 50.6	ID[bk]-N 1	*Harmony*
(i) [æm]	*	*		*			−218.6
☞ (ii) [æ̃]			*		*		−117.8
(iii) [æn]		*		*		*	−119.6

mismatch at all three stages – the input /æn/ has a lower percentage of [æ̃] outputs than the other two inputs, and this is also illustrated graphically in Figure 1.

4.3 Comparing the simulation's predictions with L1 Mandarin L2 English data

To what extent do these two simulation properties align with the L2 acquisition of English by Mandarin speakers? In support of the first simulation result, it is certainly reported anecdotally that nasal place and not vowel place is the feature that surfaces unfaithfully in L2 English errors by L1 Mandarin speakers. For example, the first author (a lifelong speaker of Beijing Mandarin) reports that English words such as *bang* and *gang* are frequently produced as [bæn] and [gæn] by L1 Mandarin speakers (or as [bæ̃] and [gæ̃]), and similarly that *gone* surfaces as [gɑŋ] (or as [gɑ̃]) – all with nasal coda place replaced to match vowel backness.

Another piece of data comes from the systematic adaptation patterns of English loanwords with mismatches in vowel and nasal place into Mandarin (Hsieh et al., 2009 and references therein). Consistent with the above anecdotal reports, loanword adaptations also suggest that the L1 Mandarin grammar, when faced with a [+/− back] loVN mismatch, will alter the place of the nasal rather than the vowel. Thus, Hsieh et al. (2009) report borrowings such as *tango* [æŋ] → *tan.ge* [an] and *Wisconsin* [ɑn] → *wei.si.kang.xing* [ɑŋ]. Similarly, when English words with coda [m] are borrowed into Mandarin, [m] tends to be replaced by [n] or [ŋ] to match the [+/− back] of the low vowel, e.g., *Nottingham*[æm] -> *nuo.ding.han*[an]. Of course, we acknowledge the long-standing debate as to the extent to which a language's loanword phonology is equivalent to or divergent from its L1 grammar, which we have assumed here as the initial L2 state (for example, compare Yip, 1993; Paradis and LaCharité, 1997; Peperkamp et al., 2008). In the present case, we point to recent work by Huang and Lin (2023), which presents evidence that L1 Mandarin learners of L2 English produce English nonce words and adapt English loanwords equivalently (and see also Broselow, 2023 on the shared uses of probabilistic cue weighting in loanword and native grammars.) We therefore take this evidence as at least suggestive of convergence between Mandarin loanword repairs and the L1 grammar, which here is transferred to an L2 scenario.

With respect to the second simulation result, here we turn to an articulatory result from Liu (2016)'s ultrasound study of L2 English words produced by adult L1 Beijing Mandarin speakers, comparing vowel + oral stop coda /d, g/ (e.g., *bad*) and vowel + nasal stop coda /n, ŋ/ (e.g., *ban*). Speakers were asked to read a list of monosyllabic words containing the target sequence in a carrier sentence while their lingual gesture and movement were imaged.

To analyze the production of nasal coda closure, the shortest distance between the tongue contour and the region of interest along the palate (either alveolar or velar region) was calculated for each word. This shortest distance (also referred to as the smallest *aperture*) was normalized across speakers, by taking into account the full range of tongue positions that an individual speaker can achieve when producing any coda (oral or nasal). Specifically, *percent lingual aperture* refers to the smallest aperture produced during a particular coda, compared to the smallest aperture measurement overall. The higher the percent lingual aperature, the greater the distance between the tongue contour and the corresponding palate region (alveolar or velar region), and hence more coda gestural reduction. The key result in Liu (2016)'s data is that, for loV-N sequences, the nasal coda was produced with

TABLE 8 Errors for novel English input/ɑn/at L2 initial state.

/ɑn/	Rhyme Harm 100	*VN 68.2	*NasV 69.2	No Coda 50.4	Max 50.9	Dep Nasal 0.97	ID[bk]N 1	Harmony
(i) [ɑn]	*	*		*				−218.6
☞(ii) [ɑŋ]		*		*			*	−119.6
(iii) [ɑ̃ŋ]			*	*		*	*	−121.6
(iv) [ɑ̃]			*		*			−120.1

TABLE 9 The asymmetry in generating the deletion candidate comparing input/æn/and/æŋ/.

	Rhyme Harm	Max	No Coda	ID[bk]-N	Harmony
(a) /æn/					
(i) [æ̃]		*			*w*(Max) *−1
(ii) [æn]			*		*w*(NoCoda) *−1
(b) /æŋ/					
(i) [æ̃]		*			*w*(Max) *−1
(ii) [æn]			*	*	[*w*(NoCoda) *−1] + [*w*(Id-PlaceN) *−1]
(iii) [æŋ]	*!		*		[*w*(RhymeHarm) *−1] + [*w*(NoCoda) *−1]

a more reduced gesture (i.e., higher percent lingual aperture) when the low vowel and the nasal had mismatched backness compared to that with matched backness. In other words, the closure of [ŋ] was more reduced than [n] following [æ], and [n] is more reduced than [ŋ] following the back vowel [ɑ] (Liu, 2016, p. 29, Figure 4.2).

Our interpretation of Liu (2016)'s data is that these Mandarin-speakers L2 developing grammar produced relatively more deletion for the mismatched loV-N sequences than the matched ones. This is of course something of a translation, from one type of behavioral data to a grammatical abstraction across symbolic variants. One way to understand this might be that each phonological output candidate can be implemented with a range of coda closure gestures (and associated percent lingual apertures), and that implementing a "deletion" candidate more often for a particular type of input will result in an average higher percent aperture. In the terms of our simulation, their production grammar mapped the input /æŋ/ relatively more often to [æ̃] the deletion candidate than they did with /æn/, and similarly /ɑŋ/ generated more [ɑ̃] outputs than /ɑn/. And this is indeed what our simulated learner does (recall Table 6).

4.4 Further L2 learning simulations, and the potential end state

Over time, the L2 learner will accrue evidence that supports faithfulness to the two types of ungrammatical Mandarin surface forms in their new lexicon: coda [m]s, and backness mismatches. In the six forms given to the learner at equal proportions, two inputs violate NoCoda[m] and two others violate RhymeHarmony, so it is roughly the case that these two markedness constraints are demoted at the same speed. The last two columns of Table 10 show this, comparing typical values for the L2 grammar at 15 and 1,000 trials, with a visual re-organization that reflects the main constraint re-orderings.

The final state grammar after 1,000 learning trials encodes all the relevant English constraint weightings, and thus is faithful to all six of the English inputs. To illustrate, the tableaux in Table 11 compare the same two mappings from the L2 initial state of Tables 7, 8 above – here we see that violations of NoCoda[m] (11a) and of RhymeHarmony (11b) are now both tolerated.

As mentioned in Section 4.1, just because the simulation can reach "perfect" L2 acquisition does not mean that any human learner can or will; we acknowledge that much more must be considered to make predictions about ultimate attainment. As a sidenote from our main arguments, we note that one approach to capturing incomplete L2 acquisition in this kind of grammatical simulation might be to impose limits on the frequency with which a learner uses errors to update their current constraint weights, and/or the number of errors that the learner is willing to process before it "fossilizes" at some particular state. For a learning approach that bears some resemblance to this view, though in an L1 context, see Tessier (2016). Another relevant simulated learning approach is found in Zhao and Li (2022), in which an unsupervised, connectionist model is given training lexicons of varying sizes, to observe something akin to progressives stages of learning and the different types and frequencies of errors that each resulting learner makes. In this work, too, the goal is to compare model-predicted errors with production data (from learner corpora), but not to determine how and when an L2 target-like end-state is achieved.

4.5 Comparing L2 production and perception

As pointed out in Section 4.1, the initial L2 learning state in the grammar that we model here is not the end state of an L1 production grammar. Our assumption is that by this point, our learner's L2 *perception* is fairly accurate. To

TABLE 10 Sample constraint weightings at different L2 learning states.

An L2 grammar...	(a) *At the initial state*	(b) *After 5 learning cycles*	(c) *After 15 cycles*	(d) *After 1000 cycles*	
RhymeHarmony	100	98.1	96.9		
NoCoda[m]	100	99.1	96.1		
MaxNasal	83.4	83.4	83.3	MaxNasal	83.5
*VN ↓	68.2	66.4	65.6	*NasalV	75.0
*NasalV ↑	67.2	69.0	69.9	*VN	60.4
Max ↑	50.6	52.4	52.3	Max	57.4
NoCoda ↓	50.4	48.6	48.7	NoCoda	43.6
Ident[bk]-V	11	11	11	Ident[bk]-V	19
Dep	1	1	1		
DepNasal ↑	0.97	0.97	1.9		
Ident[bk]-N ↑	1	1.1	1.9	Ident[bk]-N	17.5
				RhymeHarmony	10.1
				NoCoda[m]	10.2
				Dep	1
				DepNasal	1.9

TABLE 11 L2 end-state grammar given novel English inputs (compared to Tables 7, 8).

	*NasaV 75	*VN 60.4	Max 57.4	NoCoda 43.6	ID[bk]-N 17.5	NoCoda [m] 10.2	Rhyme Harm 10.1	H
(a) /æm/								
☞ (i) [æm]		*		*		*		−114.2
(ii) [æ̃]	*		*					−132.4
(iii) [æn]		*		*	*			−121.5
(b) /ɑn/								
☞ (i) [ɑn]		*		*			*	−114.1
(ii) [ɑŋ]		*		*	*			−121.5
(iii) [ɑ̃ŋ]	*			*	*			−136.1
(iv) [ɑ̃]	*		*					−132.4

support this claim, we return to the perceptual evidence from Zhang (2023). In Section 3.1.1, we reported the difference in participants' accuracy in perceiving L2 English place contrasts in nasals vs. vowels. Nevertheless, though nasal contrasts were less well discriminated, her participants still had high accuracy overall in perceiving both of these contrasts – even between those English loVN sequences which are illegal in their L1. Even for the least salient contrast, the AX discrimination task, L2 listeners had an average accuracy over 93.46% in discriminating loVN sequences that differ only in the nasal (e.g., [æn]-[æŋ]). This degree of accuracy was not significantly different from the average accuracy rate of 96.24% for L1 English listeners.

However, since Zhang (2023) did not test production data along with perception, we should ask how reasonable it is to aim our simulations at the behavior of an L2 English learner whose L2 perceptual abilities are quite advanced but whose L2 production remains highly L1-influenced? In fact, we see fairly good support for this view in the two groups of experimental participants whose data we have extracted, comparing Zhang (2023) and Liu (2016). Both of these L2 groups are L1 Beijing Mandarin-speaking educated adults, who at the time of the study were living and studying in an English-speaking country or region [Hong Kong for those speakers in Liu (2016), and Vancouver, Canada for Zhang (2023)]. Most had moved to an English-speaking environment after the age of 18, before which they were exposed to English in classroom settings, and they had all achieved sufficient English proficiency to gain acceptance to an English-speaking university (i.e., TOFEL or IELTS). Overall, we believe that the amount of L2 knowledge and experience was relatively similar for the speakers in these

TABLE 12 The grammar proposed by Luo et al. (2020).

/fAn/	*Voral-N	Rhyme Harm	*Vnasal	Ident-IO (nasal)	Ident-IO (back)
(i) [fAn]	*!	*			
(ii) [fÃn]		*!	*	*	
☞ (iii) [fãn]			*	*	*

two studies. Thus, the learning context that we have attempted to simulate above seems relatively aligned with the L2 populations that we observe from these two studies – i.e., learners with fairly target-like perception of English loVN in place, yet still with atypical L2 production strategies.

A reviewer points out an alternative interpretation – namely, that L2 English learners are so proficient in perceiving these novel contrasts not found in their L1 Mandarin simply because they are easy enough to perceive without L1 experience. In other words: perhaps it is not that Zhang (2023)'s participants had already completed significant L2 perceptual learning about loV-N sequences, but rather that their L1 production grammar's restrictions on loV-N sequences had not dampened their abilities to perceive them (In support of this possibility, see Kabak and Idsardi, 2007 for an example of English coda-onset contrasts which are ungrammatical in Korean but do *not* result in perceptual distortion among L1 Korean L2 English listeners). In the present case, we do not have direct evidence about nasal place perception among monolingual Mandarin listeners (i.e., those who have never been exposed to additional English contrasts), so we are not in a position to choose definitively between these two views. We do note, however, that nasal place contrasts have been reported in both L1 adult and infant studies to be more challenging to discriminate than most (Narayan, 2008; Narayan et al., 2010), and see especially Harnsberger (2001) on the crosslinguistic perceptual difficulties presented by novel L2 nasal place contrasts.

In the following final section, we discuss the consequences of our simulations' crucial assumptions, a comparison with alternative understandings of the L1 Mandarin phonology, and the bigger picture for this type of L1/L2 learning simulation.

5 General discussion

5.1 Alternative analyses of loV-N sequences in Mandarin and their consequences

Previous literature on Mandarin phonology often assumes that the nasal coda in VN sequences is the bearer of the place contrast (Duanmu, 2007; Hsieh et al., 2009 interalia; Luo et al., 2020). In these analyses, the backness of the low vowel is underspecified in the underlying form or else explicitly described as non-contrastive; on the surface the vowel is driven by RHYMEHARMONY to agree with the nasal's underlying specification for [+/–back]. To show how the target output is generated from the input loV-N with underspecified [+/–back] of the vowel (indicated by the archiphoneme), an example in Table 12 is adapted from Luo et al. (2020, p. 19), using a constraint set similar to the one adopted in this paper[8].

In the analysis of Luo et al. (2020), changing vowel place is the only option under consideration, and there is no motivation to alter nasal place from its UR value. A slightly different approach is adopted in Hsieh et al. (2009), though they also assume that Mandarin inputs have an unspecified low vowel that matches its backness with the nasal coda's underlying place. However, given the OT premise of Richness of the Base, Hsieh et al. (2009) also consider inputs with specified vowel place and input sequences that violate Rhyme Harmony, making the change of either the vowel place or the nasal place possible. Thus, Hsieh et al. (2009) proposes that IDENT-CPL-CODA is crucially ranked above IDENT[BK] (which targets vowels), so that here the low vowel is unfaithful to place.

Since we have adopted the opposite assumptions in this paper, we must now consider what sources of evidence there are for the underlying representation of Mandarin low vowels or coda nasals? One such piece of evidence is the romanized orthographic system *Pinyin* for Mandarin. The nasal codas [n] and [ŋ] are transcribed differently in Pinyin as *n* and *ng*, whereas the low vowel is uniformly transcribed as *ɑ*; this could well be interpreted as the latter's under-specification for place in Mandarin speakers' minds.

If we expand our view and consider the status of [+/–back] contrasts in the full Mandarin inventory of vowels and nasal codas, the picture is considerably complicated; here we rely especially on the descriptions and discussion of Xu (1980) and Duanmu (2007). With respect to high vowels, the [+/–back] contrast is maintained in open syllables, where /i/, /y/ and /u/ are all contrastive, but highV-nasal sequences are restricted to a subset which may in part obey RHYMEHARMONY (cf. [yn] and [ʊŋ], *[yŋ] and *[ʊn]). Among mid vowels there is only one contrastive vowel category, all of whose surface allophones' features including [+/–back] are dictated by surrounding segments. When followed by either nasal coda, this mid vowel surfaces invariantly as schwa (transcribed in different sources as [ə] or [ʌ]), apparently not obeying RHYMEHARMONY. Finally, the two surface low vowels are not otherwise contrastive except in the pre-nasal environment we have discussed at length in this paper.

In addition to phonotactic restrictions, several other studies have reported that both high and mid vowels undergo some degree of reduction or merger before nasal codas. With respect to high vowels, some varieties such as Shanghai Mandarin show loss of the /n∼ŋ/ contrast before [i] in both perception and production, and an overall bias toward [iŋ] (Liu and Babel, 2023), and more mixed

8 Note that the constraint Ident-IO(nas) is defined as "the corresponding segments in input and output have identical values for [nasal]" (Luo et al., 2020), which is not the same as our Ident[+/−back]-N. Ident-IO(back) is in fact a constraint targeting identical values for [+/−back].

results among Beijing and Northern Mandarin speakers (cf. Chen and Guion-Anderson, 2011). With respect to mid vowels, loss of contrast between /ən/ and /əŋ/ is reported for some varieties such as Taiwanese Mandarin and Shanghai Mandarin (Chiu et al., 2019; Faytak et al., 2020), with a bias toward /ən/.

Regardless of this complicated landscape, the grammatical upshot is that Mandarin speakers do not get clear evidence from any language-internal data as to how vowel + nasal phonotactics are imposed, because they all represent static restrictions. To return to low vowels, we can observe that lowV-nasal sequences all match for [+/–back] on the surface, but there are no alternations to demonstrate how a multi-morphemic input like /a + ŋ/ or /ɑ + n/ would be mapped to the surface. It is overall clear from the above that across vowels and Mandarin varieties, the [+/–back] place contrast is hard to maintain in both vowels and their following nasal codas, and that many markedness constraints must outweigh faithfulness to [+/–back] in various different ways. What our analysis crucially requires, among all these restrictions on backness, is only that IDENT[BK]-V outweigh IDENT[BK]-N. Regardless of any other preservation or neutralization of place contrasts, in which IDENT[BK]-V might rank above or below many other markedness constraints, all we must predict is that if an input VN sequence is driven by markedness to change one or the other of their underlying places, the vowel's place features will win.

A separate source of data that this paper has treated centrally are the multiple surface variants of /VN/ sequences in Mandarin. Once the grammar is built to produce both [an] and [ã] from the same [–back] input, it becomes harder to analyze their place feature as determined by the underlying nasal. It is not impossible to derive a mapping like /ɑn/ → [ã] as optimal, but it requires an analysis of derived and indeed displaced contrast, whereby low nasalized vowels in the output are faithful to the place of a nasal consonant that has been deleted, but its place retained; or possibly a fusion of both input vowel and nasal segments, violating a series of other Ident constraints but not losing the input consonant's nasality.

Putting derived contrast aside, any L2 account that repairs backness mismatches in favor of the input nasal's place must assume a disconnect between L1 input/output mappings on the one hand (where nasal place is retained) and loanword/perception/L2 production data on the other. This is in fact one of the key arguments of Hsieh et al. (2009), which takes it as a given that the disconnect exists and that it reveals something special of the nature of loanword phonology. Instead: our approach has to been to assume that all sources of evidence point to the vowel as the faithful bearer of [+/–back] specification, at least in the current grammar of speakers like those of Beijing Mandarin, and to reason from there as to the learning consequences of their grammatical assumptions.

5.2 Choosing the properties of our learner

The discussion in the preceding sections provided motivations for some but not all of the properties of our grammar and our learning simulations. We explained our choice of a stochastic grammar (so that the L1 grammar could include multiple surface variants) and the rationale behind our initial values for classes of constraints: with Markedness high, Faithfulness low, and a specific

TABLE 13 Reproducing the grammar in Table 9 with ranked constraints.

	Rhyme Harm 100	Max 53	No Coda 51	Id[bk]-N 1
(a) /æn/				
(i) [æ̃]		*		
☞ (ii) [æn]			*	
(b) /æŋ/				
(i) [æ̃]		*		
☞ (ii) [æn]			*	*
(iii) [æŋ]	*!		*	

relationship between IDENT[BK]-V and IDENT[BK]-N. One point that we have not addressed explicitly, however, is our choice of decision strategy – that is, adopting a Harmonic Grammar with weighted constraints, rather than a classic OT grammar of ranked constraints. But in fact using the HG-GLA system was a key to the simulation results of Section 4.

The relevant property of a weighted constraint grammar is that its choice of an optimum is affected by *all* the constraints that are violated by each candidate; there is no equivalent to the strict ranking that classic OT analyses rely on. In HG, low-weighted constraints can still exert their influence when other constraints might cancel each other out, contributing to what is frequently termed in the HG literature as a *gang effect* (see e.g., Pater, 2009; Jesney and Tessier, 2011; Breiss, 2020, among many others). This is exactly the case in the two mappings illustrated in the grammars of Table 9.

To see the importance of constraint weighting, the grammars in Table 13 reproduce those from Table 9, but now adopting *ranked* constraints. Here we still assume that constraints have numerical values and that these are perturbed at each use of the grammar, but that the classic OT EVAL component chooses the optimal candidate.

In (13a), the choice of candidates is all up to the relative ranking of MAX and NOCODA; with these one-time values of MAX and NOCODA, candidate (ii) wins, but given their similar underlying values, NOCODA will often outrank MAX and candidate (i) will win. In (13b), the situation is in fact exactly the same – all that chooses between the two first candidates is that same ranking, and the fact that Tableau (13bii) violates a lower ranked faithfulness constraint does not make it any different in the grammar's eyes than Tableau 13 (aii). Since Liu (2016)'s L2 production data suggests that the (i) and (ii) candidates do differ between situations as in Table 9, we adopt Harmonic Grammar's weighted constraints to let the faithfulness constraint violated in (9bii) influence our system's overall behavior.

Our other consequential assumption concerns the implementation of our bias between IDENT[BK]-V and IDENT[BK]-N. Following Jesney and Tessier (2011) in particular, we assume that implementing "low" faithfulness in Harmonic Grammar means an initial value that is as low as possible without being zero; thus we start all F constraints at 1. To implement a "fixed" weighting between IDENT[BK]-V and IDENT[BK]-N, we initialized vowel faithfulness at 11; given our particular parameter

TABLE 14 Output proportions from early L2 learning with adjusted weights of IDENT[BK]-V vs. N.

V[-bk] Target	Output Candidate	% at initial state	% after 5 trials
/æn/	[æ̃]	41	0.2
	[æ̃n]	20	2.1
	[æn]	39	97.7
/æŋ/	[æ̃]	100	97
	[æ̃n]	0	0
	[æn]	0	3

settings regarding stochastic noise, this spread of 10 points means that effectively IDENT[BK]-V will never act as though weighted lower than IDENT[BK]-N. We also chose a plasticity value for IDENT[BK]-N that was 10 times slower than that for IDENT[BK]-V. Together, these two settings keep IDENT[BK]-V reliably weighted above IDENT[BK]-N in the early stages of learning, at least. Note, however, that they do not impose a fully-fixed relationship between the two constraints – in fact, the final L2 grammar we report in Table 10 (d) has only 1.5 points between them.

In our L1 simulations, this initial bias between the two Ident constraints ensures that the grammar resolves backness mismatches by changing nasal and not vowel place – *without* any evidence of alternations. Then in the L2 simulations, both Ident constraints must be promoted sufficiently to allow backness violations in English, but in the meantime the lower Ident N constraint's role is to produce the asymmetry just discussed in Table 9. The extent of that asymmetry, and how long it lasts in learning, is a function of the initial value of Ident N. In our default-type setting (IDENT[BK]-V at 11 and IDENT[BK]-N at 1), we saw that deletion is initially chosen for input with matching backness (Tableau 9a) about 10% less often than for those with mismatches (9b), and that this spread is becoming smaller and disappearing after the first 15 trials and beyond. By way of comparison, if everything else is kept the same but the initial values of IDENT[BK]-V vs. N are 30 and 20 respectively, the end state of learning is also successful, but the distributions of winning candidates along the way are quite different. Table 14 shows a typical distribution of winning candidates at and near the beginning of L2 learning, just focusing on the front vowel with coronal and velar nasals.

Compared to the distributions in Section 4.3, this is a more extreme asymmetry in Table 14; both in the difference between /æn/ and /æŋ/ and between 0 and 5 trials. Here, mismatched VN inputs are almost entirely subject to coda deletion, whereas inputs that match for [+/–back] have their nasals deleted at the usual L1 rate initially, and then almost immediately become entirely protected from deletion – so that after 5 trials, the treatment of both inputs is almost diametrically opposite. We suspect that this extreme difference in the input is not a likely trajectory, at least partly given the fairly gradient nature of Liu (2016)'s result. In any event, we thus observe that while constraint initial values may feel arbitrary, they do in fact encode substantial and falsifiable predictions about the course of learning.

5.3 More general consequences for learning simulations and the study of L2 learning

The biggest picture goal of this paper has been to highlight how the details of an L1 grammatical analysis and its inherent variability can influence L2 acquisition, and how that influence can be studied formally using learning simulations.

One caveat we want to emphasize again is that the absolute numbers of learning trials, in these simulations, are merely a function of the parameter settings chosen, especially the plasticity of each constraint demoted and promoted in response to errors, and cannot reflect any direct connection with learning rates. The fact that a learner takes 15 trials to move from one qualitative stage of phonological development to another is not meaningful on its own; we only seek to show that a learner beginning at the initial state (zero trials) will first pass through one type of intermediate stage, onto another, and eventually make it to the end state with success.

Ultimately, we would like to use this kind of learning simulation to not only capture existing data about stages and asymmetries in second language phonological acquisition, but also make novel, testable predictions about L2 perception and production. For example, Luo et al. (2020) report at length about the interaction between (i) RHYMEHARMONY in loV-N sequences and (ii) nasal place sharing with a following stop onset, across a compound boundary (e.g., /fan.kai/ -> [faŋ.kai] *fan.kai* "open"). One direct extension of our analyses here is to consider how the L1 Mandarin grammar might derive this pattern, especially in light of our proposal about the source of underlying [+/–back] place contrasts. We could then use simulations to predict how learners should acquire an L2 English grammar of NC clusters, within and across morphemes (e.g., the place assimilation of *i[mp]ossible*, *i[nt]olerant* and *i[ŋk]capable*, vs. the more gradient assimilation across word boundaries, e.g., Turnbull et al., 2018). Adding the additional dimension of nasal coda place sharing into the inventory of Mandarin surface forms ([æ̃, æn, æ̃n], plus [æŋ.k] and possibly [æ̃.k]) will create further degrees of L1 end-state variability. The specific choice of a constraint set that can make RHYMEHARMONY variably violable – but only when coda-onset place sharing is at issue – will predict the stages through which such an L2 HG-GLA learner should develop, which could then be tested experimentally. The more sources of human behavioral and perceptual data we can compare with a simulated learner, the better we may understand how grammatical and other pressures combine to create second language speech patterns.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: see published data in Liu (2016) and Zhang (2023).

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

SZ: Data curation, Formal analysis, Writing—original draft, Writing—review & editing. A-MT: Conceptualization, Formal analysis, Writing—original draft, Writing—review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by SSHRC Insight Grant #435-2023-0169, awarded to A-MT.

Acknowledgments

We would like to thank the members of the UBC Child Phonology Lab especially Kaili Vesik; Molly Babel, Suyuan Liu, and Yadong Liu for their contributions to and discussion of related work; two very helpful reviewers, Baris Kabak, and audiences at Interspeech2023.

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1327600/full#supplementary-material

References

Angluin, D. (1980). Inductive inference of formal languages from positive data. *Inf. Control* 45, 117–135. doi: 10.1016/S0019-9958(80)90285-5

Becker, M., and Tessier, A. M. (2011). Trajectories of faithfulness in child-specific phonology. *Phonology* 28, 163–196. doi: 10.1017/S0952675711000133

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., and Brasher, A. (2013). The time course of perception of coarticulation. *The J. Acous. Soc. Am.* 133, 2350–2366. doi: 10.1121/1.4794366

Berwick, R. (1985). *The Acquisition of Syntactic Knowledge*. Cambridge, MA: MIT Press.

Boersma, P. (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives* (Doctoral dissertation). University of Amsterdam.

Boersma, P. (2011). A programme for bidirectional phonology and phonetics and their acquisition and evolution. *Bidirect. Optim. Theor.* 180, 33–72. doi: 10.1075/la.180.02boe

Boersma, P., and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Ling. Inq.* 32, 45–86. doi: 10.1162/002438901554586

Boersma, P., and Weenink, D. (2016). *Praat: Doing Phonetics by Computer*. Available online at: http://www.praat.org/

Breiss, C. (2020). Constraint cumulativity in phonotactics: evidence from artificial grammar learning studies. *Phonology* 37, 551–576. doi: 10.1017/S0952675720000275

Broselow, E. (2023). *Probabilistic Cue Weighting in Native and Loanword Perception. The Annual Meeting of Phonology.*

Broselow, E., Chen, S. I., and Wang, C. (1998). The emergence of the unmarked in second language phonology. *Stu. Sec. Lang. Acq.* 20, 261–280. doi: 10.1017/S0272263198002071

Chen, M. Y. (2000). Acoustic analysis of simple vowels preceding a nasal in Standard Chinese. *J. Phonetic.* 28, 43–67. doi: 10.1006/jpho.2000.0106

Chen, Y., and Guion-Anderson, S. (2011). Perceptual confusabiltiy of word-final nasals in southern min and Mandarin: implications for coda nasal mergers in Chinese. *Proc. 17th Int. Cong. Phonetic. Sci.* 22, 464–467. doi: 10.1121/1.3587899

Chiu, C., Lu, Y., Weng, Y., Jin, S., Weng, W., and Yang, T. (2019). "Uncovering syllable-final nasal merging in Taiwan Mandarin: an ultrasonographic investigation of tongue postures and degrees of nasalization," in *Proceedings of the 19th International Congress of Phonetic Sciences* (Melbourne, VIC; Canberra, ACT: Australasian Speech Science and Technology Association), 398–402.

Cohn, A. (1993). Nasalisation in English: phonology or phonetics. *Phonology* 10, 43–81. doi: 10.1017/S0952675700001731

Curtin, S., and Zuraw, K. (2002). Explaining constraint demotion in a developing system. *Proc. Boston Univ. Conf. Lang. Dev.* 26, 118–129.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization* [Ph.D. thesis]. Utercht: Utercht University.

Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Stu. Sec. Lang. Acquisi.* 26, 551–585. doi: 10.1017/S0272263104040021

Fang, Q. (2004). *Acoustic Analysis of Nasal Codas in Connected Speech in Chinese*. Beijing: Phonetic Research Report, Institute of Linguistics, CASS, 91–97.

Faytak, M., Liu, S., and Sundara, M. (2020). Nasal coda neutralization in Shanghai Mandarin: articulatory and perceptual evidence. *Lab. Phonol.* 11, 1–14. doi: 10.5334/labphon.269

Gnanadesikan, A. (2004). Markedness and faithfulness constraints in child phonology. *Constr. Phonol. Acq.* 22, 73–108. doi: 10.1017/CBO9780511486418.004

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *The J. Acoustic. Soc. Am.* 110, 489–503. doi: 10.1121/1.1371758

Hayes, B. (2004). "Phonological acquisition in optimality theory: the early stages," in *Constraints in Phonological Acquisition*, eds R. Kager, J. Pater and W. Zonneveld (Cambridge: CUP), 158–203.

Hayes, B. (2022). Deriving the wug-shaped curve: a criterion for assessing formal theories of linguistic variation. *Ann. Rev. Ling.* 8, 473–494. doi: 10.1146/annurev-linguistics-031220-013128

Hayes, B., and Wilson, C. (2008). A maximum entropy model of phonotactics and phonotactic learning. *Ling. Inq.* 39:3. doi: 10.1162/ling.2008.39.3.379

Hsieh, F. F., Kenstowicz, M., and Mou, X. (2009). Mandarin adaptations of coda nasals in English loanwords. *Loan Phonol.* 307:131. doi: 10.1075/cilt.307.05hsi

Huang, H. H., and Lin, Y. H. (2023). "English vowel quality effects on nasal germination in Mandarin Loanwords: corpus data vs. bilingual production and perception experimental results," in *[Conference Presentation] Annual Meeting of Phonology 2023*.

Jarosz, G. (2006). *Rich lexicons and Restrictive Grammars: Maximum Likelihood Learning in Optimality Theory* [Doctoral dissertation]. Baltimore, MD: Johns Hopkins University.

Jarosz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning. *J. Child Lang.* 37, 565–606. doi: 10.1017/S0305000910000103

Jesney, K., and Tessier, A. M. (2011). Biases in harmonic grammar: the road to restrictive learning. *Nat. Lang. Ling. Theor.* 22, 251–290. doi: 10.1007/s11049-010-9104-2

Kabak, B., and Idsardi, W. (2007). Perceptual distortions in the adaptation of english consonant clusters: Syllable structure or consonantal contact constraints? *Lang. Speech* 50:201. doi: 10.1177/00238309070500010201

Kawahara, S. (2020). A wug-shaped curve in sound symbolism: the case of Japanese Pokémon names. *Phonology* 37, 383–418. doi: 10.1017/S0952675720000202

Keller, F. (2006). Linear optimality theory as a model of gradience in grammar. *Grad. Gram. Gen. Persp.* 12, 270–287. doi: 10.1093/acprof:oso/9780199274796.003.0014

Legendre, G., Miyata, Y., and Smolensky, P. (1990). "Harmonic Grammar - a formal multilevel connectionist theory of linguistic wellformedness: an application," in *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society*. Cambridge: Lawrence Erlbaum, 884–891.

Levelt, C. C., Schiller, N. O., and Levelt, W. J. M. (1999). A developmental grammar for syllable structure in the production of child language. *Brain Lang.* 68, 291–299.

Liu, S., and Babel, M. (2023). "Production and perception of Mandarin /i/-nasal rhymes," in *Proceedings of the 20th International Congress of Phonetic Sciences*, eds R. Skarnitzl and J. Volín (London: Guarant International), 142–146.

Liu, Y. (2016). Oral gestural reduction of English nasal coda produced by Mandarin English-as-a-foreign-language learners. *The J. Acous. Soc. Am.* 140, 3339. doi: 10.1121/1.4970656

Luo, M., Van de Weijer, J., and Sloos, M. (2020). "A perceptually-based OT approach to opacity in Mandarin Chinese nasal rhymes," in *Explorations of Chinese Theoretical and Applied Linguistics*, eds D. Chen and D. Bell (Cambridge Scholar Publishing), 8–43.

McCarthy, J., and Prince, A. (1995). "Faithfulness and reduplicative identity," in *Papers in Optimality Theory. University of Massachusetts Occasional Papers in Linguistics 18*, eds J. Beckman, L. Dickey, and S. Urbanczyk (Amherst, MA: University of Massachusetts; GLSA), 249–384.

Narayan, C. R. (2008). The acoustic-perceptual salience of nasal place contrasts. *J. Phonetic.* 36, 191–217. doi: 10.1016/j.wocn.2007.10.001

Narayan, C. R., Werker, J. F., and Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: evidence from nasal place discrimination. *Dev. Sci.* 12, 407–420. doi: 10.1111/j.1467-7687.2009.00898.x

Nelson, M. (2019). Segmentation and UR acquisition with UR constraints. *Proc. Soc. Comput. Ling.* 2, 60–68. doi: 10.7275/zc9d-pn56

O'Hara, C. (2017). How abstract is more abstract? Learning abstract underlying representations. *Phonology* 34, 325–345. doi: 10.1017/S0952675717000161

Paradis, C., and LaCharité, D. (1997). Preservation and minimality in loanword adaptation. *J. Ling.* 33, 379–430. doi: 10.1017/S0022226797006786

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognit. Sci.* 33, 999–1035. doi: 10.1111/j.1551-6709.2009.01047.x

Pater, J., Staubs, R., Jesney, K., and Smith, B. C. (2012). "Learning probabilities over underlying representations," in *Proceedings of the Twelfth Meeting of the Special Interest Group on Computational Morphology and Phonology*, 62–71

Peperkamp, S., Vendelin, I., and Nakamura, K. (2008). On the perceptual origins of loanword adaptations: experimental evidence from Japanese. *Phonology* 25, 129–164. doi: 10.1017/S0952675708001425

Potts, C., Pater, J., Jesney, K., Bhatt, R., and Becker, M. (2010). Harmonic Grammar with linear programming: from linear systems to linguistic typology. *Phonology* 27, 77–117. doi: 10.1017/S0952675710000047

Prince, A., and Tesar, B. (2004). "Learning phonotactic distributions," in *Constraints in Phonological Acquisition,* eds R. Kager, J. Pater and W. Zonneveld (Cambridge: CUP), 245–291.

Schwartz, B., and Sprouse, R. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Lang. Res.* 12, 40–72.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Ling. Inq.* 27, 720–731.

Smolensky, P., and Géraldine, L. (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*. Cambridge: MIT Press.

Steriade, D. (2001). *The phonology of Perceptibility Effects: The P-map and Its Consequences for Constraint Organization.* Los Angeles, CA: UCLA.

Tessier, A. M. (2016). *Phonological Acquisition*. Basingstoke: Palgrave Macmillan.

Turnbull, R., Seyfarth, S., Hume, E., and Jaeger, T. F. (2018). Nasal place assimilation trades off inferability of both target and trigger words. *Lab. Phonol. J. Assoc. Lab. Phonol.* 9:15. doi: 10.5334/labphon.119

van Leussen, J. W., and Escudero, P. (2015). Learning to perceive and recognize asecond language: the L2LP model revised. *Front. Psychol.* 6:1000. doi: 10.3389/fpsyg.2015.01000

Wang, J. Z. (1993). *The Geometry of Segmental Features in Beijing Mandarin*. Delaware: University of Delaware.

Xu, S. (1980). *Putonghua Yuyin Zhishi [Phonology of Standard Chinese]*. Beijing: WenziGaige Chubanshe.

Yip, M. (1993). Cantonese loanword phonology and optimality theory. *J. East Asian Ling.* 2, 261–291. doi: 10.1007/BF01739135

Zhang, S. (2023). First language effects on second language perception: evidence from english low-vowel nasal sequences perceived by l1 Mandarin Chinese listeners. *Proc. Interspeech* 2023, 4184–4188. doi: 10.21437/Interspeech.2023-1123

Zhao, X., and Li, P. (2022). Fuzzy or clear? A computational approach towards dynamic l2 lexical- semantic representation. *Front. Commun.* 6:726443. doi: 10.3389/fcomm.2021.726443

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 22 May 2024
DOI 10.3389/flang.2024.1325597

OPEN ACCESS

EDITED BY
John Archibald,
University of Victoria, Canada

REVIEWED BY
Vladimir Kulikov,
Qatar University, Qatar
Marion Caldecott,
University of Victoria, Canada
Stuart Davis,
Indiana University, United States

*CORRESPONDENCE
Darin Flynn
✉ dflynn@ucalgary.ca

RECEIVED 21 October 2023
ACCEPTED 27 March 2024
PUBLISHED 22 May 2024

CITATION
Flynn D (2024) Redeployment in language contact: the case of phonological emphasis. *Front. Lang. Sci.* 3:1325597.
doi: 10.3389/flang.2024.1325597

Redeployment in language contact: the case of phonological emphasis

Darin Flynn*

School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

This article applies the notion of redeployment in second language acquisition to contact-induced diachronic changes. Of special interest are cases where a marked phonological contrast has spread across neighboring languages. Such cases suggest that listeners can re-weight and re-map phonetic cues onto novel phonological structures. On the redeployment view, cues can indeed be re-weighted, but phonological structures which underlie a new contrast are not expected to be fully novel; rather, they must be assembled from preexisting phonological structures. Emphatics are an instructive case. These are (mostly) coronal consonants articulated with tongue-root retraction. Phonological emphasis is rare among the world's languages but it is famously endogenous in Arabic and in Interior Salish and it has spread from these to not a few neighboring languages. The present study describes and analyzes the genesis of phonological emphasis and its exogenous spread to a dozen mostly unrelated languages—from Arabic to Iranian and Caucasian languages, among others, and from Interior Salish to Athabaskan and Wakashan languages. This research shows that most languages acquire emphatics by redeploying the phonological feature [RTR] (retracted tongue root) from preexisting uvulars. On the other hand, some languages acquire imitations of emphatics by redeploying the consonantal use of [low] from preexisting pharyngeals. Phonological emphasis is apparently not borrowed by neighboring languages where consonants lack a phonological feature fit for redeployment. The overall impression is that a language in contact with emphatics may newly adopt these sounds as [RTR] or [low] only if the relevant feature is already in use in its consonant system. This pattern of adoption in language contact supports the redeployment construct in second language acquisition theory.

KEYWORDS

Afroasiatic languages, Caucasian languages, Pacific Northwest Plateau, language contact, emphasis (phonological), uvularization, pharyngealization, redeployment

1 Introduction

The retracted coronal consonants known as emphatics (/ṭ ḍ ṣ …/) are found only in a few languages that have innovated them, notably Arabic (Wallin, 1855) and Interior Salish (Shahin, 1996), and in neighboring languages that have borrowed them (e.g., Cook, 1978; Anonby, 2020). A cross-linguistic diachronic study of these sounds may therefore sound niche, even quaint, but in practice the present study validates several complementary ideas that could hardly be broader. The first is Kabak's (2019) dictum that "second-language learning... mimics language change through language contact" (p. 221). On this view, it makes sense to study contact-induced sound shifts using a construct that has proven valuable in second-language acquisition theory, viz. Archibald's (2003; 2005; 2009; 2018; 2021; 2022; 2023) redeployment dictum that, as a rule, "new structures" are never fully so, but are rather "assembled out of the building blocks found in the L1" (2018, p. 15).

A classic example of redeployment in second-language acquisition concerns the English /l–ɹ/ contrast. This distinction is notoriously difficult for adult learners whose L1s have only one liquid phoneme, such as Japanese and Korean (Brown, 2000). Of special interest is that native speakers of Standard Chinese are relatively successful at learning English /l–ɹ/, in spite of their L1 having just one liquid phoneme (Brown, 2000). Brown suggests that these learners derive benefit from the fact that unlike Japanese and Korean, Standard Chinese distinguishes multiple series of [strident, coronal] sibilants: plain /ts ts^{h} s z/ vs. [posterior] /tʂ tʂ^{h} ʂ ʐ/ vs. [front] /tɕ tɕ^{h} ɕ (ʑ)/.[1] Setting aside the details of Brown's analysis, the basic idea is that native speakers of Chinese are able to recycle a distinctive feature from their rich sibilant system to learn English liquids. In particular, the [posterior] feature of the retroflex sibilant series may be repurposed to distinguish /ɹ/ from /l/ in L2 English. Note that [posterior] is used for /ɹ/ in L1 English (Nelson and Flynn, 2022, and references therein), but this phonological feature is not used to distinguish liquids in Standard Chinese (Duanmu, 2007).

In some cases a redeployed structure may be a poor imitation of the target structure, but succeed nonetheless at distinguishing many lexical items in the L2. For instance, Japanese and Korean do not use [posterior], so adult native speakers of these languages cannot redeploy that distinctive feature when learning the /l–ɹ/ contrast in English (cf. Brown, 2000). Paradoxically, however, they appear to be successful at learning the /s–ʃ/ contrast in most (but not all) English words (Eckman and Iverson, 2013). This is surprising because the /s–ʃ/ contrast is based on [posterior] in English (Atkey, 2002; Son, 2005; Clements, 2009, p. 50; Nelson and Flynn, 2022). This paradox is resolved not by rejecting the redeployment dictum, but by leaning into it: "learners are not really successful in acquiring E/š/. In fact, they perceive and produce E[š] by utilizing the feature [front] in their system" (Son, 2005, p. 192). That is, native speakers of Japanese and Korean learn English /ʃ/ as [front] /ɕ/. This strategy is straightforward in the case of Japanese, where [front] /ɕ/ already exists as "a palatalized consonant (Cy)" (Labrune, 2012, p. 68). In Korean, however, only the [front] affricates /tɕ tɕ* tɕ^{h}/ are well-established (Shin et al., 2012, p. 76–78, 195–196); the fricative [ɕ] is strictly an allophone of /s/ "when followed by the vowels /i/ or /j/ or the diphthong /wi/" (Shin et al., 2012, p. 70).[2] In this case, then, redeploying [front] entails a newly assembled phoneme in L2 English, e.g., *push* /p^{h}ʊɕ/.[3] What is redeployed here is the phonological use of [front] in a sibilant, not simply the feature [front], which occurs in most languages, notably in front vowels.

As these examples illustrate, phonological redeployment is akin to Lardiere's (2009) feature re-assembly model of second-language morphology. As such, a redeployment analysis can only be as strong as the evidence that a particular phonological structure is present or absent in the L1 (e.g., [posterior] in Chinese vs. Japanese) and that this structure can play a particular function in a re-assembled representation (e.g., [front] /ɕ/ in L2 English). Accordingly, the present article dwells at some length on the representation of emphatics and related sounds. The upshot is that, as McCarthy (1988) famously put it, "if the representations are right, the rules will follow" (p. 84)—a third dictum validated in the present study. That is, if one assumes the most agreed upon phonological representations for emphatics and related sounds, the redeployment construct helps to make sense of how and why emphatics have developed in and across languages.

Specifically, I will show that Interior Salish and Arabic innovated a series of coronal consonants specified [retracted tongue root] ([RTR]) and that these emphatics were borrowed as such in many neighboring languages (Tsilhqot'in, Kumzari, etc.) by redeploying the feature [RTR] from preexisting uvulars. Importantly, neighboring languages without uvulars (and without any other [RTR] consonant) did not and arguably could not participate in such redeployment. On the other hand, I will show that certain languages with pharyngeals have developed approximate imitations of emphatics. Pharyngeal consonants entail a constriction in the epilarynx and lower pharynx, traditionally represented by the phonological feature [low]. This feature can apparently be used for secondary pharyngealization in consonants, too. For example, the [RTR] emphatic consonants of Arabic were evidently borrowed into Tigre as [low] instead, by redeploying [low] from preexisting pharyngeal consonants to ejectives. The phonological use of [low] in consonants is disputable, if traditional; it is discussed at the end of this article, alongside possible alternatives.

2 The dissemination of phonological emphasis in the Pacific Northwest

This first major section describes how "Salish emphatics" (Shahin, 1996) originated in the Pacific Northwest Plateau (Section 2.1) and then spread via redeployment (Archibald, 2003 *et seq.*) to a string of unrelated languages—Tsilhqot'in (Section 2.2), Nedut'en-Witsuwit'en (Section 2.3), and X̄a'islak̓ala-X̄enaksialak̓ala (Section 2.4).

2.1 Emphasis genesis: Interior Salish

Interior Salish, located in the Pacific Northwest Plateau, is one of two major branches of the Salish family of languages (Czaykowska-Higgins and Kinkade, 1998; Cook and Flynn, 2020; Davis, 2020). Interior Salish consists of a northern branch, which includes Secwepemctsín (Shuswap), Sṫát'imcets (Lillooet), and Nłeʔkepmxcín (Thompson), and a southern branch, which includes Snchitsu'umshtsn (Coeur d'Alene) and Nxaʔamxcín (Columbia-Moses), among others. These languages have long been reported as having retracted coronal consonants and vowels (Kinkade, 1967; Sloat, 1968; Kuipers, 1974; Johnson, 1975; Cook, 1978, 1981, 1984; etc.). The sounds in question are standardly analyzed with the phonological feature [retracted tongue root] ([RTR] or [TR]) in the Interior Salish literature (Cook, 1978, 1985, 1987; Cole, 1987; Czaykowska-Higgins, 1987, 1990; Bessell and Czaykowska-Higgins,

1 These features reflect Duanmu's (2007) analysis, rather than Brown's. However, I use privative features throughout the present article, e.g., [strident], [posterior], and [front] for Duanmu's [+fricative], [−anterior], and [−back], respectively. Duanmu tentatively suggests that the laryngeal feature [aspirated] (i.e., [spread glottis]), rather than [voice], may be constrastive in the fricatives (/s^{h} ʂ^{h} ɕ^{h}/ vs. /s ʂ (ɕ)/), just as it is in the affricates (/ts^{h} tʂ^{h} tɕ^{jh}/ vs. /ts tʂ tɕ^{j}/), e.g. /s^{h}ṣ/ "die" [sz ∼ ss]; /ʂ^{h}ʂ̍/ "history" [ʂʐ ∼ ʂʂ] (p. 24). The latter suggestion is consequential for redeployment, as discussed in Archibald (2023).

2 Thus "Korean and Japanese native speakers often mispronounce the English word 'see', as they apply the allophonic rules of their native language to the pronunciation of the English word" (Shin et al., 2012, p. 71).

3 Son (2005) suggests that beginners may approximate English /ʃ/ by "using their available L1 resources to mimic it with the *sequence* /s/ + [front]" (p. 137; emphasis in original). Thanks to Bill Idsardi (p.c.) for this example.

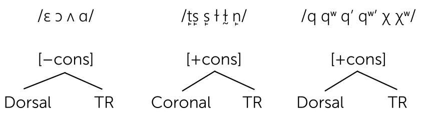


FIGURE 1
Representations of retracted vowels, retracted coronals, and uvulars adapted from Czaykowska-Higgins (1990, p. 3). TR, Tongue Root.

1991; Bessell, 1993, 1998a,b; Shahin, 1996, 2002; Ananian and Nevins, 2001; McDowell, 2004; Namdaran, 2005, 2006; etc.).

For example, Czaykowska-Higgins (1990, p. 2) reports that in Nxaʔamxcín the vowels /i u ə a/ and the coronal consonants /ts s l l̰ n/ "all have retracted counterparts," viz. /ɛ ɔ ʌ ɑ/ and /t̙s̙ s̙ ɫ l̰ n̙/; that "the "darkened" timbre of these sounds is due to uvularization," i.e., "retraction of the tongue root"; and that, "[w]hile retracted vowels and consonants may appear in morphemes or words which contain no back consonants, it is interesting to note that they may also be found (directly) adjacent to uvular segments" (Czaykowska-Higgins, 1990). That is, uvulars cause adjacent /i u ə a/ and /ts s l l̰ n/ to become retracted as [ɛ ɔ ʌ ɑ] and [t̙s̙ s̙ ɫ l̰ n̙], like the underlyingly retracted vowels and coronals. She concludes that retracted vowels, retracted coronals, and uvulars uniquely share a tongue-root retraction feature, as shown in Figure 1.[4]

> Both retracted vowels and retracted consonants are produced by retracting the root of the tongue. Since uvular consonants trigger retraction of adjacent vowels or coronal consonants, then one may assume that uvulars also involve tongue root retraction. (Czaykowska-Higgins, 1990, p. 2)

Indeed, ultrasound studies suggest that "the articulation of uvular consonants universally includes a retracted tongue root position" (Namdaran, 2006, p. 14). In particular, several ultrasound studies of the neighboring Interior Salish language St'át'imcets confirm that retracted coronals share a distinct tongue-root retraction gesture with uvulars (Namdaran, 2005, 2006; Hudu, 2008; Allen et al., 2013, p. 199–200). These studies also confirm the consensus view among phonologists that "uvulars are, in fact, dorsal as well as tongue root segments" (Czaykowska-Higgins, 1990, p. 3),[5] e.g.,

> St'át'imcets uvular consonants possess a raised and retracted tongue dorsum articulation toward the upper-pharyngeal/posterior-uvula region of the vocal tract, as well as a tongue root constriction toward the lower pharynx. (Namdaran, 2006, p. 153)

Tongue-dorsum raising is far less consistent in the retracted coronals (Namdaran, 2006). This, too, conforms with some phonologists' claim that [coronal] emphatics are [RTR], but not necessarily [dorsal] (as in Figure 1).[6]

As an important aside, Nxaʔamxcín appears to be unique among Interior Salish languages in having true pharyngeals, including voiceless /ħ ħʷ/ (Bessell, 1993, p. 93). The phonetic effect of these pharyngeals on adjacent vowels is different from that of retracted coronals and uvulars. The unrounded pharyngeal consonants /ħ ʕ ʕ̰/ cause /i u ə/ to lower as [e o a], and /a/ to be "slightly fronted" (Czaykowska-Higgins, 1990, p. 2, fn. 4). The latter effect was first reported by Kinkade (1967, p. 232): "Pharyngeals may have some effect on neighboring vowels. The most notable is a marked fronting of /a/ in immediate proximity to /ḥ/ or /ʕ/ (e.g., Cm ḥácəm *tie*)." This effect has also been reported for the pharyngeals /ħ ʕ/ in other languages such as Akkadian and Arabic (Harrell, 1957; Colarusso, 1985, p. 366; Hayward and Hayward, 1989, p. 187; Herzallah, 1990, p. 29, 59; McCarthy, 1994, p. 197; Rose, 1996, p. 87; Shahin, 2002, 2011, p. 612; Watson, 2002, p. 271–272, 277–278; Moisik, 2013, p. 484; Sylak-Glassman, 2014, p. 72). For instance, "the tongue body is front with the Arabic pharyngeals, as we can see by the adjacent front allophone of the low vowel" (McCarthy, 1991, p. 78).[7]

Pharyngeal consonants are traditionally represented by the distinctive feature [low] in phonological theory (Chomsky and Halle, 1968, p. 305; Ladefoged, 1971, pp. 92–94; Lass and Anderson, 1975, p. 18; Prince, 1975, p. 12; Rood, 1975, p. 329–333; Halle, 1983; Halle and Clements, 1983; Cole, 1991, p. 25; Coleman, 1998, p. 69; Jensen, 2004, p. 97; Calabrese, 2005, p. 59–60; Hayes, 2009, p. 87–88; Miller, 2011, p. 434; Flynn, 2012, p. 142–144; Odden, 2013, p. 54, 60; among many others). The basic idea is that the canonical low vowel /a/ corresponds to the approximant /ʕ/ in consonant positions, as shown in Figure 2.[8] Crucially, the feature [low] is considered *least marked* in syllable-nucleus position and *most marked* in syllable margins (Prince and Smolensky, 2004, p. 157). Using [low] as in Figure 2 therefore nicely captures the typological fact that all

4 Czaykowska-Higgins (1990) follows an old tradition here in taking the tongue body to be the designated articulator of not only dorsal consonants, but also vowels (Sievers, 1881, p. 93ff; Chomsky and Halle, 1968, p. 302; Sagey, 1986).

5 See also Czaykowska-Higgins (1987), Gorecka (1989), Goad (1989), Bessell (1993), Davis (1993; 1995, p. 471–472), Halle (1995, p. 18), Mahadin and Bader (1995), Rose (1996), Shahin (1996, 1997, 2002, 2011), Zawaydeh (1998), Watson (1999), Al-Raba'a and Davis (2020, p. 22ff), Abo Mokh and Davis (2020, p. 40–41), among many others.

6 See also Czaykowska-Higgins (1987), Goad (1989, 1991), Bessell (1993), Davis (1993; 1995, p. 471–472), Al-Raba'a and Davis (2020, p. 22ff), Abo Mokh and Davis (2020, p. 40–41), among others.

7 The fronting effect of pharyngeals is apparently a consequence of their "double bunching" (Catford, 1983, p. 349). Roughly, the tongue is displaced forward by the lower pharyngeal constriction which accompanies the epilaryngeal constriction in pharyngeal consonants (for details, see Catford, 1983, p. 349; Moisik, 2013, p. 482–500; Sylak-Glassman, 2014, p. 70–73; Beguš, 2021, p. 715).

8 The nucleus is not recognized in standard moraic theory (Hayes, 1989) but this syllabic constituent is essential to a wide range of phonological phenomena (Shaw, 1993, 1994, 1996, 2002; Shaw et al., 1999; Bach et al., 2005; Davis, 2006). For instance, the nucleus is the unitary structure behind diphthongs. Standard moraic theory treats bimoraic diphthongs (e.g., /ɑɪ/ in "buy", /ɔɪ/ in "boy") the same as bimoraic vowel-consonant sequences (e.g., /ɑm/ in "bomb", /ɔɹ/ in "bore"), but this uniform treatment is belied by phonological and psycholinguistic facts. To give just one example, fluent backward talkers (Cowan et al., 1985) reverse the order of bimoraic vowel-consonant sequences (e.g., /mɑb/ for "bomb", /ɹɔb/ for "bore") but they leave the components of bimoraic diphthongs in order—the nucleus is preserved as a unitary structure (e.g., /ɑɪb/ for "buy", /ɔɪb/ for "boy").

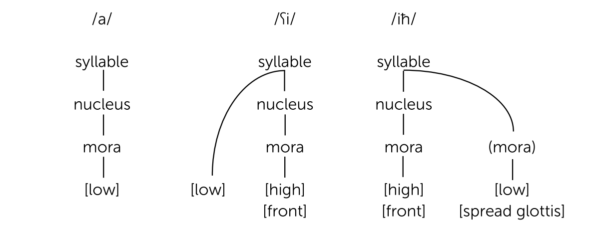


FIGURE 2
Unmarked [low] in syllable nucleus vs. marked [low] in non-nuclear positions.

spoken languages have a low vowel whereas only a small number of languages have pharyngeals.

The understanding of true pharyngeals as [low] rather than [RTR] helps to explain why adjacent non-low vowels become lower, but not necessarily more retracted, and why adjacent low vowels may even be slightly fronted, as in Nxaʔamxcín (Kinkade, 1967, p. 232; Czaykowska-Higgins, 1990, p. 2, fn. 4). However, it should be noted that the latter effects are not observed elsewhere in Interior Salish (Bessell, 1993, p. 98). The so-called pharyngeals in other Interior Salish languages turn out to be uvular approximants /ʁ̞ ʁ̞^{w} ^{ʔ}ʁ̞ ^{ʔ}ʁ̞^{w}/ (Namdaran, 2006, p. 145, and citations therein). That these uvulars have become true pharyngeals in Nxaʔamxcín is not surprising—“there is a common sound change of uvulars to pharyngeals” (Blevins, 2004, p. 198), as seen, for instance, “in every branch of Semitic” (Namdaran, 2006), in Wakashan (Jacobsen, 1969) and in Haida (Eastman and Aoki, 1978). As Weiss (2015) remarks, “the typological surveys of Simpson (2003) and Kümmel (2007) show that uvulars frequently become pharyngeals but pharyngeals don’t often become uvulars” (p. 135). “All evidence points to pharyngeals as an innovation in Southern Interior Salish due to a regular uvular to pharyngeal sound change” (Blevins, 2004, p. 198).

This brings us to the origin of retracted coronal consonants and vowels in Interior Salish. Speakers of Proto Interior Salish innovated these sounds by spreading the retracted articulation of their uvular obstruents /q q^{w} q’ q^{w}’ χ χ^{w}/ and uvular approximants /ʁ̞ ʁ̞^{w} ^{ʔ}ʁ̞ ^{ʔ}ʁ̞^{w}/ inside words (Kuipers, 1981; Cook, 1985, 1987; Van Eijk and Nater, 2020). More specifically, the emphatic series /ṭṣ ṣ .../ developed by assimilating the phonological feature [RTR] from a uvular in the same word. [RTR] assimilation arguably remains an active phonological process in certain Interior Salish languages (e.g., Cole, 1987; Czaykowska-Higgins, 1990; Ananian and Nevins, 2001; Shahin, 2002; cf. Davis, 2020, p. 458). The diachronic and synchronic spread of emphasis in Interior Salish words is a handy analogy for the fact that phonological emphasis has spread to unrelated languages to the north of Interior Salish.

2.2 Emphatics via dentals: Tsilhqot’in

The Athabaskan language Tsilhqot’in has a series of retracted coronals which patterns with uvular consonants, just like its Interior Salish neighbors to the south (Krauss, 1975; Cook, 1978, 1983, 1984, 1993a,b; Latimer, 1978, p. 237–238, 2013, p. 20; Goad, 1989; Ananian and Nevins, 2001; Hansson, 2010, p. 79–81; Bird and Onosson, 2022). Hansson (2010) gives a pointed description:

> [A]lveolar sibilants in Tsilhqot’in contrast in pharyngealization, with “sharp” ([−RTR]) /s, z, ts, ts’, dz/ vs. “flat” ([+RTR]) /sˤ, zˤ, tsˤ, tsˤ’, dzˤ/. Consonant harmony operates over precisely this distinction, making it a rare instance of secondary-articulation harmony… In Tsilhqot’in, all alveolar sibilants in a word agree in [±RTR], with the rightmost one determining their surface [RTR] value. … Tsilhqot’in also has a velar vs. uvular contrast (/k/ vs. /q/, etc.), which also appears to involve [±RTR] (Cook, 1993a), given that uvulars and “flat” sibilants have the exact same lowering and / or retraction effect on neighboring vowels (/æ/→ [ɑ], /u/ → [o], and so forth). (p. 164)

As this quote illustrates, the feature [RTR] is generally assumed for both coronal emphatics and uvulars in Tsilhqot’in (Latimer, 1978; Cook, 1984, 1993a, 2013, p. 35–37; Goad, 1989, Ananian and Nevins, 2001; Hansson, 2010).[9] Flynn and Fulop (2014, p. 215) “suggest that uvulars acted as an origin of the pharyngealization in the emphatic coronals,” such that even today, “uvulars pattern with emphatic coronals in triggering flattening consonant harmony in Tsilhqot’in, e.g., *ts’iqi, ṭṣ’iqi* [ṭṣ’əiqəi] “woman” (cf. *tsiṣaj, ṭṣiṣaj* [ṭṣəiṣɑj] “sand”; Cook, 1983, 1993a).”

It is now possible to be more concrete: the phonological feature [RTR] was redeployed as phonological emphasis from the uvulars, which date back to Proto-Athabaskan (Leer, 1979; Cook, 1981). More specifically, Tsilhqot’in speakers repurposed the [RTR] feature of their large uvular series /q q^{w} q^{h} q^{wh} q’ q^{w}’ χ χ^{w} ʁ ʁ^{w}/, turning an earlier series of dental obstruents into emphatic sibilants, under the influence of emphatic coronals (including sibilants) in neighboring Interior Salish languages. Emphatic coronals are rare sounds so it is unlikely that they developed in Tsilhqot’in independently of their use in neighboring Interior Salish languages. The examples in (1) illustrate that dental consonants, which remain intact in Dëne Sułiné (among other northern Athabaskan languages), have evolved into emphatics (written <ŝ ẑ tŝ dẑ tŝ’>) in Tsilhqot’in (Cook, 2004; Flynn and Fulop, 2014).

(1)

Dëne Sułiné	*Tsilhqot’in*		*Dëne Sułiné*	*Tsilhqot’in*	
t̪θ^{h}ɛ̃ɬ	ṭṣ^{h}ĩ́ɬ	“axe”	θɛ-	ṣɛ-	perf. conj.
-t̪θ^{h}í	-ṭṣ^{h}í	“head”	jaθ	jəṣ	“snow”
-t̪θ^{h}ə́n	-ṭṣ^{h}ə̃́	“meat”	-ðe	-ṣi	“belt, hide”
-t̪θ’i	-ṭṣ’i	“stay (pl)”	-ðá	-ẓí	“mouth”
t̪θ^{h}aj	ṣaj	“sand”	-néð	-neẓ	“long”

Flynn and Fulop (2012, 2014) explain that dental obstruents like [θ] are somewhat grave, auditorily, in the precise sense

9 Latimer (1978) first characterized the “flat” consonants in Tsilhqot’in with “the feature [RTR]” which, he tentatively suggests, “corresponds to a contraction of the styloglossus” (p. 54).

(2) *F-mutation in Witsuwit'en* (Wright et al., 2002, p. 46).

/ə/	/-t'əts/	[t'ʌts]	"incisor"	/tʰəz/	[tʰʌz]	"cane"	*cf.*	/təz/	[təz]	"driftwood"
/o/	/-t'ots/	[t'ɔts]	"peel, bark"	/tʰo/	[tʰɔ]	"water"		/toso/	[tosɔ]	"gunny sack"
/a/	/t'ats/	[t'ɑts]	"backward"	/tʰaj/	[tʰɑj]	"paternal uncle"		/taji/	[tæji]	"appointed chief"

that the high-frequency noise above 2.5 kHz is not predominant, meaning that the amplitude of noise below 2.5 kHz is at least as great. They use this acoustic property to explain the varied shifts of dentals in other northern Athabaskan languages—to laterals, which are also somewhat grave; to velars, which are also grave; and to labials and labiovelars, which are not only grave but also flat, because they involve a "downward shift of a set of formants" (Jakobson et al., 1952, p. 31; see also Trubetzkoy, 1939, p. 127ff). Crucially, tongue-root retraction is also somewhat flat (e.g., it lowers F2) and as such, it lowers the noise spectrum of otherwise acute sibilants. Flynn and Fulop (2014, p. 216) suggest that Tsilhqot'in speakers traded in their grave dentals for the flat sibilants of their Interior Salish neighbors on the basis of this lowered spectrum. See Flynn and Fulop (2014) for a broader discussion of "grave" and "flat" in sound change.

Note that an emphatic-dental connection is recognized elsewhere. The emphatic approximants <z̲ z̲'> are interdental in the Mount Currie dialect of the Interior Salish language St'at'imcets, which adjoins Tsilhqot'in (Van Eijk, 1997, p. 4; Shahin, 2002, p. 177–178). Notably, too, emphatic *ḍ* and *ẓ* are [ð̣] in many dialects of Arabic (Bellem, 2014).[10]

2.3 Phonological emphasis via fortis: Nedut'en-Witsuwit'en

Tsilhqot'in borders another Athabaskan language to the north, Nedut'en-Witsuwit'en, in which fortis consonants cause any following vowel to be lowered and retracted, an effect called F(ortis)-mutation (Cook, 1984, 1987, 1989, 1990; Story, 1984; Vaux, 1996, p. 176–177; Wright et al., 2002, p. 46–48; Hargus, 2007). The fortis set in question consists of the voiceless fricatives /ɬ s ç χʷ χ h/ as well as stops and affricates which are either [constricted glottis] ([CG]) /t' tɬ' ts' (tʃ') c' qʷ' q' ʔ/ or [spread glottis] ([SG]) /tʰ tɬʰ tsʰ cʰ (tʃʰ) qʷʰ qʰ/.[11] The examples in (2) illustrate that /ə o a/ change to [ʌ ɔ ɑ] after ejective /t'/ and aspirated /tʰ/, but not after plain /t/, which is approximately lenis [d̥]. F-mutation affects the other vowels similarly, but more complexly: /i/ changes to [e] in closed syllables (except in loans) and to [əj] in open or laryngeal-closed syllables; /e/ changes to [ɛ]; /u/ remains unchanged or else changes to [o̞] (Hargus, 2007, p. 186).

Cook (1984, 1987, 1989, 1990) suggests that fortis consonants in Nedut'en-Witsuwit'en have a secondary articulation of pharyngealization, which affects adjacent vowels like the emphatics do in Tsilhqot'in and Interior Salish.

> I have no doubt in my mind from my experience with "flattened" vowels in Chilcotin (see Cook, 1983) and "retracted vowels" in Interior Salish (see Cook, 1985), that the phonetic basis for the vowel quality in the fortis syllable is the retracted tongue root—narrowed pharyngeal cavity. (Cook, 1989, p. 139)

More specifically, Cook treats "F-mutation as a process of pharyngealization" (Cook, 1990, p. 124) which is triggered by the feature [RTR] (Cook, 1989, p. 139) or [radical] (Cook, 1990, p. 303). This implies that Nedut'en-Witsuwit'en speakers phonologized tongue-root retraction in fortis consonants. Such retraction in voiceless obstruents is well-understood as aerodynamically motivated: the pharyngeal cavity is constricted by retracting the tongue root, which increases the supraglottal pressure, which in turn serves to inhibit passive voicing in fortis consonants (Trigo, 1991; Vaux, 1992, 1996).

Interestingly, a strong prediction follows from Cook's (1989) [RTR]-analysis of F-mutation in Nedut'en-Witsuwit'en: if this analysis is correct, then the same effect on vowels is expected from non-fortis uvulars, on the assumption that all uvulars are [RTR] (see Section 2.1). As it happens, this is precisely what Cook (1990) found—"F-mutation is triggered not only by a fortis consonant, but also by any consonant of the Q-series [i.e., uvulars]" (p. 132), including lenis /q/ ([ɢ̥]) and /ʁ/ (transcribed by Cook as /G/ and /ɣ/, respectively), e.g., /qis/ [ɢ̥eis] "spring salmon" (p. 129), /peʁu/ [b̥eʁo] "his / her tooth" (p. 132), /qʰequni/ [qʰɛɢ̥oni] "leather shoe" (Cook, 1990).[12]

10 Tellingly, emphatic ḍ [ð̣] has become labiodental [f] in a subvariety of Faifi Arabic (Davis and Alfaifi, 2019). This substitution is expected on Flynn and Fulop's (2012, 2014) claim that a pharyngealized dental like [ð̣] and a labiodental like [f] are both grave and flat, auditorily. Interestingly, Shockley (2024) reports an intermediate sound in Musandam Arabic: "pharyngealized linguolabial fricative [ð̼ˤ]" (p. 16), e.g., *wað̼ˤaʕ* "situation."

11 The labio-dorsals are uncertain. They are rounded uvulars in Nedut'en-Witsuwit'en according to Story (1984, p. 25) and Cook (1989, p. 139), among others. Thus /ə/ becomes retracted and rounded as [ɔ] before the labio-dorsals (Story, 1984). Hargus (2007) notes that labio-dorsals in Nedut'en-Witsuwit'en originate from labialized uvulars in Proto-Athabaskan (p. 29, fn. 14) and that they pattern with non-labialized uvulars in a recent merger of laryngeal contrasts in Witsuwit'en (cf. p. 221–223): *q, *qʰ > qʰ; *qʷ, *qʷʰ > qʷʰ. "The mergers," Hargus (2007, p. 222, fn. 58) remarks, "are also in accord with a phonetic universal proposed by Maddieson (1997) that VOT duration correlates with degree of backness: the backer the sound the longer its VOT." However, Story (1984) reports that Proto-Athabaskan high vowels became mid before non-labialized uvulars, whereas these vowels remain high before labio-dorsals, so Hargus (2007, p. 34) argues that the latter are actually labialized velars. At any rate, "labio-velars are relatively rare, particularly when adjacent to vowels other than /ə/." (Hargus, 2007, p. 157).

12 Cook transcribes the first vowel as [e], but it is expected to be [ɛ] after the fortis uvular /qʰ/. This word is usually recorded with ɛ in the first syllable (e.g., Hargus, 2007, p. 243).

According to Hargus (2007, p. 215–218), the phonetic gesture that was phonologized in Nedut'en-Witsuwit'en fortis consonants was not tongue-root retraction, but rather larynx raising. Specifically, she identifies "synchronic F-mutation" (Story, 1984, p. 30) with [–lowered larynx], a phonological feature proposed by Trigo (1988, 1991). Larynx raising has long been associated with pharyngeal constriction, e.g., "the smaller pharynx is produced by retracting the root and raising the larynx. The vertical position of the larynx is reasonably well-correlated with the position of the tongue root" (Lindau, 1975, p. S12; see also Trigo, 1988, 1991; Moisik, 2013). So it is reasonable for Hargus (2007) to ascribe "synchronic pharyngealization" (Cook, 1989, p. 141) to [–lowered larynx].

However, there are several reasons for doubt. First, if F-mutation is caused by [–lowered larynx], why would the [+lowered larynx] "lenis consonants of the Q-series trigger F-mutation" (Cook, 1990, p. 133), as noted above? Second, the majority of fortis consonants in Nedut'en-Witsuwit'en are [spread glottis] according to Hargus (2007): "Voiceless fricatives are [+spread glottis]" (p. 217), just like "the [+spread glottis] voiceless aspirated stops / affricates" (Hargus, 2007).[13] The problem here is that [spread glottis] is normally associated with larynx lowering, not raising, according to Trigo herself (see also Esling et al., 2019, p. 18). To give just one example: "In the case of Madurese, as discussed by Trigo, it seems quite plausible that the *aspirated stops* and the voiced stops are indeed both [+LL]" (Cohn, 1993, p. 119, italics added). Third, there is no precedent or independent motivation for the phonological feature [LL] in Nedut'en-Witsuwit'en, unlike [RTR], which has long been assumed for uvulars and other sounds in Athabaskan (Latimer, 1978; Cook, 1985, 1989; Czaykowska-Higgins, 1987; Goad, 1989; see Section 2.1). Finally, it must be said that Trigo's (1988, 1991) proposed feature is considered "somewhat controversial" (Cohn, 1993, p. 118) and "tentative" (Moisik, 2013, p. 405).

The pharyngealization of fortis consonants in Nedut'en-Witsuwit'en likely occurred under the influence of secondary pharyngealization in Tsilhqot'in to the south—pharyngealized sounds are rare, so it is unlikely that they developed independently in these neighboring, closely-related languages.[14] In this contact situation, the only possibility for redeployment was the [RTR] feature of uvulars in Nedut'en-Witsuwit'en. By contrast, [–lowered larynx] had no precedent in Nedut'en-Witsuwit'en, as mentioned. As explained above, tongue-root retraction is an enhancement gesture for fortis consonants (cf. Stevens and Keyser, 1989, 2010; Keyser and Stevens, 2001, 2006), which was phonologized by repurposing the [RTR] feature of uvulars to voiceless fricatives and to stops and affricates which are either [CG] or [SG].[15] Thus all fortis consonants, along with the lenis uvulars /q ʁ/, cause lowering and retraction in following vowels, due to their [RTR] specification, like the emphatic sibilants and uvulars in neighboring Tsilhqot'in.

2.4 Phonological emphasis via fortis, again: X̄a'islak̓ala-X̄enaksialak̓ala (Haisla)

In Nedut'en-Witsuwit'en, the uvular series /q qʰ q' χ ʁ/ does not contrast with a velar series /k kʰ k' x ɣ/, but rather with a palatal one /c cʰ c' ç j/ (Wright et al., 2002; Hargus, 2007). This is the outcome of a phonetic-distancing effect called "polarization" (Keating, 1984; Ladefoged and Maddieson, 1996, p. 46). The use of palatalization to distance velars from uvulars is attested elsewhere in Athabaskan (Leer, 2011; Flynn and Fulop, 2014, p. 210), including eastern Ahtna (Kari, 1977, p. 284–285) and Hupa (Woodward, 1964, p. 200; Gordon, 1996). It is also an areal feature shared by languages to the west of Nedut'en-Witsuwit'en, including the Wakashan language X̄a'islak̓ala-X̄enaksialak̓ala (Lincoln and Rath, 1986) and the Tsimshianic languages Gitxsan (Brown et al., 2016, p. 368–369) and Sm'algyax (Dunn, 1995).

Ironically, the Nedut'en-Witsuwit'en strategy to palatalize the velar stops (in order to create distance from the uvular stops /q qʰ q'/) resulted in palatal stops /c cʰ c'/ which are more similar to the alveolar stops /t tʰ t'/. This may be a contributing factor to the shift of palatal stops to more distinct [strident] affricates in syllable onset position in the Nedut'en /U'in Wit'en dialects of Fort Babine (Wit'at) and Takla Lake (Hargus, 2007, p. 6), e.g., /cəs/ > /tʃəs/ "hook" (cf. /təz/ "driftwood"), /cʰəs/ > /tʃʰəs/ "down feathers" (cf. /tʰəz/ "cane"), /tinc'əj/ > /tintʃ'əj/ "four" (cf. /-t'əts/ "incisor").[16]

This dilemma is also evident in the neighboring Wakashan language X̄a'islak̓ala-X̄enaksialak̓ala, also known as Haisla (Lincoln and Rath, 1980, 1986; Bach, 1991, 1997). Apparently in order to create phonetic distance from the uvular stops /q, ɢ, q'/, the velar stops are strongly palatalized /kʲ ∼ c, gʲ ∼ ɟ, kʲ' ∼ c'/, so much so that they risk confusion with the alveolar stops /t, d, t'/. With this in mind, the following fact is striking:

> It is a peculiarity of Haisla that /t/ and /t'/... cause a following vocalic plain resonant to sound like after a plain uvular. (Lincoln and Rath, 1980, p. 25)
>
> In Kitimaat, with the phonemes /t/ and /t'/... a following vocalic resonant is pronounced as after an unrounded uvular... cf. /tl̩qʷ/ [tʰʌlχʷ] "soft", /ˈtiɬa/ [ˈtʰɛɪɬa] "to fish with a line and baited hook", /ˈt'm̩sdu/ [ˈt'ɑmst̬u] "stye", /ˈt'uxʷa/ [ˈt'oʊxʷa] "big wave, ocean swell", /t'l̩ːs/ [t'ʌlːs] "cranberry." (Lincoln and Rath, 1986, p. 45)

13 See also Vaux (1998), Vaux and Miller (2011), and Esling et al. (2019, p. 42–43).

14 Babine-Witsuwit'en and Tsilhqot'in share certain grammatical innovations, too (e.g., Hargus, 2007, p. 371; Cook, 2013, p. 521–522).

15 This shift may also have occurred on the basis of acoustic similarities between the stiff vocal folds of voiceless consonants in Babine-Witsuwit'en and the retracted tongue root of emphatic consonants in neighboring Tsilhqot'in. Stiff vocal folds increase the relative intensity of higher frequencies. Similarly, tongue root retraction or pharyngeal contraction increases damping of F1, which causes the spectrum to sound brighter in the high frequencies (Fulop et al., 1998; Guion et al., 2004). Thanks to Sean Fulop for helpful discussion.

16 In practice, the new U'in Wit'en affricates /tʃ tʃʰ tʃ'/ are more similar to the preexisting affricates /ts tsʰ ts'/ than the palatal stops /c cʰ c'/ were. This similarity between sibilant series is precisely what drove /ts tsʰ ts'/ to become dental /t̪θ t̪θʰ t̪θ'/ in the immediate precursor of Tsilhqot'in, as discussed in Section 2.2 (cf. Leer, 2011; Flynn and Fulop, 2014).

The fact that /t t'/ have the same lowering and retracting effect as /q ɢ q' χ/ suggests that the [RTR] feature was redeployed from the latter to the former in X̄a'islak̓ala-X̄enaksialak̓ala, under the influence of neighboring Nedut'en-Witsuwit'en, in which /t^h t'/ have precisely the same lowering and retracting effect; see (2) above. The secondary articulation of pharyngealization renders /t/ and /t'/ auditorily flat in Haisla, which presumably helps to distinguish them from the auditorily sharp /k^j ∼ c/ and /k^{j}' ∼ c'/, respectively.

Note, finally, that /d/ causes no lowering or retraction in adjacent sounds in X̄a'islak̓ala-X̄enaksialak̓ala. A secondary articulation of pharyngealization would render /d/ auditorily flat, which would presumably help to distinguish it from the auditorily sharp /g^j ∼ ɟ/. However, /d/ is evidently not [RTR] in X̄a'islak̓ala-X̄enaksialak̓ala, presumably for the same aerodynamic reason that lenis non-uvular consonants are not [RTR] in Nedut'en-Witsuwit'en (Section 2.3). As Vaux (1996, p. 178) puts it: "Phoneticians have long known that advancement of the tongue root is necessary to produce voicing in stop consonants (for a review of the literature, see Vaux, 1992)." Such tongue-root advancement is obviously antagonistic to [RTR], which helps to explain why voiced /d/ is not [RTR], unlike voiceless /t/ and ejective /t'/.[17]

2.5 Interim conclusion

The retracted consonants known as emphatics were innovated in Interior Salish and then spread to neighboring (unrelated) languages. Such cases demonstrate that listeners can re-weight and re-map phonetic cues onto novel phonological structures. On Archibald's conception of redeployment, cues can indeed be re-weighted, but phonological structures which underlie a new contrast are not expected to be fully novel; rather, they must be assembled from preexisting phonological structures: "We need to look at what cues are detected in the input, which subset of the input becomes intake, and how this intake is parsed onto phonological structures" (Archibald, 2023, p. 288). As diagrammed in Figure 3, the feature [RTR] was redeployed from uvulars to other consonants on the basis of partial acoustic/auditory similarities with emphatic sounds in a neighboring language in Tsilhqot'in, Nedut'en-Witsuwit'en, and X̄a'islak̓ala-X̄enaksialak̓ala (Haisla).

On this understanding, languages without uvulars (and without any other [RTR] consonant) are not expected to participate in the areal spread of phonological emphasis. A case in point is the Northern Athabaskan language Dakelh (a.k.a. Carrier) spoken in the Central Interior of British Columbia, Canada (Morice, 1932; Walker, 1979; Story, 1984; Bird, 2003; Gessner, 2003). Dakelh has been in direct contact for centuries with the Athabaskan languages Nedut'en-Witsuwit'en and Tsilhqot'in (Chilcotin), with the Wakashan language X̄a'islak̓ala-X̄enaksialak̓ala, and with the Interior Salish languages Secwepemctsín (Shuswap) and St̓át̓imcets (Lillooet). Phonological emphasis has spread as an areal feature across all these other languages, but Dakelh remains unaffected. For context, Dakelh is closely related to Nedut'en-Witsuwit'en, within a larger "Central British Columbia" group which also includes Tsilhqot'in and the extinct language Nicola. Proto-Athabaskan uvulars are preserved in Nedut'en-Witsuwit'en and Tsilhqot'in, as we have seen. Uvulars were also preserved in Nicola according to Boas (1924, p. 36–37). However, the historical uvulars have shifted to velars in Dakelh (Leer, 1996, p. 197; Hargus, 2007, p. 11; Flynn and Fulop, 2014, p. 202). Thus, Dakelh does not have any uvulars, nor any other type of consonant with a phonological feature to redeploy as phonological emphasis.

17 North Wakashan languages like X̄a'islak̓ala-X̄enaksialak̓ala have three series of stops and affricates: plain voiceless, [voice], and [constricted glottis] (Howe [Flynn], 1999c, 2000).

3 The dissemination of phonological emphasis from Arabic

This second major section explains how pharyngealization originated in Arabic and then spread via redeployment to other languages. Of particular interest are cases in which the borrowed sounds have a different phonological structure than the original ones in Arabic.

3.1 Emphasis genesis: Arabic

The term emphasis has long been used in Afroasiatic studies not only for a secondary pharyngeal constriction in consonants (e.g., Wallin, 1855), but also for ejection—i.e., [constricted glottis] ([CG])—in cognate consonants (Gasparini, 2021). In point of fact, ejective emphatics are reconstructed for Proto-Afroasiatic (Diakonoff, 1984; Ehret, 1995; Orel and Stolbova, 1995; Bomhard, 2008) as well as for Proto-Semitic (Martinet, 1953; Ullendorff, 1955, p. 155; Cantineau, 1960; Knudsen, 1969; Dolgopolsky, 1977; Roman, 1981; Zemánek, 1996; Fallon, 2002, p. 102; Bellem, 2007; Watson and Bellem, 2011, p. 239; Kogan, 2012, p. 61; Bellem and Watson, 2014; Huehnergard, 2019, p. 49–50, 2023, p. 141–142; Pat-El, 2019, p. 81). By contrast, Arabic-style emphatics are new-fashioned, relatively speaking (Zemánek, 1996; Kogan, 2012; Huehnergard, 2017, p. 18): "In Arabic, ... an important phonological development is the change of the "emphatic" consonants from glottalic to pharyngealized or uvularized, as in [s'] > [$s^ʕ$]" (Huehnergard and Pat-El, 2019, p. 11).

The terminological and historical coupling of ejection with pharyngealization and uvularization makes sense, phonetically: "The ejective is produced with a closed glottis, air being expelled through the constriction by raising the glottis and *narrowing the pharynx*, thereby creating an increased pressure in the mouth" (Halle and Stevens, 1971, p. 208; italics added). Indeed, Kingston (1985) measured intraoral pressure (Po) during the production of ejectives in the Ethiopian Semitic language Tigrinya and determined that larynx raising is insufficient to create the extreme intraoral pressure involved in the production of ejective stops in particular. He concluded:

> Other maneuvers which would contract the cavity, such as *retracting the tongue root*, together with an increase in the stiffness of the walls of the vocal tract to reduce passive

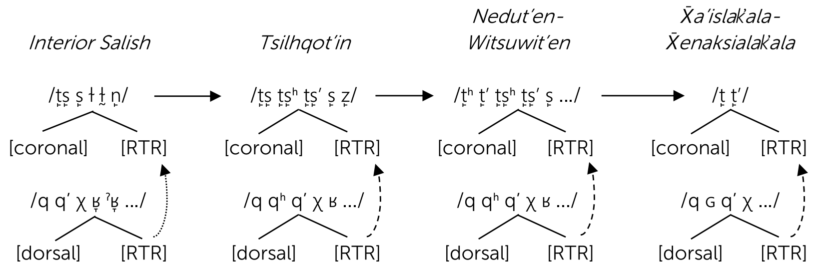


FIGURE 3
Emphatic consonants originated by assimilating [RTR] from uvulars in Interior Salish words (dotted curved-lined arrow) and spread to neighboring languages (solid straight-lined arrows) by redeploying [RTR] from uvulars in these languages (dashed curved-lined arrows).

> expansion in response to increasing Po, must also be employed if Po is to be elevated as high as it typically is in the articulation of an ejective. (p. 385; italics added)

The same point is made in Demolin's (2002) study of ejectives in another Ethiopian Semitic language: "The Amharic data suggest that additional maneuvers must be employed, such as *retracting the tongue root* or extending the magnitude of the contact in the oral cavity" (p. 470; italics added).

In other words, pharyngeal constriction and tongue root retraction are enhancement gestures for ejectives (cf. Stevens and Keyser, 1989, 2010; Keyser and Stevens, 2001, 2006; Stevens, 2002, 2005). Though enhancement properly belongs to the phonetic component of grammar, it is recognized that "enhancement gestures can become phonologized" (Keyser and Stevens, 2006, p. 61).[18] In fact, most phonologizations derive from enhancement gestures (Hyman, 2008). In the case at hand, the tongue-root retraction that accompanied ejectives in proto-Central Semitic became a proxy for the feature-defining gesture of [CG] in Proto-Arabic (for details on how phonetic proxies work, see Keyser and Stevens, 2006; Flynn, 2011; Flynn and Fulop, 2014). This proxy relation was phonologized early on, such that [RTR] replaced [CG] in nearly all forms of Arabic.[19] Specifically, [CG] */t' (t)θ' (t)s' (t)ɬ' k'/ became [RTR] /ṭ ð̣ ṣ ḍ q/ (cf. Kogan, 2012).

The phonologization of tongue root retraction in Arabic emphatics may have been a true innovation if the feature [RTR] did not previously exist, that is, if Proto-Central Semitic had no uvulars or uvularization, nor even retracted vowels which might be considered [RTR] (cf. Zemánek, 1996; Kogan, 2012; Huehnergard, 2017, p. 18). This may be the case—unlike contact phenomena, ordinary internal sound shift is not constrained by phonological redeployment (see, e.g., Blevins, 2004). In practice, however, it is difficult to prove the prior absence of uvulars—Proto-Afroasiatic had uvular stops and fricatives according to Orel and Stolbova (1995), and Proto-Semitic had "velar / uvular" stops and fricatives (Huehnergard, 2019, p. 50). In particular, the reflexes of Proto-Semitic dorsal fricatives are uvular in most languages (Huehnergard, 2019, p. 51) and pharyngeal in others (Huehnergard, 2019), which suggests that these sounds may have been uvular from the beginning. As mentioned in Section 2.1, "the typological surveys of Simpson (2003) and Kümmel (2007) show that uvulars frequently become pharyngeals but pharyngeals don't often become uvulars" (Weiss, 2015, p. 135). "As documented by Simpson (2003), uvular to pharyngeal shifts are well documented in every branch of Semitic" (Blevins, 2004, p. 198).

Proto-Central Semitic also had /ʕ ħ/, so it is tempting to think that emphatics were created instead by combining coronal consonants with pharyngeals in Arabic. However, this would imply a secondary constriction in the lower pharynx and epilarynx, whereas Arabic emphatics are well-documented with a constriction in the upper pharynx at or just below the uvula (Ali and Daniloff, 1972; Dolgopolsky, 1977; Ghazeli, 1977; Czaykowska-Higgins, 1987, p. 2; Shahin, 1997, 2002, 2011; Zawaydeh, 1998, 2003; Al-Tamimi et al., 2009, p. 612–613; Jongman et al., 2011; Zawaydeh and de Jong, 2011; Israel et al., 2012; Al-Tairi et al., 2016, 2017; Al-Solami, 2017; Al-Tairi, 2018; Freeman, 2019; Alfaifi et al., 2020; Moisik et al., 2021, p. 26; Al-Ansari and Kulikov, 2023; Kulikov et al., 2023, p. 466). Because "the 'emphatics' are pronounced as uvularized consonants," Dolgopolsky (1977, p. 1) argued that they ought to be transcribed as /tʶ dʶ sʶ/ instead of /tˤ dˤ sˤ .../.[20] McCarthy (1994), the most widely cited publication on the topic, is bullish on this point:

> Despite differences in details, the overall picture is consistent: the emphatics and *q* have a constriction in the upper pharynx similar to that of the uvular gutturals *χ* and *ʁ*. Although there are suggestions (Keating, 1988) that Arabic dialects differ in the location of the secondary constriction of emphatics (with some showing a low, *ʕ*-like constriction), this does not seem to be true; all studies, now encompassing several different dialect

18 Examples from English include [round] in /u/ (Keyser and Stevens, 2006, p. 38–40) and [strident] in /tʃ, dʒ/ (Clements, 2009, p. 50).

19 Exceptions that prove the rule include the Zabid dialect of Yemeni Arabic, where *q* can still be heard as [k'] (Naïm, 2008). Interestingly, Nakao (2022) suggests that [CG] may persist alongside [RTR] in Arabic dialects, pointing to reports of preglottalization and/or implosion in emphatics in isolated varieties of Arabic spoken in Algeria, Morocco, Palestine, and Egypt. Nakao argues that the glottalization effects which accompany emphatics in these cases are not the result of language contact.

20 Unfortunately for Dolgopolsky (1977), "[t]he symbol [ʶ] ... to mark uvularization ... has not been made part of the IPA alphabet" (Anonby, 2020, p. 297, fn. 7).

> areas, find that the emphatics have a constriction in the upper pharynx. The so-called pharyngealized consonants of Arabic should really be called uvularized. (p. 218–219)

That "Arabic emphatics are uvularized" (Al-Tairi et al., 2016, p. 1) does not exclude the possibility that these consonants may additionally involve the same epiglotto-pharyngeal constriction as pharyngeals in certain varieties (Wallin, 1855, p. 612; Laufer and Baer, 1988; Al-Tamimi and Heselwood, 2011; Hassan and Esling, 2011; Al-Tamimi, 2017). However, recall from Section 2.1 that "the tongue body is not back but front with the Arabic pharyngeals, as we can see by the adjacent front allophone of the low vowel: compare pharyngeal [ħæːl] 'condition' with uvular [χɑːl] 'maternal uncle'" (McCarthy, 1994, p. 197). Crucially, Arabic emphatics pattern with uvulars rather than with pharyngeals in this regard (Herzallah, 1990, p. 29, 59; McCarthy, 1994, p. 220; Rose, 1996, p. 87; Shahin, 2002, 2011, p. 612, 615–616; Watson, 2002, p. 272; Moisik, 2013, p. 484). This suggests that Arabic emphatics (and uvulars) are not simply specified with the same phonological feature as pharyngeals, say [low] (see Section 2.1; cf. Lass, 1984, p. 87–88; Odden, 2013, p. 54, 60).[21]

The simplest solution is to assume that uvularized emphatics and uvulars uniquely share a different phonological feature, viz. [RTR] (Czaykowska-Higgins, 1987; Goad, 1991; Davis, 1993, 1995, p. 471–472; Mahadin and Bader, 1995; Zawaydeh, 1998; Watson, 1999; Halle et al., 2000, p. 425–426, 408, fn. 429; Ananian and Nevins, 2001; Hansson, 2010, p. 141–142, 161, 198; Slimani, 2018; Al-Bataineh, 2019; Al-Raba'a and Davis, 2020, p. 22ff; Alwabari, 2020; Jaradat, 2020; Al-Taisan, 2022; Habib, 2022, p. 16; Gebski, 2023). As the preceding citations illustrate, "the feature [RTR]... is the most agreed upon feature for emphatics in the literature" (Alwabari, 2020, p. 75). Likewise: "[+RTR] is the widely used feature specification for pharyngealization at least in Arabic that reflects the activity of the retraction of the root of the tongue" (Al-Tamimi, 2017, p. 29). Note that [RTR] is a phonological feature, so it disappears in the gesture-calculations component of the phonetics, where the feature-defining tongue-retraction gesture is accompanied by robust enhancement gestures, including pharynx constriction (cf. Stevens and Keyser, 1989, 2010; Keyser and Stevens, 1994, 2001, 2006; Stevens, 2002; Flynn, 2011, and references therein). Thus, Davis (1995, p. 471) describes "the feature [RTR] … as entailing a constriction in the upper pharynx," after Czaykowska-Higgins (1987) and Goad (1991).[22]

Finally, Rose (1996) and Shahin (2002) claim that Arabic pharyngeals are [RTR], too. This claim obscures the fact that pharyngeals cause low vowels to be slightly fronted, and non-low vowels to be lowered, but not necessarily retracted (see Section 2.1). It also obscures the fact that pharyngeals show [RTR] allophones in words with emphatics or uvulars (Card, 1983, p. 16; Anonby, 2020, p. 281). Moreover, "that emphatics, uvulars and pharyngeals share the feature [RTR] … is phonologically problematic because it does not account for the free occurrences of emphatics and pharyngeals in Arabic" (Al-Tairi, 2018, p. 65; cf. p. 33–39 on co-occurrence restrictions affecting gutturals in roots). Davis (1995) shows that Palestinian Arabic has a long-distance regressive dissimilation rule of "depharyngealization" whereby "the first of the two consonants containing [RTR] loses that feature" (p. 481). He illustrates the application of this rule to underlying emphatics and uvulars, e.g., /ṣadaqa/ "charity" (p. 482), /χabaṣ/ "mixed randomly" (p. 483), /ṣabaʁ/ "he dyed" (p. 480). Crucially, pharyngeals do not participate in this process:

> The phenomenon of depharyngealization … strongly supports the view that uvulars and emphatics are characterized by a common underlying [RTR] feature. This view has been argued for previously by Czaykowska-Higgins (1987) and Goad (1991). Moreover, it is revealing that the depharyngealization phenomenon is not triggered by the occurrence of a primary pharyngeal [ħ] or [ʕ] in the root, but only by a uvular. This supports Goad's (1991) specific proposal that primary pharyngeals do not have the feature [RTR] underlyingly. (Davis, 1995, p. 483)

Ironically, the next sections present language-contact cases in which Arabic sounds are borrowed by redeploying the "wrong" features—[low] for emphatics (Sections 3.2, 3.5), and [RTR] for pharyngeals (Section 3.3).

3.2 Phonological emphasis via ejectives: consonantal [low] in Semitic and beyond

As explained above, pharynx constriction works with larynx raising to pressurize the trapped air in ejectives. Arabic phonologized the upper pharyngeal constriction of ejectives as [RTR], creating retracted coronals /ṭ ṣ .../ and uvular /q/ from earlier ejectives */t' ts' ... k'/. This secondary pharyngeal constriction then spread as an areal feature across Northwest Semitic languages (Hebrew, Aramaic, and Phoenician) via their glottalized consonants (Huehnergard and Rubin, 2012, p. 269). It also spread—again,

21 Hoberman (1988), Halle (1989, p. 18), and Kenstowicz and Louriz (2009) assume [constricted pharynx] for both emphatics and pharyngeals. Prince (1975, p. 12) takes "[+low] (perhaps better is [+C.P.]) as the feature shared by /ṭ ṣ q/" in Tiberian Hebrew. The feature [constricted pharynx] is discussed further below.

22 Likewise, Napiorkowska (2021) treats emphatics as [RTR], defined as "constriction of the upper pharynx" (p. 326, fn. 8). As such, [RTR] is roughly equivalent to other features proposed for Arabic emphatics, such as [rhizo-lingual] (Brame, 1970, p. 15–17), [upper pharynx] (Czaykowska-Higgins, 1987, p. 13; Hess, 1998, p. 268–271), [constricted tongue root] (Stevens, 1998, p. 251–254), [retracted tongue back] (Zawaydeh, 1999), [retracted] (Sylak-Glassman, 2014, p. 137), and [−ATR] (Gasparini, 2021, p. 17–18; Archangeli and Pulleyblank, 1994, p. 20–21). Goad (1989) and Elorrieta Puente (1991) argue that emphatics and uvulars are [RTR], but warn that uvulars may not be [RTR] in all languages (see also Trigo, 1991, p. 122). Vaux (1994, p. 251–256) and Bin-Muqbil (2006) claim that emphatics are [+RTR], whereas uvulars are [−ATR, −RTR]. Purnell and Raimy (2015, p. 526) treat pharyngealization as [RTR], but uvulars as [back].

via ejectives—to more distantly related languages within Semitic and Afroasiatic more generally. For example, "the pharyngealized articulation of Berber emphatics is ascribed to the influence of Arabic" (Zemánek, 1996, p. 18).

Remarkably, the influence of Arabic on ejectives in neighboring languages can be observed even in our present time. The North Ethiopic language Tigre (Palmer, 1956) and the Modern South Arabian language Ḥarsusi (Al Bulushi, 2019) are apparently partway through the shift from ejection to secondary pharyngeal constriction under the influence of Arabic. More specifically, Rose (1996, p. 92–97) and Bulakh and Kogan (2011, p. 7–8) claim that [CG] ejectives in Tigre and Ḥarsusi have become [RTR] under the influence of Arabic emphatics:[23]

> To the best of my knowledge, Tigre and Harsusi, a Modern South Arabian language (Johnstone, 1977), are the only two languages in which ejectives lower vowels. My solution to the lowering facts requires positing an [RTR] feature on ejectives, yet [RTR] normally defines emphatics and not ejectives. … Interestingly, the two languages which do show retraction next to ejectives have considerable contact with Arabic and could plausibly be influenced by the behavior of emphatics in Arabic. This is supported by Fre Woldu's (1986) study, in which he shows that perceptually, Tigrinya ejectives are judged by Sudanese Arabic speakers to be almost indistinguishable from emphatics. (Rose, 1996, p. 94)

Of special interest is that Tigre previously had pharyngeal /ʕ ħ/, but no uvulars. On Rose's (1996) assumption that pharyngeals are [RTR] (Section 3.1), she could claim that [RTR] was redeployed from pharyngeals to [CG] ejectives, making them pharyngealized. However, if pharyngeals are not [RTR] (see Sections 2.1 and 3.1 and references therein), then Tigre could not have redeployed [RTR] in this way, because there is no precedent for [RTR] in the language—Tigre had no [RTR] uvulars, as just mentioned, nor is there clear evidence of [RTR] contrasts in Tigre vowels. There is, therefore, only one possibility for redeployment: the [RTR] secondary articulation of Arabic must have been borrowed as a [low] secondary articulation in Tigre, by redeploying the marked consonantal use of [low] in Tigre pharyngeals (Section 2.1). That is, the pharyngeal constriction which enhances the feature-defining gesture of [CG] in ejectives (Halle and Stevens, 1971, p. 208) was phonologized in Tigre ejectives as [low].

Keyser and Stevens (1994) define [low] as lowering the tongue body (p. 231), but they remark that constricting the lower pharynx serves to enhance the acoustic manifestation of [low] (p. 226), and further, that an enhancement gesture like pharyngeal constriction is more reliable than a feature-defining gesture like tongue-body lowering—"unlike feature-defining gestures, enhancement gestures are never subject to overlap severe enough to mask their acoustic consequences" (Keyser and Stevens, 2006, p. 57–58). Chomsky and Halle (1968) first suggested that [low] could be used for consonants with secondary pharyngealization because "the superimposed articulation... in pharyngealization is [a]-like" (p. 305).

As it happens, there is compelling evidence for the redeployment of [low] from pharyngeals to ejectives in Tigre. As Rose (1996) concedes, "Lowenstamm and Prunet argue that the feature [+low] … is prosodically spread, from syllable to syllable" (p. 93)—"c'est le noeud [low] qui se propage" (Lowenstamm and Prunet, 1988, p. 23). Faust (2017, p. 3) has described "Tigre Lowness Harmony" more recently as follows:

> Tigre displays... five phonetically-stable vowels [i, u, e, o, a], and one phonetically-unstable one, of generally low quality, realized as [ə, ɛ, ʌ, a] depending on the context.... [A]s noted by Palmer (1956), the quality of that vowel is [a], rather than one of the higher qualities, if one of three conditions holds:
>
> i. A stable vowel [aː] follows anywhere in the word, and no other stable vowel interferes.
> ii. The onset of its syllable is an ejective [t', k', ts', ʧ'] or a pharyngeal [ħ, ʕ] consonant.
> iii. One of these consonants follows anywhere in the word.

That ejectives are specified [low] is supported by the fact that they pattern with phonetically-stable [a] and with the pharyngeals [ħ ʕ] in triggering an [a]-allophone of the phonetically-unstable vowel. The latter vowel is analyzed by Palmer (1956, p. 565) as "a short half open central vowel" /ɐ/ underlyingly. Crucially, Palmer is explicit that the [low] allophone of /ɐ/ is "a short open *front* vowel" [a] (Palmer, 1956; italics added). He indicates that a "retracted" vowel allophone is triggered by /u, w/, and /k, g/ (p. 567–568), but the pharyngeals and ejectives cause no special retraction on /ɐ/; they only cause it to be more open and more front. This indicates that pharyngeals and ejectives are specified [low] in Tigre, as suggested by Lowenstamm and Prunet (1988, p. 23–25), and not [RTR], contra Rose (1996, p. 92–97), and Bulakh and Kogan (2011, p. 7–8).

Tellingly, non-ejective /ʔ/ is also variably pharyngealized in Tigre. As Moisik et al. (2012) describe,

> Tigre (Semitic) has an optional process that neutralizes the contrast between /ʔ/ and /ʕ/ in the presence of pharyngeals and ejectives anywhere else in the word. For example, /ʔaddaħa/ "noon" is variably realized as [ʕaddaħa] or [ʔaddaħa] (Raz, 1983, p. 5; see also McCarthy, 1994). Critically, /h/ and /ħ/ do not show neutralization under the same conditions. (p. 11)

That is, [CG] /ʔ/ becomes [ʕ] by assimilating the marked consonantal use of [low] from a pharyngeal or an ejective in the same word. This is a variable phonological process, but it is similar to the redeployment strategy in Tigre diachrony: the [CG] ejectives "assimilated" the marked consonantal use of [low] in pharyngeals. The fact that /ʔ/ synchronically (if variably) assimilates [low] from a pharyngeal or ejective, but not from /a/, finds a parallel in redeployment, too: [low] emphatic consonants were created by redeploying the [low] feature of pharyngeal consonants, not by redeploying the [low] feature of the vowel /a/. As discussed by Martinez et al. (2023, p. 390), "redeployment within systems" (e.g., [low] within the consonant system) is privileged over "redeployment

23 I will argue shortly that Tigre ejectives are not [RTR], but [low].

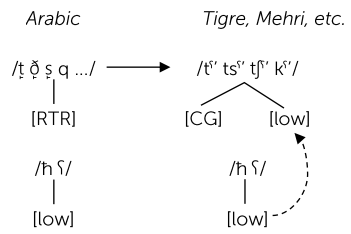


FIGURE 4
Emphatic consonants are [RTR] in Arabic, unlike its pharyngeals, which are [low], but emphatic sounds have been borrowed as [low] in Tigre and Mehri, among others (solid straight-lined arrows) by redeploying [low] from pharyngeals in these languages (dashed curved-lined arrow).

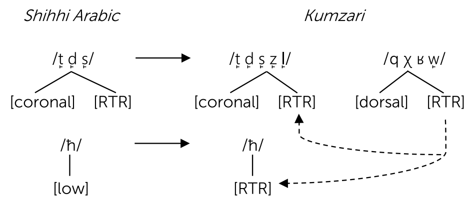


FIGURE 5
/ħ/ is [low] in Shihhi Arabic, unlike the coronal emphatics, which are [RTR]. Kumzari has acquired both /ħ/ and coronal emphatics as [RTR] (solid straight-lined arrows), by redeploying [RTR] from its uvulars (dashed curved-lined arrows).

across systems" (e.g., [low] from the vowel system to the consonant system).[24]

To broaden the discussion, Tigre is just one of many Afroasiatic languages that have changed their [CG] ejectives to emphatics under the areal influence of Arabic (Hebrew, Aramaic, Phoenician, Berber, etc.). As already mentioned, this change can be observed in progress in other present-day languages, such as Modern South Arabian languages Mehri (Watson and Bellem, 2011; Naïm and Watson, 2013; Watson and Heselwood, 2016; Ridouane and Gendrot, 2017), Ḥarsusi (Johnstone, 1977; Al Bulushi, 2019), and Soqotri (Kogan and Bulakh, 2019, p. 283). As mentioned above, Proto-Afroasiatic (Diakonoff, 1984; Bomhard, 2008) and Proto-Semitic (Kogan, 2012; Huehnergard and Pat-El, 2019) are reconstructed with [low] pharyngeals, but not necessarily with [RTR] uvulars (cf. Huehnergard, 2019, p. 50), so it is likely that some of the languages mentioned above adopted phonological emphasis like Tigre, by redeploying the marked consonantal use of [low] in preexisting pharyngeals, as diagrammed in Figure 4. To give just one potential example, Ridouane and Gendrot (2017) report the following for Mehri as spoken in Salalah, Oman:[25]

> Ejectives were shown to pattern together with uvulars and pharyngeals as a natural class defined by the feature [+low]. One very important characteristic of this class of segments is that it systematically triggers the diphthongization of following long high vowels /iː/ and /uː/ to [aj] and [aw], respectively, and the lowering of long /eː/ into /aː/. (p. 142)

3.3 Pharyngeals via uvulars: [RTR] in Kumzari

In "Emphatic consonants beyond Arabic," Anonby (2020) reports on Kumzari, an endangered Indo-European language spoken mainly in Oman. Kumzari has uvular obstruents, viz. /q χ ʁ/, which is not uncommon in (Southwestern) Iranian languages, but it also has a new series of alveolar emphatics /ṭ ḍ ṣ ẓ ḷ/ and a new pharyngeal fricative /ħ/, due to the influence of Arabic. On the redeployment view, Kumzari speakers must have created the emphatics and pharyngeal by redeploying the [RTR] feature of their historical uvulars, as diagrammed in Figure 5.[26]

This predicts that the emphatics are uvularized as [RTR] (see Sections 2 and 3.1), rather than pharyngealized as [low] (see Section 3.2). More daringly, it also predicts that the pharyngeal is [RTR], like the emphatics. Both of these predictions appear to be confirmed by Anonby (2020):

> Uvularization is the main articulatory basis for emphasis in Kumzari. The alveolar emphatics ṭ ḍ ṣ ẓ ḷ exhibit strong, simultaneous posterior secondary articulation, with uvularization dominating but bounded by a unified stricture all the way from the pharynx up to the velum. … The remaining member of the Kumzari emphatic series, however, is a pharyngeal consonant *ḥ* [ħ]. Although *ḥ* is not uvularized, its behavior suggests that it should be classed as an emphatic. (p. 297)
>
> The uvularized alveolar emphatics, uvular consonants *x q ġ*, the pharyngeal *ḥ*, and the uvularized allophone of *r* all cause preceding as well following non-back vowels to be retracted (*ā* [aː] → [ɑː], *a* [ɐ] → [ʌ]). In the case of non-low vowels, they cause lowering in the transition between the

24 See Nelson (2023) for a possible case of redeployment across systems in adult language acquisition.

25 The reverse influence is rare, but al-Kathīrī (2019) reports on a variety of Oman Arabic that has changed its [RTR] /ṭ/ and /q/ to [CG] /t͡ʃʷ'/ and /k'/, respectively, under the influence of neighboring Modern South Arabian languages with ejectives. Crucially, Arabic has long lost its historical emphatics, but it has preserved /ʔ/. Evidently, the feature [CG] was redeployed from /ʔ/ to /ṭ q/ in this variety of Oman Arabic, creating the new ejectives /t͡ʃʷ' k'/ under the influence of surrounding Modern South Arabian languages.

26 As Kahn (1976) says of closely-related Persian: "whereas Persian does not have any pharyngealized or pharyngeal consonants, it does have a post-velar stop/approximant /q/" (p. 27). Proto-Iranian is not reconstructed with a phonemic distinction between velars and uvulars (Skjærvø, 2009, p. 51), but its lone dorsal fricative was probably uvular *χ (Cantera, 2017, p. 482), such that most branches have uvular fricatives and many eventually developed uvular stops (see Bashir, 2009; Edelman and Dodykhudoeva, 2009; Windfuhr and Perry, 2009, etc.).

> vowel and consonant (*i* [iː] → [iə] before a consonant, [əi] after a consonant; *ē* [eː] → [eə] before a consonant, [əe] after a consonant). (p. 298)

The uvularized nature of Kumzari emphatics is not surprising, given that Arabic has uvularized emphatics, too (Anonby, 2020, p. 282; see Section 3.1 above). The fact that Kumzari /ħ/ retracts vowels the same as emphatics and uvulars is more significant. Recall that "the tongue body is not back but front with the Arabic pharyngeals, as we can see by the adjacent front allophone of the low vowel: compare pharyngeal [ħæːl] 'condition' with uvular [χɑːl] 'maternal uncle'" (McCarthy, 1994, p. 197). This fronting effect is not as pronounced in the Shihhi Arabic that surrounds Kumzari (cf. Anonby, 2020, p. 301–302), but according to Bernabela (2011) it remains the case that /ħ/ (the only pharyngeal in Shihhi Arabic) does not cause retraction in low vowels, e.g., *yiftaḥ* [ˈjɪftaħ] "he opens" (p. 29), *ḥasan* [ˈħasæn] "Hasan (proper name)" (p. 30); *maḥḥ* [maħ] "with her" (p. 93, fn. 164), *ḥafiz* [ˈħaːfiz] "(shop)keeper" (p. 94, fn. 166). By contrast, emphatics and uvulars cause low-vowel retraction in Shihhi Arabic:

> In the vicinity of one of the velarised consonants *ṣ*, *ṭ* or *ḍ*, *a* is usually backed [ɑ] and velarised: *qōṣar* [ˈqɔːsɑɫ] "need"; *manṭaqih* [ˈmɛntɑqiʰ] "area"; *manḍarih* [ˈmɛnđɑɽiʰ] "mirror". The uvulars *q* and *x* have the same backing effect: *qarnēn* [qɑɽˈneːn] "two horns"; *xallnu* [ˈxɑlnʊ] "let us." (Bernabela, 2011, p. 30)

Anonby (2020, p. 309) explains that Kumzari speakers created many emphatics by "diffusing" phonological emphasis from uvular obstruents (Ar. *qyās* > K. *qyāṣ* "measurement;" Middle Persian (MP) *suxr* > *ṣirx* "red;" etc.) and from /w/, which he therefore analyzes as labio-uvular (Ar. *walā* > *waḷa* "or;" Middle Persian (MP) *sabz* > *sawz* > *ṣawz* "green;" etc.). Crucially, /ħ/ was created in the same way (e.g., Ar. *qahwa(t)* > K. *qaḥwē* "coffee") and, in turn, the new /ħ/ also "diffused" phonological emphasis (e.g., Ar. *sāḥir* "magician" > K. *ṣāḥar* "sorcerer"). This strongly suggests that in Kumzari the pharyngeal fricative shares the same phonological property as emphatics and uvulars, viz. [RTR].

Anonby agrees that Kumzari's historical uvulars played a key role in its adoption of Arabic emphatics:

> In Kumzari, an Indo-European language in close contact with Arabic,... a core set of alveolar emphatics is also found, but is characterized by uvularization as a dominant secondary articulation. In keeping with a uvular place of articulation, the consonants *x* ["voiceless uvular fricative" (p. 296)] and *q*, as well as uvular *w*, have a clear role in the historical diffusion of emphasis; and evidence for a historical spread of emphasis from pharyngeal *ḥ* is also found. (p. 322–312)

However, he hesitates to implicate [RTR] in Kumzari's adoption of pharyngeal /ħ/ and its involvement in the diffusion of phonological emphasis, because this phonological feature "is typically limited to emphatics with secondary articulations in synchronic accounts of emphasis" (p. 309). On the other hand, he concedes that pharyngeals may present [RTR] allophones in words with emphatics and uvulars in Arabic (p. 280), and he suggests that "in Cairo Arabic and Palestinian Arabic, there is even a contrast available between plain and emphatic pharyngeals" (p. 280–281), so in principle, nothing prevents Anonby from treating Kumzari /ħ/ as [RTR], like uvulars and alveolar emphatics.[27]

3.4 Phonological emphasis without redeployment in Northern Songhay?

The preceding section argued that Kumzari speakers acquired the [RTR] emphatics of their Arabic neighbors by redeploying the [RTR] feature of preexisting uvulars. By contrast, Section 3.2 argued that speakers of Tigre, which previously lacked uvulars, acquired the [RTR] emphatics of their Arabic neighbors as [low] instead, by redeploying the marked consonantal use of [low] in preexisting pharyngeals. Beyond cases like these, I have made a sincere effort to look for falsifying evidence—languages which have acquired emphatic consonants in language contact, with no previous uvulars or pharyngeals or any other type of consonant with a phonological feature that might be redeployed as phonological emphasis.

Coming closest are Northern Songhay languages spoken in Saharan oases across Algeria, Niger, and Mali: Korandje, Tasawaq, Tagdal, and Tadaksahak. These languages have each adopted a series of pharyngealized coronals under the areal influence of Berber and Arabic, in spite of Proto-Songhay having no uvulars or pharyngeals (Nicolaï, 1981; Souag, 2020, p. 646). However, Souag (2010) remarks that "Proto-Northern Songhay had probably already developed a phoneme q, judging by the pan-Northern sound change k > q /_o (Nicolaï, 1981)." As Nicolaï (1981) explains, the development of /q/ in Proto-Northern Songhay probably occurred under the areal influence of Tamasheq, a variety of Tuareg Berber, but this development was nonetheless an internal sound change, rooted in the difficulty of maintaining the Songhay phonemic contrast between /k/ and /kʷ/ before /o/, so "it remains possible that the shift in question occurred independently of language contact" (Nicolaï, 1981, p. 359). Souag (2012) describes the internal sound shift *k > q /_o as "a genuine shared innovation" (p. 184) in Northern Songhay and perhaps the strongest phonological evidence for this subgrouping within the larger family of languages. In short, it seems that Proto-Northern Songhay had /q/ before various descendants borrowed coronal emphatics from Berber and Arabic, so we can assume that the feature [RTR] was redeployed from their /q/ to facilitate this borrowing, as diagrammed in Figure 6.

3.5 Phonological emphasis via alveolopalatals: consonantal [low] in Northwest Caucasian

Section 3.2 described how a language like Tigre, which had [low] pharyngeals but no [RTR] uvulars, apparently borrowed the [RTR] emphatics of its Arabic neighbor as [low] instead. Section 3.3 described how a language like Kumzari, which had an [RTR] uvular but no [low] pharyngeal, apparently borrowed the [low]

27 Even laryngeals are [RTR] in certain languages, such as Nedut'en-Witsuwit'en (Section 2.3). Howe [Flynn] (1999a,b, 2000) argues that uvulars and laryngeals are both [RTR] in the Wakashan language Oowekyala.

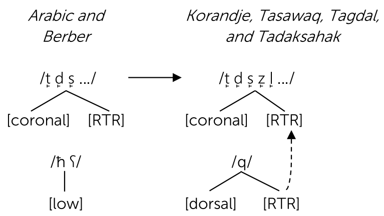


FIGURE 6
Northern Songhay languages have acquired the [RTR] emphatics of their Arabic and Berber neighbors (solid straight-lined arrow), by redeploying [RTR] from a preexisting uvular (dashed curved-lined arrows).

pharyngeal of its Arabic neighbor as [RTR] instead. Kumzari also borrowed Arabic emphatics as [RTR], by redeploying this feature from preexisting uvulars. Section 3.4 suggested that emphatics were similarly borrowed into several Northern Songhay languages. The present section describes a more equivocal case: the borrowing of Arabic emphatics into a Northwest Caucasian language which had a [low] pharyngeal as well as [RTR] uvulars.

For historical context, many Circassians were exiled from the Caucasus to the Ottoman Empire after the Russo-Circassian war (Natho, 2009). Notably, the Israeli village "Kfar Kama was established in 1878 by 1150 Circassian immigrants of the Shapsugh tribe and is located in the eastern Lower Galilee" (Reichel, 2010, p. 255). Natho (2009) remarks:

> About 3,000 Shapsughs now live prosperously in Kfar-Kama... The children are taught Arabic, Circassian, Hebrew, and English languages in their school. … Remarkable is the fact that the inhabitants of this village are purely Circassian, and that all of them old and young speak Circassian fluently. (Natho, 2009, p. 517–518)

The Shapsugh dialect of Adyghe spoken in Kfar Kama, Israel, has the following inventory of sounds:

(3) *Phoneme inventory in Israeli Shapsugh Adyghe* (adapted from Colarusso, 1988, p. 424).

p		t	ts		**tsˤʷ (∼ tɕˤʷ)**	tʃ		c ∼ kʲ	q	qʷ			
b		d	dz		**dzˤʷ (∼ dʑˤʷ)**	dʒ		ɟ ∼ gʲ					
p'		t'	ts'		**tsˤʷ' (∼ tɕˤʷ')**	tʃ'	tɬ'	c' ∼ kʲ'				ʔ	ʔʷ
f	fʷ		s	**sˤ (∼ ɕˤ)**	**sˤʷ (∼ ɕˤʷ)**		ɬ	ç ∼ xʲ	χ	χʷ	ħ	(h)	
			z	**zˤ (∼ ʑˤ)**	**zˤʷ (∼ ʑˤʷ)**		ɮ	ʝ ∼ ɣʲ	ʁ	ʁʷ			
			s'	**sˤ' (∼ ɕˤ')**	**sˤʷ' (∼ ɕˤʷ')**								
m		n									ã	ə̃	
	w	r						j			a	ə	

Of special interest are the emphatic sibilants shown in boldface font. These were documented in 1973 by Catford after spending 5 weeks working with Shapsugh speakers in Kfar Kama at the invitation of the Israeli Ministry of Education (Catford, 1984, p. 27). Catford's report is confirmed by Colarusso (1988, p. 22–23):

> Professor Catford has informed me that the younger members of the village of Kafr Kama in Israel, who speak a form of Shapsugh, have substituted pharyngealized alveolar spirants for the alveo-palatal series. Tapes kindly provided to me by Miss Wendy Orent of Boston University and Mr. Alexander Borg of Hebrew University confirm Catford's observation. This substitution may have been aided by the presence of pharyngealized coronals in the neighboring Arabic dialects. For Israeli Shapsegh the contrast between alveolars, pharyngealized alveolar, and rounded pharyngealized alveolars, as in /sa/ "knife," /sˤa/ "100," and /sˤʷa/ "skin, hide," present data which show that rounding and pharyngealization are not mutually exclusive.

As mentioned in this quote, the pharyngealized sibilants correspond to alveolopalatals in other Shapsugh dialects (Colarusso, 1988, p. 421–436). Perhaps for this reason, Catford transcribed the emphatic sibilants as pharyngealized alveolopalatals (p.c., Colarusso, 1988, p. 75, n. 7). "Some speakers may in fact have this articulation," Colarusso (1988) wrote, but he added: "The specimens which I have heard of Israeli Shapsugh... appear to have a lamino-alveolar articulation [+anterior, –high], with pharyngealization" (p. 75, n. 7). Wallis (1987), who conducted fieldwork in Kfar Kama in the 1970's, also recorded emphatic sibilants as alveolar, e.g., *sˤ'ə* "to make," *psˤasˤa* "girl" (p. 85).[28]

Colarusso (1988, p. 23) claimed that the pharyngealized sibilants in Israeli Shapsugh are specified [constricted pharynx], defined as "a narrowing of the *lower* pharynx" (Perkell, 1971, p. 124; italics added). Perkell (1971) argued for a total abandonment of [low] in favor of this new feature. Keating (1988) dismissed the proposed replacement as "a short-lived move" (p. 15), but it should be noted that Stuart Davis favors [constricted pharynx] over [low] to characterize pharyngeals in Arabic (Davis, 1993, 1995, p. 471; Abo Mokh and Davis, 2020, p. 40–41). Like other Circassian languages, Israeli Shapsugh has /ħ/, so it is possible that the relevant phonological structure was redeployed from this pharyngeal fricative to the alveolopalatal sibilants under the influence of pharyngealized coronals in Arabic. The structure in question could be the feature [constricted pharynx], as Colarusso suggests, or else the marked consonantal use of [low] in syllable margins.

There is too little information on Israeli Shapsugh to be confident that its pharyngealized sibilants /sˤ zˤ zˤ' sˤʷ zˤʷ sˤʷ' tsˤʷ dzˤʷ tsˤʷ'/ are phonologically [low], but it is significant that these sounds developed from earlier alveolopalatals /ɕ ʑ ɕ' ɕʷ ʑʷ ɕʷ' tɕʷ dʑʷ tɕʷ'/ and that some alveolopalatalization may persist (Colarusso, 1988, p. 75, n. 7). As mentioned earlier,

28 Wallis recorded the rounded emphatic in sˤʷəz "woman" as an "alveopalatal retroflexed sibilant" (p. 85), but evidently not all her consultants were from Kfar Kama: "Field work as the basis for this paper was done in the village of Kafr Kama, Israel, 1971–1979. More recent work has been done with speakers now living in the Circassian community in the Paterson, N.J. area of the U.S." (Wallis, 1987, p. 89, n. 1).

[low] pharyngeals cause non-low vowels to become lower, but not necessarily more retracted, and adjacent low vowels may even be slightly fronted, as in the Interior Salish language Nxaʔamxcín (Section 2.1) and in Arabic (Section 3.1). A fronting effect is also reported for pharyngeals in Northwest Causasian languages such as Abkhaz, Kabardian, and Tsakhur (Trubetzkoy, 1931; Catford, 1983; Colarusso, 1985, 1988, 1992, p. 31, 2013, p. 98–99; Sylak-Glassman, 2014, p. 72–73; Beguš, 2021, p. 716). For instance, Andersson et al. (2023) report that Cwyzhy Abkhaz has "slightly palatal" (p. 271) [round] sibilants which they transcribe as alveolopalatal [ɕᶣ ʑᶣ tɕᶣʰ dʑᶣ tɕᶣ'] (p. 269–270). Crucially, they find the same "front rounded secondary articulation" (p. 6) in [low, round] /ħʷ/, i.e., [ħᶣ]. This Abkhaz dialect lacks [low, round] /ʕʷ/, but Chirikba (2014) reports that in certain varieties, *ʕʷ has changed into /jʷ/, phonetically [ɥ] (see fn. 7 for an articulatory explanation).

More pointedly, certain Northwest Caucasian languages have consonants with secondary pharyngealization. Notably, "in Ubykh there is a series of pharyngealized labials, /pˤ/, /bˤ/, /pˤ'/, /mˤ/, /vˤ/, and /wˤ/, in addition to the two pharyngealized uvular series, plain /qˤ, qˤ', χˤ, ʁˤ/ and rounded /qˤʷ, qˤʷ', χˤʷ, ʁˤʷ/" (Colarusso, 1988, p. 48; see also Beguš, 2021, p. 700–701). Similar to pharyngeals, Caucasian emphatics are known to cause "slight front coloring" (Colarusso, 2013, p. 98) in adjacent vowels, an effect called "emphatic softening" (Trubetzkoy, 1931) or "emphatic palatalization" (Trubetzkoy, 1969, p. 131–132; Catford, 1983, 1992; Colarusso, 1985, p. 366, 1988, p. 26, 2013; Rose, 1996, p. 98; Comrie, 2005; Bellem, 2009, p. 98–99; Moisik, 2013; Sylak-Glassman, 2014, p. 71–72; Beguš, 2021, p. 715–716). As mentioned, such effects are less perplexing if the emphatic feature is [low], rather than [RTR]. Again, see fn. 7 for an articulatory explanation.

Moreover, Ubykh already has [RTR] /q q' χ ʁ/ and [round, RTR] /qʷ qʷ' χʷ ʁʷ/, so the pharyngealized counterparts must involve an additional feature, say [low, RTR] /qˤ, qˤ', χˤ, ʁˤ/ and [low, round, RTR] /qˤʷ, qˤʷ', χˤʷ, ʁˤʷ/. Pace Halle et al. (2000, p. 408–410) and Purnell and Raimy (2015, p. 526), among others, the velar-uvular contrast cannot be understood as [front]-[back] instead, freeing up [RTR] to characterize secondary pharyngealization in Ubykh uvulars. This is because the [front]-[back] dimension is contrastive not only among velars (/k g k'/ vs. /kʲ gʲ kʲ'/), but also among uvulars (/q q' χ ʁ/ vs. /qʲ qʲ' χʲ ʁʲ/; Colarusso, 1988, p. 438; Beguš, 2021, p. 700–701).

Tellingly, "there are no palatalized, pharyngealized uvulars" (Colarusso, 1988, p. 274). Thus, in Ubykh, the [front] (palatalized) uvulars do not contrast for [low] (pharyngealization), unlike the [round] (labialized) uvulars. Similarly, in Cwyzhy Abkhaz (Andersson et al., 2023), [front] and [round] are both contrastive across coronals (e.g., /ʃ ʃʲ ʃʷ/), velars (e.g., /k kʲ kʷ/), and uvulars (e.g., /χ χʲ χʷ/), but [front] is not contrastive in the [low] pharyngealized uvulars and pharyngeals; only [round] is: /χˤ' χˤʷ' ħ ħʷ/ (Colarusso, 1988, p. 268; Chirikba, 2014, p. 298). The lack of a [front] contrast among [low] consonants in Ubykh and Abkhaz is surely related to "emphatic palatalization," mentioned above. Under such palatalized-pharyngealized phonetic conditions, it is difficult to establish or maintain a [front] contrast among pharyngeals and pharyngealized consonants. It is challenge enough to distinguish plain /χ/, say, from [front] /χʲ/, [low] /χˤ/, [round] /χʷ/, and [low, round] /χˤʷ/ in Ubykh (Beguš, 2021, p. 700–701) and Abkhaz (Andersson et al., 2023, p. 268).

As mentioned, Colarusso (1988, 2013) entertains [low] for certain "adytal pharyngeals" /ħ ʕ/ in Caucasian languages, but he rejects the use of this feature for pharyngealized uvulars in Ubykh, because the tongue body is not always low in these sounds, so he adopts Perkell's (1971) [constricted pharynx] instead to represent pharyngealization. On the other hand, he suggests [advanced tongue root] instead of [front] for palatalized uvulars and velars (e.g., Colarusso, 1988, p. 438, 2013, p. 98). As discussed in Sections 2 and 3.1, the vast majority of theorists treat uvulars as [retracted tongue root] (Latimer, 1978; Czaykowska-Higgins, 1987; Cook, 1989, p. 139; Goad, 1989; Davis, 1993, 1995; Mahadin and Bader, 1995; Rose, 1996; Shahin, 1996, 1997, 2002, 2011; Zawaydeh, 1998; Watson, 1999; Halle et al., 2000, p. 425–426, 408, fn. 8; Rose and Walker, 2004, p. 484–485; Hansson, 2010, p. 141–142, 161, 198; Slimani, 2018; Al-Bataineh, 2019; Al-Raba'a and Davis, 2020, p. 22ff; Alwabari, 2020; Jaradat, 2020; Al-Taisan, 2022; Habib, 2022, p. 16; Alqahtani and Almoaily, 2023; Gebski, 2023; etc.). For instance, Halle et al. (2000) explicitly describe Arabic *q* as an emphatic with "consonantal [RTR]" and even suggest "a prohibition *[+RTR, +ATR]" (p. 408, fn. 9). The point is: it is somewhat inconsistent to avoid using [low] for pharyngealized uvulars in Ubykh while also using [ATR] for palatalized uvulars in the same language.[29]

In sum, using [low] rather than [RTR] helps to explain "the seemingly anomalous palatal or fronting bias of pharyngeals and pharyngealization, most famously embodied by the "emphatic palatalization" of Caucasian languages" (Moisik, 2013, p. 558). Critically, alveolopalatals are usually palatalized or [front] (see Section 1), so their pharyngealization in Israeli Shapsugh makes more sense in terms of [low] than [RTR], too. If so, speakers of Israeli Shapsugh may have borrowed Arabic emphatics like speakers of Tigre (Section 3.2), by redeploying the consonantal use of [low] in preexisting pharyngeals, as diagrammed in Figure 7.

4 Conclusion

> [S]econd-language acquisition and bilingualism provide us with methodological utilities to inspect sound patterns because patterns that emerge when sound systems meet are not only familiar to us from the native language of the speaker or listener, but are also reflective of the universal laws of

29 Likewise, Sylak-Glassman (2014) argues against the use of [low] for pharyngealized uvulars in Ubykh, because the tongue body is not necessarily low in these sounds. He suggests using a new feature [constricted epilarynx] instead. Critically, he does not blink at palatalized uvulars in the same language (p. 22, 26, 112–3). He suggests that these palatalized sounds are specified [+front, +retracted, +raised, +open] (p. 128, 137–8, 141, 145). His features [+raised] and [+open] are somewhat at cross purposes, but not nearly so much as the other features. The "forward movement of the tongue body" (p. 137) of [+front] in /qʲ qʲ' χʲ ʁʲ/ is directly opposed to the "retraction of the tongue body" (Sylak-Glassman, 2014) of [+retracted] and to the "backward" (Sylak-Glassman, 2014) tongue movement of [+raised]. The point here is not to criticize Sylak-Glassman's proposed features—contrastive palatalization in uvulars is bound to involve partly antagonistic gestures in any feature system. The point is: using [front] for palatalized uvulars in Ubykh is comparable to using [low] for pharyngealized uvulars in the same language.

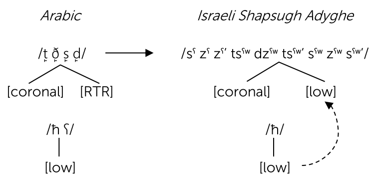


FIGURE 7
Emphatic consonants are [RTR] in Arabic, unlike its pharyngeals, which are [low], but emphatic sounds have been borrowed as [low] in Israeli Shapsugh (solid straight-lined arrow) by redeploying [low] from its pharyngeal (dashed curved-lined arrow).

> phonetics and human cognition. At the crossroads of unity and variation across the languages of the world, studying second-language sound patterns therefore gives us a unique window of opportunity to understand the nature of linguistic processes and representations as well as the extent of human grammars. All of these shape "patterns" that linguists are fond of because, after all, patterns are manifestations of how we get to know what we know. For one thing, second-language acquisition is expected to mimic linguistic change through language contact, albeit—and perhaps luckily—observable within an individual's life span. (Kabak, 2019, p. 250)

Spurred by reflections like Kabak's, I have applied the notion of redeployment in second language acquisition to contact-induced diachronic changes. Of particular interest are cases where a marked phonological contrast has spread across neighboring languages. Such cases suggest that listeners can re-weight and re-map phonetic cues onto novel phonological structures. On the redeployment view, cues can indeed be re-weighted, but phonological structures which underlie a new contrast are not expected to be fully novel; rather, they must be assembled from preexisting phonological structures (Archibald, 2003, 2005, 2009, 2018, 2021, 2022, 2023; Archibald et al., 2022).

Emphatics prove to be an instructive case. These typologically marked consonants were innovated in Interior Salish (Section 2.1) and Arabic (Section 3.1), and were then borrowed into neighboring (unrelated) languages. Most phonologists consider the original emphatics to be [RTR], like uvulars, "entailing a constriction in the upper pharynx" (Davis, 1995, p. 471), and the emphatics were evidently borrowed as such in many languages. In Tsilhqot'in (Section 2.2), Nedut'en-Witsuwit'en (Section 2.3), X̄a'islak̓ala-X̄enaksialak̓ala (Section 2.4), Kumzari (Section 3.3), and Northern Songhay languages (Section 3.4) among others, the feature [RTR] was redeployed from preexisting uvulars to other consonants on the basis of partial acoustic /auditory similarities with emphatic sounds in neighboring languages. Importantly, languages without uvulars (and without any other [RTR] consonant) did not and arguably could not participate in the areal spread of phonological emphasis. For example, Dakelh (Section 2.5) did not have uvulars, nor any other type of consonant with a phonological feature to redeploy as emphasis, so it has not adopted emphatic consonants in spite of prolonged contact with five languages with these sounds (Sections 2.1–2.4).

On the other hand, it was found that languages with pharyngeals may borrow emphatics differently (Sections 3.2, 3.5). Pharyngeal consonants entail a constriction in the epilarynx and lower pharynx, traditionally represented by the phonological feature [low]. This feature can apparently be used for secondary pharyngealization, too. For example, Tigre had no uvulars so the [RTR] emphatic consonants of Arabic were arguably borrowed as [low] instead, by redeploying [low] from preexisting pharyngeal consonants to [CG] ejectives (and to [CG] /ʔ/ in words with [low] consonants). To clarify: Tigre redeployed the phonological use of [low] in a consonant, not simply the feature [low], which presumably occurs in most spoken languages, notably in low vowels. Similarly, recall from Section 1 that some native speakers of Korean appear to learn English [posterior] /ʃ/ as [front] /ɕ/. What gets redeployed in that case is the phonological use of [front] in a sibilant, not simply the feature [front], which occurs in most languages, notably in front vowels.

A background assumption here is that redeploying a feature within the consonant system is easier than redeploying a feature from the vowel system to the consonant system. Take Soqotri (Kogan and Bulakh, 2019), one of several Modern South Arabian languages which have acquired emphatics under the influence of Arabic, as discussed in Section 3.2. Soqotri phonology has long distinguished laryngeals /h, ʔ/ from [low] pharyngeals /ħ, ʕ/, but it does not distinguish velars from [RTR] uvulars (Kogan and Bulakh, 2019, p. 283).[30] However, Soqotri phonology does distinguish /e, o/ from [RTR] /ɛ, ɔ/ (Kogan and Bulakh, 2019, p. 285–286).[31] Interestingly, Soqotri speakers apparently acquired the [RTR] emphatics of Arabic as [low] instead, by redeploying their use of [low] in pharyngeal consonants, rather than by redeploying their use of [RTR] in mid vowels. So for instance, /ɛ/ has two allophones according to Kogan and Naumkin (2014, p. 58): "open mid-front [ɛ]" and "open front [a] ('average European *a*');" "the first is the basic allophone appearing in neutral environments, the second is conditioned by the proximity of emphatics and pharyngeals" (Kogan and Naumkin, 2014). The fact that Soqotri emphatics cause vowel lowering to front [a], not back [ɑ], suggests that they are—like pharyngeals—[low] rather than [RTR].

Conversely, recall from Section 3.3 that /ħ/ is arguably [low] in Shihhi Arabic, but Kumzari speakers did not redeploy the feature [low] from their vowel system to acquire /ħ/ from Shihhi Arabic. Rather, Kumzari speakers redeployed the feature [RTR] from their preexisting uvulars to acquire /ħ/ as [RTR] instead. As Anonby (2020, p. 297) writes: "Although *ḥ* is not uvularized, its behavior suggests that it should be classed as an emphatic" (p. 297). So for instance: "The uvularized alveolar emphatics, uvular consonants *x q g*, the pharyngeal *ḥ*, and the uvularized allophone of *r* all cause preceding as well following non-back vowels to be retracted (ā [aː] → [ɑː], a [ɐ] → [ʌ])" (p. 298). By contrast, in Shihhi Arabic [low] /ħ/ does not have the same retraction effect on vowels as [RTR] emphatic consonants do (Bernabela, 2011, p. 30).

Critically, most languages distinguish several height levels in vowels, such as /a/ vs. /e, o/ vs. /i, u/, and many languages further distinguish /ɛ, ɔ/, so the phonological features [low] and [RTR] are frequently active in vowel systems. By contrast, these

30 Soqotri <q> is /kˤ'/, i.e., [stop, dorsal, low, constricted glottis].

31 The /o–ɔ/ contrast has a low functional load; e.g., "*hɔ* as form of address vs. *ho* 'I'" (Kogan and Bulakh, 2019, p. 283).

phonological features are relatively rare in consonant systems. The overall impression is that in contact situations, a language may newly adopt emphatics or pharyngeals as [RTR] or [low] only if the relevant feature is already in use in its consonant system. This supports the redeployment construct in second language acquisition theory (Archibald, 2003 *et seq*.). It also dovetails with the discussion in Martinez et al. (2023, p. 390): "redeployment within systems" (e.g., [RTR] or [low] within the consonant system) is privileged over "redeployment across systems" (e.g., [RTR] or [low] from the vowel system to the consonant system).

5 Envoi on consonantal [low]

Finally, it must be acknowledged that using [low] to represent epiglotto-pharyngeal constriction in consonants is disputable, albeit traditional (Chomsky and Halle, 1968, p. 305; Ladefoged, 1971, p. 92–94; Lass and Anderson, 1975, p. 18; Prince, 1975, p. 12; Rood, 1975, p. 329–333; Halle, 1983; Halle and Clements, 1983; Cole, 1991, p. 25; Coleman, 1998, p. 69; Jensen, 2004, p. 97; Calabrese, 2005, p. 59–60; Hayes, 2009, p. 87–88; Miller, 2011, p. 434; Flynn, 2012, p. 142–144; Odden, 2013, p. 54, 60; among many others). This distinctive feature was originally intended to be relatively abstract and implementable in both vowels and consonants with various articulators in the phonetics (hyoglossus muscles, jaw lowering, larynx raising, etc.). In practice, however, [low] is often narrowly defined as "a lowered tongue body" (Sagey, 1986, p. 278).

In Feature Geometry, too, various possibilities were originally contemplated for [low] in the tree—it might be located directly under the Place node (Clements, 1985), or under a Height node (Hyman, 1988, p. 269; Odden, 1991, p. 265; Lahiri, 2018, p. 234), or a Vowel place node (Goad, 1991), or a Pharyngeal node (McCarthy, 1988, p. 105). However, most assume that [low] is located under a Dorsal node or under a [dorsal] feature (Sagey, 1986, p. 61; Steriade, 1987, p. 597; Keyser and Stevens, 1994, p. 231; Halle, 1995; Avery and Idsardi, 2001, p. 68; Hall, 2007, p. 313), or else under a Tongue Body node alongside [dorsal] (Halle et al., 2000). This narrow conception of [low] is ill-suited to represent pharyngeals and pharyngealization according to some theorists (McCarthy, 1991, p. 43; see also Lee, 1995, p. 343).

As mentioned in Section 3.5, Perkell (1971) proposed to replace [low] with [constricted pharynx], defined as "a narrowing of the lower pharynx" (p. 124). Alternative replacements include [lower pharynx] (Czaykowska-Higgins, 1987, p. 13), [laryngopharynx] (Hess, 1998, p. 268–271), [constricted epilaryngeal tube] (Moisik and Esling, 2011; Moisik et al., 2012; Al-Tamimi, 2017; Al-Tairi, 2018; Esling et al., 2019), and [constricted epilarynx] (Sylak-Glassman, 2014). These various features were introduced to model the phonetic realities of pharyngealization more accurately than [low]. But phonological features were never intended to be used directly in the phonetics:

> [W]hile the input to the gesture-calculations component is a phonological representation, the output is not. Rather, the output is a series of instructions to the musculature. This entails that the phonological representation disappears at this point, being replaced by motor instructions. Hence, if the birthplace of lexical representation is in the lexicon, its demise is in the gesture-calculations component. (Keyser and Stevens, 2006, p. 36)

Moreover, even theorists who reduce distinctive features to particular defining gestures still place greater importance on other accompanying gestures in the phonetics. As already mentioned, Keyser and Stevens (1994) define [low] as a tongue-body feature and locate it as such in their feature tree (p. 231), but they remark that constricting the lower pharynx serves to enhance the acoustic manifestation of [low] (p. 226). Crucially, enhancement gestures like pharyngeal constriction are introduced in the phonetics, not in the phonology, and as such, these gestures prove to be more reliable phonetic cues than feature-defining gestures like tongue-body lowering:

> [W]hile feature-defining gestures are, in certain contexts, subject to severe weakening up to and including obliteration, enhancement gestures are far more robust and are apparently never obliterated… We hypothesize that overlap is responsible for the deviations in careful speech. We also suppose that, unlike feature-defining gestures, enhancement gestures are never subject to overlap severe enough to mask their acoustic consequences. (Keyser and Stevens, 2006, p. 57–58)

It turns out to be relatively common for an enhancement gesture to serve as a proxy for a phonological feature whose defining gesture is obliterated in the phonetics (e.g., Stevens and Keyser, 1989, 2010; Keyser and Stevens, 2001, 2006; Stevens, 2002; Flynn, 2011, and references therein). So it remains defensible to use the traditional feature [low] to represent pharyngeals and certain emphatics (Cole, 1991, p. 25; Coleman, 1998, p. 69; Jensen, 2004, p. 97; Calabrese, 2005, p. 59–60; Hayes, 2009, p. 87–88; Odden, 2013, p. 54, 60; etc.), on the understanding that this feature is implemented with additional gestures in the phonetics, such as jaw lowering (Nolan, 1995) and larynx raising (Esling, 1999), and that an enhancement gesture like pharyngeal constriction acts as a proxy for [low] in certain phonetic contexts (Keyser and Stevens, 1994, p. 231). This may be the case in Tigre ejectives (Section 3.2) and perhaps in Northwest Caucasian emphatics (Section 3.5), where the tongue body is indeed not always low.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

DF: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

Conflict of interest

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

Publisher's note

References

Abo Mokh, N., and Davis, S. (2020). "What triggers 'imala: focus on a Palestinian variety with phonological analysis," in *Perspectives on Arabic Linguistics XXXII: Papers selected from the Annual Symposium on Arabic Linguistics, Tempe, Arizona, 2018*, ed. E. V. Gelderen (Amsterdam: John Benjamins), 35–54.

Al Bulushi, H. S. M. (2019). *The Phonology of Modern South Arabian Harsusi of Oman* (MA thesis). University of Texas, Austin, TX, United States.

Al-Ansari, N., and Kulikov, V. (2023). Acoustics of Arabic uvulars and emphatic coronals: evidence for uvularization of emphatics in Qatari Arabic. *Al-'Arabiya* 55–56, 37–66. Available online at: https://muse.jhu.edu/article/921322

Al-Bataineh, H. (2019). Emphasis harmony in Arabic: a critical assessment of feature-geometric and optimality-theoretic approaches. *Languages* 4:79. doi: 10.3390/languages4040079

Alfaifi, A. H., Cavar, M. E., and Lulich, S. M. (2020). Tongue root position in Hijazi Arabic voiceless emphatic and non-emphatic coronal consonants. *Proc. Meet. Acoust.* 42:e060004. doi: 10.1121/2.0001391

Ali, L. H., and Daniloff, R. G. (1972). A contrastive cinefluorographic investigation of the articulation of emphatic-non emphatic cognate consonants. *Stud. Linguist.* 26, 81–105. doi: 10.1111/j.1467-9582.1972.tb00589.x

al-Kathīrī, Ā. i. A. A. (2019). "al-Lahja al-Kathīriyya (a?-?aniyya) fī bādiyat □ufār, 'Umān: Dirāsa lugawiyya wa?fiyya," in *Presented at: AIDA (Association Internationale de Dialectologie Arabe) 13th International Conference* (Kutaisi, GA: Akaki Tsereteli State University), 10–13.

Allen, B., Pulleyblank, D., and Ajíbóyè, Ọ. (2013). Articulatory mapping of Yoruba vowels: an ultrasound study. *Phonology* 30, 183–210. doi: 10.1017/S0952675713000110

Alqahtani, M., and Almoaily, M. (2023). Debuccalization in Gulf Pidgin Arabic: OT parallelism or harmonic serialism. *J. Semit. Stud.* 68, 139–164. doi: 10.1093/jss/fgac022

Al-Raba'a, B. I. M., and Davis, S. (2020). The typology of pharyngealization in Arabic dialects focusing on a rural Jordanian variety. *J. Univ. Lang.* 21, 1–42. doi: 10.22425/jul.2020.21.2.1

Al-Solami, M. A. (2017). Ultrasound study of emphatics, uvulars, pharyngeals and laryngeals in three Arabic dialects. *Can. Acoust.* 45, 25–35. Available online at: https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2603

Al-Tairi, H. (2018). *Tongue Retraction in Arabic: An Ultrasound and Acoustic Study* (doctoral dissertation). University of Auckland, Auckland, New Zealand.

Al-Tairi, H., Brown, J., Watson, C., and Gick, B. (2016). "Arabic emphatics are uvularized," in *Presented at: 2016 Annual Meeting on Phonology (University of Southern California, Los Angeles)*. Available online at: https://researchspace.auckland.ac.nz/bitstream/handle/2292/32344/final_versionwithcoversheet.pdf (accessed August 20, 2023).

Al-Tairi, H., Brown, J., Watson, C., and Gick, B. (2017). "Tongue retraction in Arabic: an ultrasound study," in *Proceedings of the 2016 Annual Meeting on Phonology*, eds. K. Jesney, C. O'hara, C. Smith and R. Walker (Washington, DC: Linguistic Society of America), 1–12.

Al-Taisan, H. A. A. (2022). *Velar, Uvular and Pharyngeal Alternations in Hasawi Arabic: a Harmonic Serialism Optimality Theoretic Approach* (doctoral dissertation). University of Essex, Essex, United Kingdom.

Al-Tamimi, F., Alzoubi, F., and Tarawnah, R. (2009). A videofluoroscopic study of the emphatic consonants in Jordanian Arabic. *Folia Phoniatrica et Logopaedica* 61, 247–253. doi: 10.1159/000235644

Al-Tamimi, F., and Heselwood, B. (2011). "Nasoendoscopic, videofluoroscopic and acoustic study of plain and emphatic coronals in Jordanian Arabic," in *Instrumental Studies in Arabic Phonetics*, eds. Z. Hassan and B. Heselwood (Amsterdam: John Benjamins), 163–192.

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: implications for formal representations. *Lab. Phonol.* 8:28. doi: 10.5334/labphon.19

Alwabari, S. (2020). *Phonological and Physiological Constraints on Assimilatory Pharyngealization in Arabic: Ultrasound Study* (doctoral dissertation). University of Ottawa, Ottawa, ON, Canada.

Ananian, C. S., and Nevins, A. I. (2001). *Postvelar Harmonies: a Typological Odyssey*. Cambridge, MA: M.I.T. Available online at: https://www.academia.edu/download/31034685/10.1.1.1.7874.pdf (accessed July 14, 2023).

Andersson, S., Vaux, B., and Pysipa, Z. (2023). Cwyzhy Abkhaz. *J. Int. Phonet. Assoc.* 53, 266–286. doi: 10.1017/S0025100320000390

Anonby, E. (2020). Emphatic consonants beyond Arabic: the emergence and proliferation of uvular-pharyngeal emphasis in Kumzari. *Linguistics* 58, 275–328. doi: 10.1515/ling-2019-0039

Archangeli, D., and Pulleyblank, D. (1994). *Grounded Phonology*. Cambridge, MA: MIT Press.

Archibald, J. (2003). Learning to parse second language consonant clusters. *Can. J. Linguist.* 48, 149–177. doi: 10.1017/S0008413100000633

Archibald, J. (2005). Second language phonology as redeployment of L1 phonological knowledge. *Can. J. Linguist.* 50, 285–314. doi: 10.1353/cjl.2007.0000

Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Sec. Lang. Res.* 25, 231–233. doi: 10.1177/0267658308100284

Archibald, J. (2018). "Transfer, contrastive analysis and interlanguage phonology," in *The Routledge Handbook of Contemporary English Pronunciation*, eds. O. Kang, R. I. Thomson and J. M. Murphy (Abingdon, VA: Routledge), 9–24.

Archibald, J. (2021). Ease and difficulty in L2 phonology: a mini-review. *Front. Commun.* 6:26529. doi: 10.3389/fcomm.2021.626529

Archibald, J. (2022). Segmental and prosodic evidence for property-by-property transfer in L3 English in Northern Africa. *Languages* 7:28. doi: 10.3390/languages7010028

Archibald, J. (2023). Phonological redeployment and the mapping problem: cross-linguistic E-similarity is the beginning of the story, not the end. *Sec. Lang. Res.* 39, 287–297. doi: 10.1177/02676583211066413

Archibald, J., Yousefi, M., and Alhemaid, A. (2022). Redeploying appendices in L2 phonology: illusory vowels in L1 Persian and Arabic acquisition of English s+ C initial clusters. *J. Monoling. Biling. Speech* 4, 76–108. doi: 10.1558/jmbs.20388

Atkey, S. (2002). "The acquisition of non-native segmental contrasts: a look at English speakers' acquisition of Czech palatal stops," in *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-Language Speech*, eds. A. James and J. Leather (Klagenfurt: University of Klagenfurt), 22–31.

Avery, P., and Idsardi, W. J. (2001). "Laryngeal dimensions, completion and enhancement," in *Distinctive Feature Theory*, ed. T. A. Hall (Berlin: Mouton de Gruyter), 41–70.

Bach, E. (1991). Representations and operations in Haisla phonology. *Proc. West. Conf. Linguist.* 4, 29–35.

Bach, E. (1997). "Some questions from Haisla morphology and phonology [unpublished report]," in *Paper Presented at the University of British Columbia Colloquium Series in Linguistics*. Vancouver, BC.

Bach, E., Howe [Flynn], D., and Shaw, P. A. (2005). "On epenthesis and moraicity in North Wakashan," in *Presented at: 79th Annual Meeting of the Linguistic Society of America*. Oakland, CA.

Bashir, E. (2009). "Wakhi," in *The Iranian Languages*, ed. G. Windfuhr (London/New York, NY: Routledge), 825–857.

Beguš, G. (2021). "Segmental phonetics and phonology in Caucasian languages," in *The Oxford Handbook of Languages of the Caucasus*, ed. M. Polinsky (Oxford: Oxford University Press), 688–728.

Bellem, A. (2007). *Towards a Comparative Typology of Emphatics Across Semitic and Into Arabic Dialect Phonology* (doctoral dissertation). University of London, London, United Kingdom.

Bellem, A. (2009). "The "problem" of pharyngealization and its role in the sound systems of North-East Caucasian languages," in *Presented at: International Workshop on Pharyngeals and Pharyngealization*. Newcastle.

Bellem, A. (2014). "Triads, emphatics and interdentals in Arabic sound system typology," in *Arab and Arabic Linguistics: Traditional and New Theoretical Approaches*, ed. M. E. B. Giolfo (Oxford: Oxford University Press), 9–41.

Bellem, A., and Watson, J. C. E. (2014). "Backing and glottalisation in three SWAP language varieties," in *Arab and Arabic Linguistics: Traditional and New Theoretical Approaches*, ed. M. E. B. Giolfo (Oxford: Oxford University Press), 169–207.

Bernabela, R. S. (2011). *A Phonology and Morphology Sketch of the Šihhi Arabic Dialect of ʔlGēdih (Oman)* (MA thesis). Leiden University, Leiden, Netherlands.

Bessell, N. J. (1993). *Towards a Phonetic and Phonological Typology of Postvelar Articulation* (doctoral dissertation). University of British Columbia, Vancouver, BC, Canada.

Bessell, N. J. (1998a). Local and non-local consonant-vowel interaction in Interior Salish. *Phonology* 15, 1–40. doi: 10.1017/S0952675798003510

Bessell, N. J. (1998b). "Phonetic aspects of retraction in Interior Salish," in *Salish Languages and Linguistics: Theoretical and Descriptive Perspectives*, eds. E. Czaykowska-Higgins and M. D. Kinkade (The Hague: Mouton de Gruyter), 125–152.

Bessell, N. J., and Czaykowska-Higgins, E. (1991). "The phonetics and phonology of postvelar sounds in Moses-Columbia Salish (Nxa'amxcin)," in *Presented at: Annual Meeting of the Canadian Linguistics Association*. Kingston, ON: Queen's University.

Bin-Muqbil, M. S. (2006). *Phonetic and Phonological Aspects of Arabic Emphatics and Gutturals* (Doctoral dissertation). University of Wisconsin-Madison, Madison, WI, United States.

Bird, S. (2003). *The Phonetics and Phonology of Lheidli Intervocalic Consonants* (Doctoral dissertation). University of Arizona, Tucson, AZ, United States.

Bird, S., and Onosson, S. (2022). A phonetic case study of Tsilhqot'in/z/and/zʔ/. *J. Int. Phonet. Assoc. First View* 2022, 1–34. doi: 10.1017/S0025100322000093

Blevins, J. (2004). *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Boas, F. (1924). Vocabulary of the Athapascan Tribe of Nicola Valley, British Columbia. *Int. J. Am. Linguist.* 3, 36-38. doi: 10.1086/463747

Bomhard, A. R. (2008). "A sketch of Proto-Afrasian phonology," in *Semitio-Hamitic Festschrift*, eds. A. B. Dolgopolsky and H. Jungraithmayr (Sprache und Oralität in Afrika 24) (Berlin: Reimer), 79–92.

Brame, M. K. (1970). *Arabic Phonology: Implications for Phonological Theory and Historical Semitic* (doctoral dissertation). MIT, Cambridge, MA, United States.

Brown, C. A. (2000). "The interrelation between speech perception and phonological acquisition from infant to adult," in *Second Language Acquisition and Linguistic Theory*, ed. J. Archibald (Oxford: Blackwell), 4–63.

Brown, J., Davis, H., Schwan, M., and Sennott, B. (2016). Gitksan. *J. Int. Phonet. Assoc.* 46, 367–378. doi: 10.1017/S0025100315000432

Bulakh, M., and Kogan, L. (2011). Arabic influences on Tigre: a preliminary evaluation. *Bullet. Sch. Orient. Afri. Stud.* 74, 1–39. doi: 10.1017/S0041977X10000698

Calabrese, A. (2005). *Markedness and Economy in a Derivational Model of Phonology*. Storrs, CT: University of Connecticut.

Cantera, A. (2017). "The phonology of Iranian," in *Handbook of Comparative and Historical Indo-European Linguistics*, eds. K. Jared, J. Brian and F. Matthias (Berlin, Boston, MA: De Gruyter Mouton), 481–503.

Cantineau, J. (1960). *Études de linguistique arabe*. Paris: C. Klincksieck.

Card, E. A. (1983). *A phonetic and phonological study of Arabic emphasis* (Doctoral dissertation). Cornell University, Ithaca, NY, United States.

Catford, J. C. (1983). "Pharyngeal and laryngeal sounds in Caucasian languages," in *Vocal Fold Physiology: Contemporary Research and Clinical Issues*, eds. D. M. Bless and J. H. Abbs (San Diego, CA: College Hill Press), 344–350.

Catford, J. C. (1984). "Instrumental data and linguistic phonetics," in *Topics in Linguistic Phonetics in Honour of E. T. Uldall*, eds. J. a. W. Higgs and R. Thelwall (Ulster: Ulster University), 23–48.

Catford, J. C. (1992). "Caucasian phonetics and general phonetics," in *Caucasologie et mythologie comparée: Actes du Colloque international du C.N.R.S. - IVe Colloque de Caucasologie (Sèvres, 27–29 Juin 1988)* (Paris: Peeters), 193–216.

Chirikba, V. A. (2014). "Sadzskij dialekt abxazskogo jazyka i ego govory," in *Sadzy: Istoriko-etnograficheskie ocherki*, ed. S. D. Inal-Ipa (Sukhumi), 274–374.

Chomsky, N., and Halle, M. (1968). *The Sound Pattern of English*. New York, NY: Harper & Row.

Clements, G. N. (1985). The geometry of phonological features. *Phonol. Yearb.* 2, 225–252. doi: 10.1017/S0952675700000440

Clements, G. N. (2009). "The role of features in phonological inventories," in *Contemporary Views on Architecture and Representations in Phonological Theory*, eds. E. Raimy and C. E. Cairns (Cambridge, MA: MIT Press), 19–68.

Cohn, A. C. (1993). "Voicing and vowel height in Madurese: a preliminary report," in *Tonality in Austronesian Languages*, eds. J. A. Edmondson and K. J. Gregerson (Honolulu, HI: University of Hawaii Press), 107–121.

Colarusso, J. (1985). Pharyngeals and pharyngealization in Salishan and Wakashan. *Int. J. Am. Linguist.* 51, 366–368. doi: 10.1086/465895

Colarusso, J. (1988). *The Northwest Caucasian Languages: A Phonological Survey*. New York, NY: Garland.

Colarusso, J. (1992). *A Grammar of the Kabardian Language*. Calgary, AB: University of Calgary Press.

Colarusso, J. (2013). "The typology of gutturals," in *Base articulatoire arrière*, eds. J. L. Léonard and S. Naïm (Munich: LINCOM), 93–109.

Cole, J. (1987). *Planar Phonology and Morphology* (doctoral dissertation). Massachusetts Institute of Technology, Cambridge, MA, United States.

Cole, J. (1991). *Planar Phonology and Morphology*. New York, NY: Garland Pub.

Coleman, J. (1998). *Phonological Representations: Their Names, Forms and Powers*. Cambridge: Cambridge University Press.

Comrie, B. (2005). Introduction to Caucasian. *Lingua* 115, 1–4. doi: 10.1016/j.lingua.2003.07.009

Cook, E.-D. (1978). "Flattening in Chilcotin and Interior Salish: a case of an areal rule," in *Presented at: Annual Meeting of the Canadian Linguistic Association*. London, ON.

Cook, E.-d. (1981). Athapaskan linguistics: proto-Athapaskan phonology. *Ann. Rev. Anthropol.* 10, 253–273. doi: 10.1146/annurev.an.10.100181.001345

Cook, E.-D. (1983). Chilcotin flattening. *Can. J. Linguist.* 28, 123–132. doi: 10.1017/S0008413100024075

Cook, E.-D. (1984). "Pharyngealization and related phenomena," in *Presented at: 9th International Conference on Salishan and Neighboring Languages*. Victoria, BC.

Cook, E.-D. (1985). "Pharyngealization in the Northwestern languages," in *Presented at: 24th Conference on American Indian Languages*. Washington, DC.

Cook, E.-D. (1987). "Shuswap vowels and Proto-Salish from an Athabaskan point of view," in *Presented at: 12th International Conference on Salishan and Neighboring Languages*. Vancouver, BC.

Cook, E.-D. (1989). "Articulatory and acoustic correlates of pharyngealization: evidence from Athapaskan," in *Theoretical Perspectives on Native American Languages*, eds. D. B. Gerdts and K. Michelson (Albany, NY: State University of New York Press), 133–148.

Cook, E.-D. (1990). Consonant classes and vowel qualities in Babine. *Can. J. Linguist.* 35, 123–143. doi: 10.1017/S0008413100013542

Cook, E.-D. (1993a). Chilcotin flattening and autosegmental phonology. *Lingua* 91, 149–174. doi: 10.1016/0024-3841(93)90011-K

Cook, E.-D. (1993b). Phonetic and phonological features of approximants in Athabaskan and Eskimo. *Phonetica* 50, 234–244. doi: 10.1159/000261944

Cook, E.-D. (2004). *A Grammar of Dëne Sųłiné (Chipewyan)*. Winnipeg, MB: Algonquian and Iroquoian Linguistics.

Cook, E.-D. (2013). *A Tsilhqút'ín Grammar*. Vancouver, BC: UBC Press.

Cook, E.-D., and Flynn, D. (2020). "Indigenous languages in Canada," in *Contemporary Linguistic Analysis*, eds. J. Archibald and W. O'grady (Toronto, ON: Pearson), 298–314.

Cowan, N., Braine, M. D. S., and Leavitt, L. A. (1985). The phonological and metaphonological representation of speech: evidence from fluent backward talkers. *J. Mem. Lang.* 24, 679–698. doi: 10.1016/0749-596X(85)90053-1

Czaykowska-Higgins, E. (1987). *Characterizing Tongue Root Behavior [Unpublished Report]*. Cambridge, MA: MIT.

Czaykowska-Higgins, E. (1990). "Retraction in Moses-Columbia Salish," in *25th International Conference on Salish and Neighboring Languages*, ed M. D. Kinkade (Vancouver, BC: Department of Linguistics, University of British Columbia), 81–96. Available online at: https://lingpapers.sites.olt.ubc.ca/

Czaykowska-Higgins, E., and Kinkade, M. D. (1998). "Salish languages and linguistics," in *Salish Languages and Linguistics: Theoretical and Descriptive Perspectives*, eds. E. Czaykowska-Higgins and M. D. Kinkade (The Hague: Mouton de Gruyter), 1–70.

Davis, H. (2020). "Salish languages," in *The Routledge Handbook of North American Languages*, eds. D. Siddiqi, M. Barrie, C. Gillon, J. D. Haugen and É. Mathieu (New York, NY: Routledge), 452–472.

Davis, S. (1993). "Arabic pharyngealization and phonological features," in *Perspectives on Arabic Linguistics V: Papers From the Fifth Annual Symposium on Arabic Linguistics*, eds. M. Eid and C. Holes (Amsterdam: John Benjamins), 149–162.

Davis, S. (1995). Emphasis spread in Arabic and Grounded Phonology. *Linguist. Inq.* 26, 465–498.

Davis, S. (2006). "Syllabic constituents," in *the Encyclopedia of Language and Linguistics*, eds. K. Brown and R. E. Asher (Amsterdam: Elsevier).

Davis, S., and Alfaifi, A. (2019). A different path to [f]: labiodentalization in Faifi Arabic. *Pap. Hist. Phonol.* 4, 45-61. doi: 10.2218/pihph.4.2019.3072

Demolin, D. (2002). "The search for primitives in phonology and the explanation of sound patterns: the contribution of fieldwork studies," in *Laboratory Phonology 7*, eds. C. Gussenhoven and N. Warner (Berlin: Mouton de Gruyter), 455–514.

Diakonoff, I. M. (1984). "An evaluation of Eblaite," in *Studies on the language of Ebla (Papers Presentet at the Colloquium "La Lingua di Ebla e la Linguistica Semitica," 24–26 June 1982)*, ed. P. Fronzaroli (Florence: Ist. di Linguistica e di Lingue Orientali), 1–10.

Dolgopolsky, A. B. (1977). Emphatic consonants in Semitic. *Israeli Oriental Studies* 7, 1–13.

Duanmu, S. (2007). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

Dunn, J. A. (1995). *Sm'algya: A Reference Dictionary and Grammar for the Coast Tsimshian Language*. Juneau: Sealaska Heritage Foundation.

Eastman, C. M., and Aoki, P. K. (1978). "Phonetic segments in Haida (Hydaburg dialect)," in *Linguistic and Literary Studies in Honor of Archibald A. Hill*, eds. M. A. Jazayery, E. C. Polomé and W. Winter (Berlin: Mouton), 237–249.

Eckman, F., and Iverson, G. K. (2013). The role of native language phonology in the production of L2 contrasts. *Stud. Sec. Lang. Acquisit.* 35, 67–92. doi: 10.1017/S027226311200068X

Edelman, D. J. I., and Dodykhudoeva, L. R. (2009). "The Pamir languages," in *The Iranian Languages*, ed. G. Windfuhr (London/New York, NY: Routledge), 773–786.

Ehret, C. (1995). *Reconstructing Proto-Afroasiatic (Proto-Afrasian): Vowels, Tone, Consonants, and Vocabulary*. Berkeley, CA: University of California Press.

Elorrieta Puente, F. J. (1991). The feature specification of uvulars. *Proc. West Coast Conf. Form. Linguist.* 10, 139–149.

Esling, J. H. (1999). The IPA categories "pharyngeal" and "epiglottal": laryngoscopic observations of pharyngeal articulations and larynx height. *Lang. Speech* 42, 349–372. doi: 10.1177/00238309990420040101

Esling, J. H., Moisik, S. R., Benner, A., and Crevier-Buchman, L. (2019). *Voice Quality: The Laryngeal Articulator Model*. Cambridge: Cambridge University Press.

Fallon, P. D. (2002). *The Synchronic and Diachronic Phonology of Ejectives*. New York, NY: Routledge.

Faust, N. (2017). Get that into your head: Tigre vowel harmonies as templatic. *Glossa* 2:95. doi: 10.5334/gjgl.466

Flynn, D. (2011). Floating yet grounded: feature transmutation in Optimality Theory. *Can. J. Linguist.* 56, 377–402. doi: 10.1017/S000841310000205X

Flynn, D. (2012). *Phonology: The Distinctive Features of Speech Sounds*. University of Calgary. Available online at: https://ucalgary.ca/sites/default/files/teams/466/flynn12_distinctive_features.pdf (accessed September 8, 2023).

Flynn, D., and Fulop, S. (2012). Dentals are grave. *Can. Acoust.* 40, 20–21. Available online at: https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2042

Flynn, D., and Fulop, S. (2014). Acoustic phonetic features in Athabaskan sound change. *Diachronica* 31, 192–222. doi: 10.1075/dia.31.2.02fly

Fre Woldu, K. (1986). Evidence of auditory similarity between Tigrinya ejective/t'/and Arabic emphatic/ʔ/. *Orientalia Suecana* 33, 123–138.

Freeman, A. (2019). *Rhotic Emphasis and Uvularization in Moroccan Arabic* (doctoral dissertation). University of Pennsylvania, Philadelphia, PA, United States.

Fulop, S., Kari, E., and Ladefoged, P. (1998). An acoustic study of the tongue root contrast in Degema vowels. *Phonetica* 55, 80–98. doi: 10.1159/000028425

Gasparini, F. (2021). Emphasis, glottalization and pharyngealization in Semitic and Afroasiatic. *Kervan Int. J. Afri. Asian Stud.* 25, 3–30. doi: 10.13135/1825-263X/6247

Gessner, S. (2003). *The Prosodic System of the Dakelh (Carrier) Language* (doctoral dissertation). University of British Columbia, Vancouver, BC, United Kingdom.

Ghazeli, S. (1977). *Back Consonants and Backing Coarticulation in Arabic* (doctoral dissertation). University of Texas, Austin, TX, United States.

Goad, H. (1989). On the feature [rtr] in Chilcotin: a problem for the feature hierarchy. *Coyote Pap.* 9, 20–31.

Goad, H. (1991). "[Atr] and [Rtr] are different features," in *Proceedings of the Tenth West Coast Conference on Formal Linguistics*, ed. D. Bates (Stanford, CA: CSLI Publications), 163–173.

Gordon, M. (1996). The phonetic structures of Hupa. *Univ. Calif. Work. Pap. Phonet.* 93, 164–187.

Gorecka, A. (1989). *Phonology of articulation* [Dissertation]. MIT, Cambridge, MA, United States.

Guion, S. G., Post, M. W., and Payne, D. L. (2004). Phonetic correlates of tongue root vowel contrasts in Maa. *J. Phonet.* 32, 517–542. doi: 10.1016/j.wocn.2004.04.002

Habib, R. (2022). Metathesis in Syrian Arabic: types and conditioning factors. *J. Univ. Lang.* 23, 1–34. doi: 10.22425/jul.2022.23.1.1

Hall, T. A. (2007). "Segmental features," in *The Cambridge Handbook of Phonology*, ed. P. De Lacy (Cambridge: Cambridge University Press), 311–334.

Halle, M. (1983). On distinctive features and their articulatory implementation. *Nat. Lang. Linguist. Theor.* 1, 91–105. doi: 10.1007/BF00210377

Halle, M. (1989). *The Intrinsic Structure of Speech Sounds* [Unpublished report]. Cambridge, MA: MIT.

Halle, M. (1995). Feature geometry and feature spreading. *Linguist. Inq.* 26, 1–46.

Halle, M., and Clements, G. N. (1983). *Problem Book in Phonology*. Cambridge, MA: MIT Press.

Halle, M., and Stevens, K. N. (1971). A note on laryngeal features. *MIT Res. Lab. Electr. Quart. Progr. Rep.* 101, 198–213.

Halle, M., Vaux, B., and Wolfe, A. (2000). On feature spreading and the representation of place of articulation. *Linguist. Inq.* 31, 387–444. doi: 10.1162/002438900554398

Hansson, G. Ó. (2010). *Consonant Harmony: Long-Distance Interaction in Phonology*. Berkeley, CA: University of California Press.

Hargus, S. (2007). *Witsuwit'en Grammar: Phonetics, Phonology, Morphology*. Vancouver, BC: UBC Press.

Harrell, R. S. (1957). *The Phonology of Colloquial Egyptian Arabic*. New York, NY: American Council of Learned Societies.

Hassan, Z., and Esling, J. H. (2011). "Investigating the emphatic feature of Iraqi Arabic," in *Instrumental Studies in Arabic Phonetics*, eds. Z. Hassan and B. Heselwood (Amsterdam: John Benjamins), 217–234.

Hayes, B. (1989). Compensatory lengthening in moraic phonology. *Linguist. Inq.* 20, 253–306.

Hayes, B. (2009). *Introductory Phonology*. West Sussex: Wiley-Blackwell.

Hayward, K. M., and Hayward, R. J. (1989). 'Guttural': arguments for a new distinctive feature. *Trans. Philol. Soc.* 87, 179–193. doi: 10.1111/j.1467-968X.1989.tb00626.x

Herzallah, R. S. (1990). *Aspects of Palestinian Arabic phonology: a nonlinear approach* (Doctoral dissertation). Cornell University, Ithaca, NY, United States.

Hess, S. A. (1998). *Pharyngeal Articulations* (doctoral dissertation). University of California, Los Angeles, CA, United States.

Hoberman, R. D. (1988). Emphasis harmony in a modern Aramaic dialect. *Language* 64, 1–26. doi: 10.2307/414783

Howe [Flynn], D. (1999a). "On debuccalization and the Pharyngeal place of laryngeals," in *Presented at: Annual Meeting of the Linguistic Society of America*. Los Angeles, CA.

Howe [Flynn], D. (1999b). "Two types of laryngeals: evidence from the Northwest," in *Presented at: Northwest Linguistics Conference* (Victoria, BC: University of Victoria).

Howe [Flynn], D. (1999c). "[Voice] in Wakashan," in *Presented at: 14th International Conference on Historical Linguistics*. Vancouver, BC.

Howe [Flynn], D. (2000). *Oowekyala Segmental Phonology* (doctoral dissertation). University of British Columbia, Vancouver, BC, United Kingdom.

Hudu, F. (2008). The low vowel and retraction in St'át'imcets: an ultrasound investigation. *SKY J. Linguist.* 21, 67–81. Available online at: https://www.linguistics.fi/julkaisut/SKY2008/Hudu_NETTIVERSIO.pdf

Huehnergard, J. (2017). "Arabic in its Semitic context," in *Arabic in Context*, ed. A. Al-Jallad (Berlin: Brill), 3–34.

Huehnergard, J. (2019). "Proto-Semitic," in *The Semitic Languages*, eds. J. Huehnergard and N. A. Pat-El (Abingdon, VA; New York, NY: Routledge), 49–79.

Huehnergard, J. (2023). "Proto-semitic and Egyptian," in *Ancient Egyptian and Afroasiatic: Rethinking the Origins*, eds. V. Almansa-Villatoro and S. Štubnová Nigrelli (University Park, PA: Pennsylvania State University Press/Eisenbrauns), 139–160.

Huehnergard, J., and Pat-El, N. a. (2019). "Introduction to the Semitic languages and their history," in *The Semitic Languages*, eds. J. Huehnergard and N. A. Pat-El (Abingdon, VA/New York, NY: Routledge), 1–21.

Huehnergard, J., and Rubin, A. D. (2012). "Phyla and waves: models of classification of the Semitic languages," in *The Semitic Languages*, eds S. Weninger (Berlin, Boston, MA: De Gruyter Mouton), 259–278.

Hyman, L. (1988). Underspecification and vowel height transfer in Esimbi. *Phonology* 5, 255–273. doi: 10.1017/S0952675700002293

Hyman, L. M. (2008). Enlarging the scope of phonologization. *UC Berkeley Phonol. Lab Ann. Rep.* 2008, 382–409. doi: 10.5070/P73ZM91694

Gebski, W. (2023). The phonology of the Judaeo-Arabic dialect of Gabes (Southern Tunisia). *J. Semit. Stud.* 68, 165–197. doi: 10.1093/jss/fgac024

Israel, A., Proctor, M., Goldstein, L., Iskarous, K., and Narayanan, S. (2012). "Emphatic segments and emphasis spread in Lebanese Arabic: a real-time magnetic resonance imaging study," in *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH 2012)*, ed. I. S. C. Association (Red Hook, NY: Curran Associates), 2175–2178.

Jacobsen, W. H. Jr. (1969). Origin of the Nootka pharyngeals. *Int. J. Am. Linguist.* 35, 125–153. doi: 10.1086/465049

Jakobson, R., Fant, G., and Halle, M. (1952). *Preliminaries to Speech Analysis*. Cambridge, MA: MIT Press.

Jaradat, M. (2020). Gutturals, emphatics, and the phenomenon of emphasis spread in fallaahi Jordanian Arabic: a non-linear analysis. *J. Lang. Linguist. Stud.* 16, 1656–1679. doi: 10.17263/jlls.850979

Jensen, J. T. (2004). *Principles of Generative Phonology: an Introduction*. Amsterdam: John Benjamins.

Johnson, R. (1975). *The Role of Phonetic Detail in Coeur d'Alene Phonology* (doctoral dissertation). Washington State University, Pullman, WA, United States.

Johnstone, T. M. (1977). *Harsusi Lexicon*. London, UK: Oxford University Press.

Jongman, A., Herd, W., Al-Masri, M., Sereno, J., and Combest, S. (2011). Acoustics and perception of emphasis in Urban Jordanian Arabic. *J. Phonet.* 39, 85–95. doi: 10.1016/j.wocn.2010.11.007

Kabak, B. (2019). "A dynamic equational approach to sound patterns in language change and second language acquisition: the (un)stability of English dental fricatives illustrated," in *Patterns in Language and Linguistics*, eds. B. Beatrix and M. F. Ruth (Berlin/Boston, MA: Mouton de Gruyter), 221–254.

Kahn, M. (1976). *Borrowing and Variation in a Phonological Description of Kurdish* (doctoral dissertation). University of Michigan, Ann Arbor, MI, United States.

Kari, J. (1977). Linguistic diffusion between Tanaina and Ahtna. *Int. J. Am. Linguist.* 43, 274–288. doi: 10.1086/465499

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286–319. doi: 10.2307/413642

Keating, P. A. (1988). *A Survey of Phonological Features*. Bloomington, IN: Indiana University Linguistics Club.

Kenstowicz, M., and Louriz, N. (2009). Reverse engineering: emphatic consonants and the adaptation of vowels in French loanwords into Moroccan Arabic. *Brill's Ann. Afroasiat. Lang. Linguist* 1, 41–74. doi: 10.1163/187666309X12491131130701

Keyser, S. J., and Stevens, K. N. (1994). Feature geometry and the vocal tract. *Phonology* 11, 207–236. doi: 10.1017/S0952675700001950

Keyser, S. J., and Stevens, K. N. (2001). "Enhancement revisited," in *Ken Hale: A Life in Language*, ed. M. Kenstowicz (Cambridge, MA: MIT Press), 271–291.

Keyser, S. J., and Stevens, K. N. (2006). Enhancement and overlap in the speech chain. *Language* 82, 33–63. doi: 10.1353/lan.2006.0051

Kingston, J. (1985). *The Phonetics and Phonology of the Timing of Oral and Glottal Events* (dissertation). University of California, Berkeley, Berkeley, CA, United States.

Kinkade, M. D. (1967). Uvular-pharyngeal resonants in Interior Salish. *Int. J. Am. Linguist.* 33, 228–234. doi: 10.1086/464965

Knudsen, E. E. (1969). "Spirantization of velars in Akkadian," in *Lišān mitḫurti: Festschrift Wolfram Freiherr von Soden zum 19. VI. 1968 gewidmet von Schülern und Mitarbeitern*, eds. W. Röllig and M. Dietrich (Kevelaer: Neukirchen-Vluyn), 147–155.

Kogan, L. (2012). "Proto-Semitic phonetics and phonology," in *The Semitic Languages*, ed. S. Weninger (Berlin, Boston, MA: De Gruyter Mouton), 54–150.

Kogan, L., and Bulakh, M. (2019). "Soqotri," in *The Semitic Languages*, eds. J. Huehnergard and N. A. Pat-El (Abingdon: Routledge), 280–320.

Kogan, L., and Naumkin, V. (2014). The vowels of Soqotri as a phonemic system. *Proc. Semin. Arab. Stud.* 44, 57–79. Available online at: https://www.jstor.org/stable/43782851

Krauss, M. (1975). *Chilcotin Phonology: A Descriptive and Historical Report, With Recommendations for a Chilcotin Orthography* (unpublished report). University of Alaska, Fairbanks, AK, United States.

Kuipers, A. (1974). *The Shuswap Language: Grammar, Texts, Dictionary*. The Hague: Mouton.

Kuipers, A. (1981). On reconstructing the Proto-Salish sound system. *Int. J. Am. Linguist.* 47, 323–335. doi: 10.1086/465702

Kulikov, V., Mohsenzadeh, F. M., and Syam, R. M. (2023). Effect of emphasis spread on VOT in coronal stops in Qatari Arabic. *J. Int. Phonet. Assoc.* 53, 456–469. doi: 10.1017/S0025100321000256

Kümmel, M. (2007). *Konsonantenwandel: Bausteine zu einer Typologie des Lautwandels und ihre Konsequenzen für die vergleichende Rekonstruktion*. Wiesbaden: Reichert.

Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago, IL: Chicago University Press.

Ladefoged, P., and Maddieson, I. (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lahiri, A. (2018). "Predicting universal phonological contrasts," in *Phonological Typology*, eds. L. M. Hyman and F. Plank (Berlin/Boston, MA: de Gruyter), 229–272.

Lardiere, D. (2009). Some thoughts on the contrastive analysis of features in second language acquisition. *Sec. Lang. Res.* 25, 173–227. doi: 10.1177/0267658308100283

Lass, R. (1984). *Phonology: an Introduction to Basic Concepts*. Cambridge: Cambridge University Press.

Lass, R., and Anderson, J. M. (1975). *Old English Phonology*. Cambridge: Cambridge University Press.

Latimer, R. (1978). *A Study of Chilcotin Phonology* (MA thesis), University of Calgary, Calgary, AB, Canada.

Laufer, A., and Baer, T. (1988). Emphatic and pharyngeal sounds in Hebrew and in Arabic. *Lang. Speech* 31, 181–208. doi: 10.1177/002383098803100205

Lee, S.-H. (1995). "Orals, gutturals, and the jaw," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, eds. B. Connell and A. Arvaniti (Cambridge: Cambridge University Press), 343–360.

Leer, J. (1979). Proto-Athabaskan verb stem variation, part one: phonology. *Alaska Native Lang. Center Res. Pap.* 1, 1–100.

Leer, J. (1996). "The historical evolution of the stem syllable in Gwich'in (Kutchin/Loucheux) Athabaskan," in *Athabaskan Language Studies: Essays in Honor of Robert W. Young*, eds. E. Jelinek, S. Midgette, K. Rice and L. Saxon (Albuquerque, NM: University of New Mexico Press), 193–234.

Leer, J. (2011). "The palatal series in Athabascan-Eyak-Tlingit, with an overview of the basic sound correspondences," in *The Dene-Yeniseian Connection*, eds. J. Kari and B. A. Potter (Fairbanks, AK: Alaska Native Language Center), 168–193.

Lincoln, N. J., and Rath, J. C. (1980). *North Wakashan Comparative Root List*. Ottawa, ON: National Museums of Canada.

Lincoln, N. J., and Rath, J. C. (1986). *Phonology, Dictionary and Listing of Roots and Lexical Derivates of the Haisla Language of Kitlope and Kitimaat, B.C.* Ottawa, ON: National Museums of Canada.

Lindau, M. (1975). Larynx height in Kwa. *J. Acoust. Soc. Am.* 58:12. doi: 10.1121/1.2001959

Lowenstamm, J., and Prunet, J.-F. (1988). "Tigre vowel harmonies," in *Rapport Annuel du Groupe de Recherche sur la Linguistique Africaniste au CRSH 1987-88.* Montreal: Université du Québec à Montréal.

Maddieson, I. (1997). "Phonetic universals," in *The Handbook of Phonetic Sciences*, eds. J. Laver and W. J. Hardcastle (Oxford: Blackwell), 619–639.

Mahadin, R. S., and Bader, Y. (1995). Emphasis assimilation spread in Arabic and feature geometry of emphatic consonants. *J. Semit.* 7, 87–113.

Martinet, A. (1953). Remarques sur le consonantisme sémitique. *Bulletin de la Société de Linguistique de Paris* 49, 67–78.

Martinez, R. M., Goad, H., and Dow, M. (2023). L1 phonological effects on L2 (non-)naïve perception: a cross-language investigation of the oral-nasal vowel contrast in Brazilian Portuguese. *Sec. Lang. Res.* 39, 387–423. doi: 10.1177/02676583211044953

McCarthy, J. J. (1988). Feature geometry and dependency: a review. *Phonetica* 43, 84–108. doi: 10.1159/000261820

McCarthy, J. J. (1991). "Semitic gutturals and distinctive feature theory," in *Perspectives on Arabic Linguistics III: Papers from the Third Annual Symposium on Arabic Linguistics*, eds. B. Comrie and M. Eid (Amsterdam: John Benjamins), 63–91.

McCarthy, J. J. (1994). "The phonetics and phonology of Semitic pharyngeals," in *Papers in Laboratory Phonology III: Phonological Structure and Phonetic Form*, ed. P. A. Keating (Cambridge: Cambridge University Press), 191–233.

McDowell, R. (2004). *Retraction in Montana Salish Lateral Consonants: an Ultrasonic Study* (MA thesis). UBC, Vancouver, BC, Canada.

Miller, A. (2011). "The representation of clicks," in *The Blackwell Companion to Phonology*, eds. M. Van Oostendorp, C. J. Ewen, E. Hume and K. Rice (Hoboken, NJ: Wiley-Blackwell), 416–439.

Moisik, S. (2013). *The Epilarynx in Speech* (Ph.D. dissertation). University of Victoria, Victoria, BC, Canada.

Moisik, S., Czaykowska-Higgins, E., and Esling, J. H. (2012). The epilaryngeal articulator: a new conceptual tool for understanding lingual-laryngeal contrasts. *McGill Work. Pap. Linguist.* 22, 1–15. Available online at: https://www.mcgill.ca/mcgwpl/files/mcgwpl/moisik2012.pdf

Moisik, S., and Esling, J. H. (2011). "The 'whole larynx' approach to laryngeal features," in *Proceedings of the 17th International Congress of Phonetic Sciences*, eds. W. S. Lee and E. Zee (Hong Kong: City University of Hong Kong), 1406–1409.

Moisik, S. R., Czaykowska-Higgins, E., and Esling, J. H. (2021). Phonological potentials and the lower vocal tract. *J. Int. Phonet. Assoc.* 51, 1–35. doi: 10.1017/S0025100318000403

Morice, A. G. (1932). *The Carrier Language*. Mödling bei Wien, St. Gabriel: Verlag der Internationalen Zeitschrift "Anthropos".

Naïm, S. (2008). "Compléments à Remarks on the spoken Arabic of Zabid," in *Presented at: 8th AIDA Conference*. Essex.

Naïm, S., and Watson, J. (2013). "La corrélation occlusive laryngovélaire dans des variétés néo-arabes et sud-arabiques," in *Base articulatoire arrière*, eds. N. Samia and L. Jean Léo (Muenchen: Lincom Academic Publishers), 133–155.

Nakao, S. (2022). "[+constricted glottis] reflexes of and q in contact situations: contact-induced change or inheritance?," in *Studies on Arabic Dialectology and Sociolinguistics: Proceedings of the 13th International Conference of AIDA, June 10-13, 2019*, eds. G. Chikovani and Z. Tskhvediani (Kutaisi, GA: Akaki Tsereteli State University), 387–396.

Namdaran, N. (2005). An ultrasonic investigation of retracted consonants in St'at'imcets (Lillooet Salish). *J. Acoust. Soc. Am.* 117, 2490–2491. doi: 10.1121/1.4787865

Namdaran, N. (2006). *Retraction in St'?t'imcets: an Ultrasonic Investigation* (MA thesis). University of British Columbia, Vancouver, BC, Canada.

Napiorkowska, L. (2021). "Modelling variation in the Neo-Aramaic dialect of Azran with Articulatory Phonology," in *Studies in the grammar and lexicon of Neo-Aramaic*, eds. G. Khan and P. M. Noorlander (Cambridge: Open Book Publishers), 319–334.

Natho, K. I. (2009). *Circassian History*. Bloomington, IN: Xlibris.

Nelson, B. C. (2023). Phonological redeployment for [retracted tongue root] in third language perception of Kaqchikel stops. *Front. Lang. Sci* 2:1253816. doi: 10.3389/flang.2023.1253816

Nelson, B. C., and Flynn, D. (2022). "Chrump's on Chwitter? A fresh look at posteriorization in English," in *Proceedings of the 2021 Annual Conference of the Canadian Linguistic Association*, eds. A. Hernández and C. Plyley (Halifax, NS: Saint Mary's University), 19.11–19.15.

Nicolaï, R. (1981). *Les dialectes du songhay: contribution à l'étude des changements linguistiques*. Paris: SELAF (Société d'études linguistiques et anthropologiques de France).

Nolan, F. (1995). "The role of the jaw active or passive? Comments on Lee," in *Phonology and Phonetic Evidence: Papers in Laboratory Phonology IV*, eds. B. Connell and A. Arvaniti (Cambridge: Cambridge University Press), 361–367.

Odden, D. (1991). Vowel geometry. *Phonology* 8, 261–289. doi: 10.1017/S0952675700001408

Odden, D. (2013). *Introducing Phonology*. Cambridge: Cambridge University Press.

Orel, V. E., and Stolbova, O. V. (1995). *Hamito-Semitic Etymological Dictionary: Materials for a Reconstruction*. Leiden: Brill.

Palmer, F. R. (1956). 'Openness' in Tigre: a problem in prosodic statement. *Bullet. Sch. Orient. Afri. Stud.* 18, 561–577. doi: 10.1017/S0041977X00088054

Pat-El, N. a. (2019). "The Semitic language family: a typological perspective," in *The Semitic Languages*, eds. J. Huehnergard and N. A. Pat-El (Abingdon, VA/New York, NY: Routledge), 80–94.

Perkell, J. S. (1971). Physiology of speech production: a preliminary study of two suggested revisions of the features specifying vowels. *MIT Res. Lab. Electr. Quart. Progr. Rep.* 102, 123–139.

Prince, A. (1975). *The Phonology and Morphology of Tiberian Hebrew* (doctoral dissertation). MIT, Cambridge, MA, United States.

Prince, A., and Smolensky, P. (2004). *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Purnell, T., and Raimy, E. (2015). "Distinctive features, levels of representation, and historical phonology," in *The Oxford Handbook of Historical Phonology*, eds. P. Honeybone and J. Salmons (Oxford: Oxford University Press), 522–544.

Raz, S. (1983). *Tigre Grammar and Texts*. Malibu, CA: Undena.

Reichel, N. (2010). The role of the educational system in retaining Circassian identity during the transition from Ottoman control to life as Israeli citizens (1878-2000). *Israel Affairs* 16, 251–267. doi: 10.1080/13537121003643896

Ridouane, R., and Gendrot, C. (2017). On ejective fricatives in Omani Mehri. *Brill's J. Afroasiat. Lang. Linguist.* 9, 139–159. doi: 10.1163/18776930-00901008

Roman, A. (1981). De la langue arabe comme un modèle général de la formation des langues sémitiques et de leur évolution. *Arabica* 28, 127–161. doi: 10.1163/157005881X00186

Rood, D. S. (1975). The implications of Wichita phonology. *Language* 51, 315–337. doi: 10.2307/412858

Rose, S. (1996). Variable laryngeals and vowel lowering. *Phonology* 13, 73–117. doi: 10.1017/S0952675700000191

Rose, S., and Walker, R. (2004). A typology of consonant agreement as correspondence. *Language* 80, 475–531. Available online at: https://www.jstor.org/stable/4489721

Sagey, E. (1986). *The Representation of Features and Relations in Nonlinear Phonology* (dissertation). MIT, Cambridge, MA, United States.

Shahin, K. N. (1996). "Salish emphatics," in *Presented at: 31st International Conference on Salishan and Neighboring Languages*. Vancouver, BC.

Shahin, K. N. (1997). "Acoustics of pharyngealization vs. uvularization harmony," in *Perspectives on Arabic Linguistics X: Papers From the Tenth Annual Symposium on Arabic Linguistics*, eds. R. R. Ratcliffe and E. Mushira (Amsterdam: John Benjamins), 215–238.

Shahin, K. N. (2002). *Postvelar Harmony*. Amsterdam: John Benjamins.

Shahin, K. N. (2011). "Pharyngeals," in *The Blackwell Companion to Phonology*, eds. M. Van Oostendorp, C. J. Ewen, E. Hume and K. Rice (Hoboken, NJ: Wiley-Blackwell), 604–627.

Shaw, P. A. (1993). "Templatic evidence for the syllable nucleus," in *Proceedings of the 23rd Annual Meeting of the North East Linguistic Society*, ed. A. Schafer (Amherst, MA: Graduate Linguistic Student Association), 463–477.

Shaw, P. A. (1994). "The prosodic constituency of minor syllables," in *The Proceedings of the Twelfth West Coast Conference on Formal Linguistics*, eds. E. Duncan, D. Farkas and P. Spaelti (Stanford, CA: Center for the Study of Language and Information), 117–132.

Shaw, P. A. (1996). "The non-nuclear status of syllabic obstruents in Berber," in *Presented at: Annual Meeting of the Linguistic Society of America*. San Diego, CA.

Shaw, P. A. (2002). On the edge: obstruent clusters in Salish. *UBC Work. Pap. Linguist.* 10, 119–136. Available online at: https://lingpapers.sites.olt.ubc.ca/wscla-volumes/

Shaw, P. A., Blake, S. J., Campbell, J., and Sheperd, C. (1999). Stress in h?q?mi?m (Musqueam) Salish. *UBC Work. Pap. Linguist.* 2, 131–163.

Shin, J., Kiaer, J., and Cha, J. (2012). *The Sounds of Korean*. Cambridge: Cambridge University Press.

Shockley, M. D. (2024). Ru'us al-Jibāl Arabic in context: a proposal for an expanded typology of Southeastern Arabian dialects. *J. Semit. Stud.* 69, 1–27. doi: 10.1093/jss/fgad026

Sievers, E. (1881). *Grundzüge der Phonetik*. Leipzig: Breitkopf and Hartel.

Simpson, A. (2003). *Gutturals in Diachronic Perspective: the Case of Pharyngeal Merger and Loss in Semitic and Beyond* (MA thesis). University of California, Berkeley, Berkeley, CA, United States.

Skjærvø, P. O. (2009). "Old Iranian," in *The Iranian Languages*, ed. G. Windfuhr (London/New York, NY: Routledge), 43–195.

Slimani, K. (2018). Emphasis spread in the Djelfa dialect of Algerian Arabic. *Brill's J. Afroasiat. Lang. Linguist.* 10, 285–307. doi: 10.1163/18776930-01002004

Sloat, C. (1968). *Phonological Redundancy Rules in Coeur d'Alene*. Ann Arbor, MI: University Microfilms.

Son, S. A. (2005). *The Acquisition of English Obstruents by Korean Speakers of English as a Second Language* (doctoral dissertation). University of Delaware, Newark, DE, United States.

Souag, L. (2010). *Arabic, Berber, and Songhay in Tabelbala and Siwa* (doctoral dissertation). School of Oriental and African Studies, University of London, London, United Kingdom.

Souag, L. (2012). The subclassification of Songhay and its historical implications. *J. Afri. Lang. Linguist.* 33, 181–213. doi: 10.1515/jall-2012-0008

Souag, L. (2020). "Songhay," in *The Oxford Handbook of African Languages*, eds. R. Vossen and G. Dimmendaal (Oxford: Oxford University Press), 644–653.

Steriade, D. (1987). "Locality conditions and feature geometry," in *Papers of Proceedings of the North East Linguistic Society 17* (Amherst, MA: GLSA Publications), 595–617.

Stevens, K. N. (1998). *Acoustic Phonetics*. Cambridge, MA: MIT Press.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *J. Acoust. Soc. Am.* 111, 1872–1891. doi: 10.1121/1.1458026

Stevens, K. N. (2005). "Features in speech perception and lexical access," in *The Handbook of Speech Perception*, eds. D. B. Pisoni and R. E. Remez (Malden, MA: Blackwell), 125–155.

Stevens, K. N., and Keyser, S. J. (1989). Primary features and their enhancement in consonants. *Language* 65, 81–106. doi: 10.2307/414843

Stevens, K. N., and Keyser, S. J. (2010). Quantal theory, enhancement and overlap. *J. Phonet.* 38, 10–19. doi: 10.1016/j.wocn.2008.10.004

Story, G. (1984). *Babine and Carrier Phonology: A Historically Oriented Study*. Dallas, TX: Summer Institute of Linguistics.

Sylak-Glassman, J. (2014). *Deriving Natural Classes: The Phonology and Typology of Post-Velar Consonants* (doctoral dissertation). University of California, Berkeley, Berkeley, CA, United States.

Trigo, L. (1988). *On the Phonological Derivation and Behavior of Nasal Glides* (doctoral dissertation). MIT, Cambridge, MA, United States.

Trigo, L. (1991). On pharynx-larynx interactions. *Phonology* 8, 113–136. doi: 10.1017/S0952675700001299

Trubetzkoy, N. S. (1931). Die Konsonantensysteme der ostkaukasischen sprachen. *Caucasica* 8, 1–52.

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Prague: Travaux du Cercle linguistique de Prague, 7.

Trubetzkoy, N. S. (1969). *Principles of Phonology*. Berkeley, CA: University of California Press.

Ullendorff, E. (1955). *The Semitic Languages of Ethiopia: a Comparative Phonology*. London: Taylor's (Foreign) Press.

Van Eijk, J. (1997). *The Lillooet Language: Phonology, Morphology, Syntax*. Vancouver, BC: University of British Columbia Press.

Van Eijk, J., and Nater, H. (2020). "Some notes on Proto-Salish phonology," in *Presented at: Proceedings of the 55th International Conference on Salish and Neighbouring Languages*. Vancouver, BC.

Vaux, B. (1992). "Adjarian's Law and consonantal ATR in Armenian," in *Proceedings of the Fourth International Conference on Armenian Linguistics*, ed. J. a. C. Greppin (Delmar, NY: Caravan), 271–293.

Vaux, B. (1994). *Armenian phonology* [Doctoral dissertation]. Harvard University, Cambridge, MA, United States.

Vaux, B. (1996). The status of ATR in feature geometry. *Linguist. Inq.* 27, 175–182.

Vaux, B. (1998). The laryngeal specifications of fricatives. *Linguist. Inq.* 29, 497–511. doi: 10.1162/002438998553833

Vaux, B., and Miller, B. (2011). "The representation of fricatives," in *The Blackwell Companion to Phonology*, eds. M. Van Oostendorp, C. J. Ewen, E. Hume and K. Rice (Hoboken, NJ: Wiley-Blackwell), 669–693.

Walker, R. (1979). "Central Carrier phonemics," in *Contributions to Canadian Linguistics*, ed. D. Zimmerly (Ottawa, ON: National Museums of Canada), 93–107.

Wallin, G. A. (1855). Über die Laute des Arabischen und ihre Bezeichnung. *Zeitschrift der Deutschen Morgenländischen Gesellschaft* 9, 1–69.

Wallis, E. E. (1987). The dynamics of vocalic harmony in Shapsugh Circassian. *Word* 38, 81–90. doi: 10.1080/00437956.1987.11435881

Watson, J. C., and Heselwood, B. (2016). "Phonation and glottal states in Modern South Arabian and San'ani Arabic," in *Perspectives on Arabic Linguistics XXVIII: Papers from the Annual Symposium on Arabic Linguistics, Gainesville, Florida, 2014*, eds. Y. A. Haddad and E. Potsdam (Amsterdam: John Benjamins), 3–37.

Watson, J. C. E. (1999). The directionality of emphasis spread in Arabic. *Linguist. Inq.* 30, 289–300. doi: 10.1162/002438999554066

Watson, J. C. E. (2002). *The Phonology and Morphology of Arabic*. Oxford: Oxford University Press.

Watson, J. C. E., and Bellem, A. (2011). "Glottalisation and neutralisation in Yemeni Arabic and Mehri: An acoustic study," in *Instrumental Studies in Arabic Phonetics*, eds. Z. Hassan and B. Heselwood (Amsterdam: John Benjamins), 235–256.

Weiss, M. (2015). "The comparative method," in *The Routledge Handbook of Historical Linguistics* (Abingdon: Routledge), 127–145.

Windfuhr, G., and Perry, J. R. (2009). "Persian and Tajik," in *The Iranian Languages*, ed. G. Windfuhr (London/New York, NY: Routledge), 416–544.

Woodward, M. F. (1964). "Hupa phonemics," in *Studies in California Linguistics*, ed. W. Bright (Berkeley, CA: University of California Press), 199–216.

Wright, R., Hargus, S., and Davis, K. (2002). On the categorization of ejectives: data from Witsuwit'en. *J. Int. Phonet. Assoc.* 32, 43–77. doi: 10.1017/S002510030200 0142

Zawaydeh, B., and de Jong, K. (2011). "The phonetics of localising uvularisation in Ammani-Jordanian Arabic," in *Instrumental Studies in Arabic Phonetics*, eds. Z. Hassan and B. Heselwood (Amsterdam: John Benjamins), 257–276.

Zawaydeh, B. A. (1998). "Gradient uvularization spread in Ammani-Jordanian Arabic," in *Perspectives on Arabic Linguistics XI*, eds. E. Benmamoun, N. Haeri and M. Eid (Amsterdam: John Benjamins), 117–141.

Zawaydeh, B. A. (1999). *The Phonetics and Phonology of Gutturals in Arabic* (doctoral dissertation). Indiana University, Bloomington, IN, United States.

Zawaydeh, B. A. (2003). "The interaction of the phonetics and phonology of gutturals," in *Phonetic Interpretation: Papers in Laboratory Phonology VI*, eds. J. Local, R. Ogden, D. and R. Temple (Cambridge: Cambridge University Press), 279–292.

Zemánek, P. (1996). *The Origins of Pharyngealization in Semitic*. Prague: Enigma Corporation.

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 13 December 2024
DOI 10.3389/flang.2024.1338625

Check for updates

OPEN ACCESS

EDITED BY
Jennifer Cabrelli,
University of Illinois Chicago, United States

REVIEWED BY
Miquel Llompart,
Pompeu Fabra University, Spain
John H. G. Scott,
University of Maryland, United States

*CORRESPONDENCE
Katharina S. Schuhmann
✉ Katharina.Schuhmann@uni-oldenburg.de

RECEIVED 14 November 2023
ACCEPTED 21 August 2024
PUBLISHED 13 December 2024

CITATION
Schuhmann KS and Smith LC (2024) From formalism to intuition: probing the role of the trochee in German nominal plural forms in L1 and L2 German speakers.
Front. Lang. Sci. 3:1338625.
doi: 10.3389/flang.2024.1338625

From formalism to intuition: probing the role of the trochee in German nominal plural forms in L1 and L2 German speakers

Katharina S. Schuhmann[1]* and Laura Catharine Smith[2]

[1]Department of German, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany,
[2]Department of German and Russian, Brigham Young University, Provo, UT, United States

Accounting for plural formation in Standard German (SG) nominals has proven to be a challenging endeavor. Numerous formalisms and models have been proposed and intensely debated over the past decades. The fundamental difficulty lies in the fact that German has a large number of suffix allomorphs, some of which can be used with or without stem-vowel fronting/raising (umlaut). Current research suggests that, at the segmental level, it is impossible to fully predict how plurality will be marked for a given singular form. At the suprasegmental level, however, the vast majority of German plurals, except plurals ending in <-s> /-s/, exhibit a specific prosodic shape word-finally: a strong-weak pattern, i.e., a sequence of a stressed syllable followed by an unstressed syllable. In other words, German plurals tend to end in a disyllabic trochee. Previous experimental investigations have sought to provide empirical evidence in favor of various formal models. To date, these experimental studies have focused primarily on the segmental composition of plural suffixes. It remains untested—and thus largely unknown—whether the prosodic pattern at the interface between morphology and phonology is an active, productive part of the grammar of first language (L1) and second language (L2) users of German across proficiency levels. We therefore set out to test whether users actively apply the trochaic principle in the production and comprehension of German plural nouns. To this end, we tested L1 German and L1 English-L2 German users across four proficiency levels on a non-word plural elicitation task, in which they produced plural forms for non-words, akin to a *wug*-test; L2 users additionally completed a plural elicitation task with existing German nouns. All users then participated in a grammatical acceptability judgment task, in which they rated German nouns with various incorrect and correct (i.e., SG) plural forms on a Likert scale. L2 learners produced more trochaic plural forms as proficiency increased, and more advanced users showed a stronger correlation between their ratings and plural forms depending on the forms' correctness and prosody. We further analyzed how prosodic patterns varied with morphological context across proficiency levels, before discussing how the data can be accounted for within various models of German plurals.

KEYWORDS

morphophonology, prosody, trochee, German plurals, L1, L2 development, allomorphs, schema

1 Introduction

Nominal plural inflection in Standard German (SG), henceforth German, is highly complex and has spurred many theoretical accounts over the past several decades (e.g., Augst, 1979; Bittner, 1991; Köpcke, 1988, 1993; Mugdan, 1977; Neef, 1998; Trommer, 2021; Wiese, 1996, 2009; Wunderlich, 1999; Wurzel, 1998; *among others*). The (ensuing) debates have attempted to explain the morphology and morphophonology of the exponents, i.e., markers, of German nominal plurals, and how these markers might be acquired. The fundamental difficulty lies in the fact that SG exhibits a large number of plural allomorphs, including six different suffix allomorphs, some of which can also be combined with stem vowel fronting and/or raising, i.e., umlaut,[1] U. As schematized in Table 1, German espouses nine phonetically distinct options for marking nominal plural formation when the suffix choices and the possibility of umlaut are combined, i.e., <-en>, <-n>, <-e>, <-er>, <-s>, -Ø, U+<-e>, U+<-er>, and U+-Ø, leaving aside exceptional cases, such as suppletion.

The pattern of nominal plural data in German in Table 1 presents a challenge for theoretical models as well as for questions of learnability and models of language acquisition. Indeed, it is impossible to fully predict how plurality will be marked for a given singular form at the segmental level. Nonetheless, linguists have identified co-occurrence restrictions between certain suffixes and umlaut, e.g., the suffixes <-n> /-n/, <-en> /-ən/, and <-s> /-s/ can never co-occur with umlaut (Augst, 1979; Wurzel, 1998; see also Trommer, 2021 for discussion, and Archibald, 2022 for an experimental L2 study). In addition, numerous tendencies have been noted regarding the type of plural marking a given singular form is likely to take, depending on features such as noun gender (e.g., feminine nouns are more likely to end in *-(e)n*) or the word-final phonological composition of the singular form (e.g., no explicit ending for masculine and neuter nouns ending in *-el, -en,* or *-er*) (see Section 2 below). At the suprasegmental level, however, the vast majority of German plurals, except those ending in <-s> /-s/, share a specific prosodic shape, namely, they overwhelmingly end in a strong-weak syllable sequence ($\sigma_{strong}\sigma_{weak}$), i.e., a sequence of a stressed and an unstressed, reduced syllable, often analyzed as a disyllabic trochaic foot (e.g., Neef, 1998; Smith, 2004, 2020, 2022; Wiese, 1996, 2009).

Descriptive accounts of German nominal plurals that take syllabic or prosodic structure into account typically agree on this predominant strong-weak disyllabic structure word-finally. Theoretical accounts, however, differ in whether and how they capture this metrical pattern in their formal models (see Section 2). Indeed, most theories and experimental investigations have focused on the segmental composition of plural markers, i.e., the distribution and patterning of the plural suffix exponents, often leaving aside the role of prosody. In this study, we focus on the prosodic structure and whether it might play an active role in the grammar of first language (L1) and second language (L2) German users with regard to plural formation. To this end, we tested L1 and L2 German speakers to examine whether their production and perception of plural forms provide experimental support for the role of prosody in nominal plural formation. For the L2 German users, we present data from four different proficiency levels in a cross-sectional design to test whether and how the production and perception of the prosodic pattern in plural forms changes with increasing language experience. This design allows us to test whether L2 users show the target-like prosodic pattern from the start, and to trace their acquisitional path. A longitudinal or cross-sectional study enables us to examine whether this prosodic pattern emerges suddenly, all-at-once, or whether it develops progressively as learners' language competencies grow.

This article thus presents behavioral, psycholinguistic data that are in line with formal models of German plurals in which the grammar of L1 German speakers, and with increasing proficiency adult L1 English-L2 German learners, contains a prosodic condition for German nominal plural formation. Although users are not explicitly aware of this prosodic requirement, the results of this study demonstrate an overwhelming tendency for plurals to end in a disyllabic trochee, i.e., a sequence of stressed-unstressed syllables word-finally. This prosodic pattern is shown for speakers' productions and in their judgments of the well-formedness of correct (trochaic) and incorrect (trochaic and non-trochaic) plural forms where L1 and—to varying degrees based on proficiency—L2 users rate trochaic forms as more well-formed compared to non-trochaic forms.

This paper is organized as follows. Section 2 provides a brief review of previous theoretical research on German plural formation. Next, Section 3 presents the methodology of the study which tests L1 and L2 German users' prosodic intuitions for German plural formation in (a) a non-word plural elicitation task (and, for L2 speakers only, also a plural elicitation task with existing German nouns), and (b) a well-formedness judgment task. In Section 4, we present the results of the study, including how prosody and morphology interact. Based on these findings, Section 5 answers the research questions and addresses theoretical implications, limitations, and open questions. The paper concludes in Section 6.

2 Background and literature review

2.1 German plurals: rule-based approaches and tendencies

As noted, accounting for German nominal plural inflection is challenging not only because of the large number of plural allomorphs—several suffix allomorphs used with or without stem vowel fronting and/or raising (umlaut)—but also because it does not seem possible to fully predict how a given singular will be marked for plural. Nevertheless, a number of studies have revealed tendencies for nouns to select their plural marker based on specific features and characteristics of words. These features include the gender of a noun and word-final phonological characteristics in the singular form (Augst, 1979; Mugdan, 1977; Köpcke, 1988; Wurzel, 1984; Duden-Grammatik, 1995), especially with respect to the level of sonority of the segments (e.g., Laaha, 2011). Köpcke (1988) also discusses semantic features such as animacy. However, these

1 Note that only the vowels represented by <a, o, u> and the diphthong <au> are able to form an umlaut: /aː, a, oː, ɔ, uː, ʊ, aʊ/ become /ɛː, ɛ, øː, œ, yː, ʏ, ɔɪ/, respectively.

TABLE 1 Overview of nominal plural allomorphs in SG.

Affix type	Umlaut pattern	Example without umlaut	Example with umlaut
-e [ə]	variable (w/ or w/o umlaut)	der Hund—Hund-**e** (*the dog, dogs*)	der Hut—H**ü**t-**e** [u]—–[y] (*the hat, hats*)
-er [ɐ]	always umlaut (U-able V)	das Kind—Kind-**er** (*the child, children*)	das Blatt—Bl**ä**tt-**er** [a]—–[ɛ] (*the piece of paper, papers*)
-en [ən, n̩]	never umlaut	die Frau—Frau-**en** (*the woman, women*)	
-n [n]	never umlaut	die Tasse—Tasse-**n** (*the cup, cups*)	
-Ø (*no overt ending*)	variable (w/ or w/o umlaut)	der Keller—Keller-Ø (*the basement, basements*)	der Apfel—**Ä**pfel-Ø [a]—–[ɛ] (*the apple, apples*)
-s [s]	never umlaut	das Kino—Kinos (*the movie theater, movie theaters*)	

Grayed out cells indicate non-existent plural formations.

tendencies do not always denote the pluralization of a noun in a "deterministic" way that applies in all cases. Most of the time, these tendencies are more "probabilistic" in nature (Szagun, 2001). Examples of deterministic plural rules include (based on Szagun, 2001, *inter alia*):

a) Nouns ending in <-e> ([ə]) form the plural with the suffix <-n> /-n/, e.g., *der/die Bote-n* "messenger-s", *die/die Blume-n* "flower-s", *das/die Auge-n* "eye-s"
b) Non-feminine nouns ending in *-el, -en, -er, -chen*, and *-lein* take the null-suffix -Ø, e.g., *der/die Eimer-Ø* "bucket-s", *das/die Steuer-Ø* "steering wheel-s" (cf. feminine *die/die Steuer-n* "taxes")
c) Nouns ending in a syllable with an unstressed full vowel select for the plural suffix <-s> /-s/, e.g., *der/die Park-s* "park-s", *die/die Oma-s* "grandma-s", *das/die Auto-s* "car-s".

More commonly, however, these so-called "rules" simply express tendencies that hold only with a certain probabilistic likelihood. For instance, feminine nouns form the plural with <-n> /-n/ in 73% of cases, regardless of the word-final segmental characteristics, e.g., *Leiter-n* "ladder-s",[2] while masculine and neuter nouns form the plural with <-n> /-n/ or <-en> /-ən/ in 9% and 4% of cases, respectively (Duden-Grammatik, 1995, see also Szagun, 2001). Thus, plural allomorphy patterns—including the distribution of gender-based plural exponents—show tendencies rather than rules.

2.2 German plurals and prosodic structure

One additional and crucial tendency, however, is often overlooked in general descriptions of German plurals despite a widespread agreement among researchers. It is the observation that syllables and their organization into feet, i.e., metrical or prosodic structures, also play a role in German plural formation (e.g., Eisenberg, 2020; Köpcke, 1988, 1993; Laaha, 2011; Smith, 2020; Wegener, 1995; cf. Binanzer and Wecker, 2020). Indeed, the vast majority of German plurals, except plurals with the suffix <-s> /-s/ (see further details below), end in the specific prosodic shape illustrated in Table 1, namely a word-final sequence of a stressed and unstressed reduced syllable (e.g., Eisenberg, 2020; Neef, 1998; Smith, 2004, 2020; Wiese, 1996, 2009; *among many others*). Assuming that syllables are organized into feet, this generalization can be expressed as plural forms predominantly showing a word-final disyllabic trochee (see below for formal analyses). This generalization holds for plural nouns regardless of the shape of their singular form. For example, all nominal plural forms in Table 1 show a word-final trochee (the period "." marks syllable boundaries): *Hü.t-e* "hats", *Kin.d-er* "children", *Frau.-en* "women", and *Äp.fel-Ø* "apples". This is regardless of the underlying shape of the corresponding singular forms, e.g., non-trochaic singular *Hut* "hat" alongside trochaic *Ap.fel* "apple".

Although this is a strong prosodic *tendency*, there are exceptions. First, as noted above, plurals marked by <-s> /-s/ do not fall under this prosodic generalization; this means that *-s* plurals for singular forms that end in final trochee will show a final trochee in the plural, e.g., *Sto.ry-s* "stories", while those whose singular forms do not end in final trochee will not show a final trochee in the plural, e.g., *Klub-s* "clubs". For all other (monomorphemic) plurals formed using the suffix options in Table 1, the prosodic generalization captures most, albeit not all, plural forms. Consider one group of exceptional cases, namely singulars ending in a specific derivational suffix (e.g., *-ung*, *-chen*, *-keit/-heit,* etc.), which each require a specific plural suffix. For instance, words ending in *-ung* always take the plural suffix *-en*, as in ˈ*Ta.gung.-en* "conference-s", thus yielding non-trochaic plural forms. Words ending in *-chen* always take the plural suffix -Ø, as in ˈ*Mäd.chen-Ø* "girl-s" or ˈ*Vö.gel.chen-Ø* "little bird-s", but may or may not result in a trochaic plural form. Moreover, the suffix *-in*, which marks animate objects as feminine, always requires the *-(n)en* plural ending, as in ˈ*Kun.din.-nen* "fem. customer-s", a non-trochaic plural form.

2 It should be noted that nouns ending in <-e> ([ə]) always form the plural with <-n> /-n/, including feminine nouns ending in <-e> ([ə]).

While it may be argued that the disyllabic trochee is epiphenomenal for plural formation in German, a similar prosodic tendency, i.e., a preference for trochaic forms, has been found in other lexical and morphological classes in German, both historically and in modern times.[3] More generally, it has been proposed that German has a "preference for the trochaic foot" (Féry, 1994, p. 31), and that the trochee is the most important foot, followed—far behind—by the dactyl in German (Eisenberg, 2020). German, just as English, has been labeled a so-called "trochaic" language with regard to word stress, as native German words are most commonly stressed on the penultimate syllable. This can be explained by the fact that most native words have two syllables, of which the final one is reduced and unstressed (Domahs et al., 2014, p. 62).

Even beyond the tendency for German to be trochaic broadly speaking, Eisenberg (2020) proposes that the trochee is the most grammaticalized foot in both inflectional and derivational morphology. This tendency to draw on the trochee is a feature that German has inherited from its Germanic origins where entire lexical classes and morphological functions were reshaped in alignment with the trochaic foot, from high vowel deletion in Old English, Old High German, and Old Saxon nouns and verbs, to Old Frisian Vowel Balance (Smith, 2024). Likewise, the trochee has been found to shape plurals and diminutives in Dutch (Booij, 1998; van der Hulst and Kooij, 1998; Smith, 2004), as well as in various German dialects (Wiese, 2009; Smith, 2020). Similar findings are noted for other morphological processes including noun shortening, e.g., *Universität—Uni* "university" (e.g., Itô and Mester, 1997; Schuhmann, 2015). In all of these cases, the critical importance of the trochee (whether syllabic or moraic, i.e., based on syllable weight) in shaping these morphophonological processes is demonstrated by a greater prevalence of the trochee in these specific forms than found more generally in the lexicon.[4] Whether this is epiphenomenal or teleological in nature (for instance via templates) is beyond the scope of the current study. Nevertheless, these phenomena, historical and modern, highlight the persistent association of the trochee with morphological functions since the earliest records of German.

Reflecting on the modern German language, Féry (1994) likewise argues that a disyllabic trochee is built into both inflectional and derivational morphology whenever possible. To illustrate this, adjectival inflections are largely trochaic such that monosyllabic adjectives like 'green' become trochaic forms such as *'grü.n-e / 'grü.n-er / 'grü.n-es* "green, fem. or pl. / green masc. / green, neutr." in attributive position and indefinite contexts. In *definite* contexts in attributive position, adjectives end in *-e*, e.g., trochaic *'grü.n-e*; all adjectival plural forms in attributive position end in *-en*, e.g., trochaic *'grü.n-en*. Note, however, that predicative adjectives are uninflected; thus, typical monosyllables like *'grün* remain non-trochaic in this context. Despite this predominance of trochaic forms in adjectival inflection, variation also exists. Disyllabic adjectives ending in *-er* and *-en* such as *'hei.ter* "cheerful" or *'troc.ken* "dry" allow both trochaic and dactylic inflected forms, e.g., *'hei.tres* and *'hei.te.res* "cheerful, neutr.", respectively (Eisenberg, 2020, p. 149).[5]

Stems have typically been described as trochaic in German. Yet, stems can be analyzed as trochaic either with respect to syllables (disyllabic) or moras (bimoraic), a different weight unit. Bimoraic stems are heavy, binary monosyllables (Féry, 1994). Thus, German stems can be said to be mostly trochaic if this allows for both disyllabic and bimoraic trochees. Another nominal inflection that shows some trochaic forms is genitive singular (Eisenberg, 2020). Yet, only weak masculine (and some neuter) nouns such as *Mensch* "human" require a trochaic form for the genitive singular and the plural form, i.e., *des/die 'Men.sch-en* (Eisenberg, 2020, p. 149). Besides weak masculine nouns, only nouns in sibilants require trochaic genitive singular forms due to phonotactic reasons, e.g., *des Ti.sch-es* "table, gen. sgl." (Eisenberg, 2020, p. 149). Otherwise, genitive singular forms do not have to be trochaic—in contrast to the basic pluralization pattern.

In sum, German is an overwhelmingly trochaic language for which the trochee has been analyzed as an unmarked prosodic constituent (Féry, 1994). While many lexical forms within inflectional and derivational morphology—in addition to plain stems—can be described as trochaic, the disyllabic trochee has been considered a "necessary condition" ("notwendige Bedingung") only for plurals (Eisenberg, 2020, p. 149). Overall, this indicates that the prosodic trochaic pattern in the plural formation of German nouns is firmly grammaticalized and applies even more consistently than in other sub-parts of the German language system.

2.3 The importance of prosody in L1 and L2 acquisition

Work on the acquisition of German plural formation by children with specific language impairment (SLI) has shed light on the role of prosodic requirements for plural formation and

3 In German, the trochee can take the form of bimoraic monosyllables, e.g., *Heu* "hay", or disyllables, e.g., *'Va.ter* "father" (Féry, 1994, p. 1). Of these two possible trochees, only the disyllabic trochee is associated with German plurals.

4 Indeed, McCarthy and Prince's (1996) work on Prosodic Morphology first made the connection between prosodic units like the foot in a variety of morphological processes, demonstrating that the foot can be influential not only in stress placement, but in shaping morphological patterns in a language. Psycholinguistic studies such as those discussed in Section 2.3 substantiate this intersection between prosody and morphological processes outlined in various theoretical approaches.

5 Similar patterns can be seen in possessive determiners. For instance, when the inflectional suffix *-e* is added to *'eu.er* "your, pl. non-fem", the resulting inflected form is trochaic, i.e., *'eu.re* "your, fem. sg./pl." Just as with adjectives, variation can be seen with certain forms. For example, when *'eu.er* or *'un.ser* "our" receive the inflectional suffixes *-e*, *-es*, or *-er*, the resulting inflected forms may or may not include the stem-final *-e-*: both *'eu.e.re, 'eu.re* and *'un.se.re, 'uns.re* are possible. With the inflectional suffixes *-em* and *-en*, either the stem-final or the suffix *-e-* might not surface in the inflected forms, e.g., *'un.se.ren / 'uns.ren / 'un.sern* (Duden-Grammatik, 1998, p. 336). Thus, certain inflected possessive determiners also show variation between trochaic and dactylic forms. Eisenberg (2020) further proposes that a dactyl occurs in inflected comparative forms such as *'klei.ne.ren*, in which an *-en* inflectional suffix is added to the comparative form *'klei.ner* "small, comp."

plural acquisition (Kauschke et al., 2013) and the role of prosody in morphosyntactic difficulties more generally (Domahs et al., 2013). In a plural elicitation task with existing and non-words, Kauschke et al. (2013) showed that German-speaking children with SLI (mean age 7.5) made more mistakes when forming plurals and produced fewer "prosodically optimal" plurals compared to age- and vocabulary-matched children without SLI. These findings underscore that the children with SLI had reduced sensitivities for the prosodic requirements of German plural marking. Kauschke et al. (2013, p. 574) argue that morphological processes are influenced by prosody, and that even more fundamentally, "morphological acquisition may in part be linked to the fact that children's acquisition of grammatical morphemes is closely tied to the development of prosodic representations (Demuth, 2009)." In another clinical study with a German-speaking patient showing primary progressive aphasia, Domahs et al. (2017) report that the patient was more likely to interpret a stimulus as a plural form if it included any of the cues of Köpcke's schema model (discussed below) and if it conformed to the "optimal prosodic plural form" (Domahs et al., 2017, p. 206, *translation KS*). Such studies point to the critical role of prosody in German for shaping morphological processes, representations, and acquisition, including plural formation.

The overall importance of prosody in L1 acquisition of morphology has been shown cross-linguistically. Demuth (2009) synthesizes cross-linguistic research on child L1 acquisition of prosodic morphology in support of the so-called Prosodic Licensing Hypothesis. The findings support the notion that frequently reported variability in the production of grammatical morphemes can be explained by the children's developing prosodic representations. For instance, Demuth (1994) argued that children acquiring the Bantu language Sesotho initially produce specific morphological prefixes only when they are footed into a disyllabic foot. It was proposed that children would first acquire morphemes prosodified as part of a foot, such as internal clitics. Further data from Gerken (1996) and Demuth and McCullough (2009) on children learning English support this approach, where it was shown that the children were more likely to produce determiners when they were part of a prosodic foot, i.e., prosodically licensed. Interestingly, cross-linguistic comparisons of L1 acquisition from English, German, Spanish, and Italian (Lleó and Demuth, 1999) show that determiners are produced earlier by children learning Romance languages than children learning Germanic languages. In Spanish, determiners are cliticized to the nouns as proclitics, while in German, a determiner is prosodified as a separate foot when it occurs in its full, unreduced, and un-cliticized form (Lleó and Demuth, 1999; Wiese, 1996). Thus, Lleo and Demuth's analyses provide further evidence for the proposal that the appearance of grammatical inflectional morphemes depends on prosodic licensing during child L1 acquisition.

Goad and White (2019) have put forth a similar proposal in which prosody constrains the acquisition of grammatical morphemes during L2 development. According to their Prosodic Transfer Hypothesis (PTH), L2 learners' non-target-like forms can be accounted for by the transfer of their L1 prosodic constraints. In their current revised PTH, they further argue that not only the production but also the comprehension and processing of L2 grammatical morphology is constrained by the transfer of learners' L1 prosodic structures to their L2.

2.4 Formal accounts and acquisition/processing models of German plurals

A wide variety of theoretical analyses and models have been proposed to account for German nominal plural formation. A large number of accounts have focused on the plural allomorphs and restrictions on their co-occurrence with stem vowel changes, i.e., umlaut, while the role of prosody varies from "not discussed" to central. The following discussion presents an overview of the approaches, ranging from the debate about dual- vs. single-route models, to schematic approaches, and generative phonological analyses.

2.4.1 Models of L1 language acquisition and processing

In terms of modeling L1 acquisition, much of the debate about L1 German plural acquisition over the past two and a half decades has centered on whether the acquisition, processing, and use of the many German plural forms are best explained by a dual- or single-route model. The generativist perspective tends to favor an intrinsic, dual-mechanism perspective (Clahsen, 1999; Marcus et al., 1995), which is modeled on the inflectional system in English. This perspective suggests that, on the one hand, regular forms are generated and comprehended via a rule-based mechanism that combines or decomposes the word-stem and an affix ("stem+affix"). Irregular forms, on the other hand, are memorized and stored holistically as discrete lexical entries in the mental lexicon, but they "pattern according to phonological similarities" (Köpcke et al., 2021, p. 4). However, attempting to define what counts as the "regular" plural form in German, or whether there even is a "regular" plural form in German, is contentious. Meanwhile, analyses of English straightforwardly map the most frequent forms onto the "regular" forms.

For German, proponents of the dual-route model initially argued that the *-s* plural is the "regular" plural, which is, despite its low frequency of about 4% (Szagun, 2001) to 6% (Zaretsky et al., 2013), processed fastest among all plural markers in German. Dual-route proponents have also argued that the *-s* plural is the "regular" plural based on Clahsen's original claim that *-s* was the most frequently overgeneralized plural marker in children; this claim, however, has since been refuted (see Laaha et al., 2006, p. 277 for a brief overview). According to this dual-route model, all other plural forms are irregular plurals (Clahsen, 1999; Clahsen et al., 1992; Marcus et al., 1995; for empirical evidence, see Beretta et al., 2003; Clahsen et al., 1997; Sonnenstuhl et al., 1999; for an overview, see Clahsen et al., 2003; see also Köpcke et al., 2021). Slightly later versions of this dual-mechanism model concede that the *-n* plural for feminine nouns ending in schwa has a special status among all irregular plurals due to its highly predictable nature (Penke and Krause, 2002; Sonnenstuhl-Henning, 2010; cf. Köpcke et al., 2021, p. 4). Other approaches include a race-model variant of the dual-route view (e.g., Baayen and Schreuder, 1999; Schreuder and

Baayen, 1995; see also Pinker and Ullman, 2002; Laaha et al., 2006), which assumes that both a parsing and a retrieval route operate at the same time. Thus, according to such race-model variants, both memory and rule-application processes apply in parallel. The winning route depends on the characteristics of the individual inflected words, such as their frequency, their morphology, the transparency of their phonology and semantics, and their lexical neighborhood effects (cf. Laaha et al., 2006, p. 273).

Conversely, proponents of the single-route model argue that children construct grammatical structures from their input and that all forms are processed in the same way, via one mechanism rather than two. These include the schema-based models of Köpcke et al. (discussed below), constructivist theories, and connectionist models (see Köpcke et al., 2021, p. 5 for a brief overview). For instance, Hahn and Nakisa (2000) provide simulation and experimental data in support of single-route connectionist systems. Laaha et al. (2006) present results from a plural elicitation study with children that can be explained by either a single-route model or a race variant of the dual-route model: Rules that are more productive and transparent are more likely to be used than the lexical route (Laaha et al., 2006, p. 299). Their data also support their hypothesized role of the degree of productivity in the speed of acquisition of the plural form of a given noun.

This literature has typically focused on factors that constrain the acquisition of the various plural suffix options as outlined in Table 1 (i.e., *-e, -er, -en, -n*, *-Ø*, *-s*), with or without umlaut. To the best of our knowledge, so far, the prosodic pattern in German plurals has not played a central role in this debate of single- vs. dual-route models and their variants. However, syllabic or prosodic aspects are at least mentioned—or take center-stage—in the specific schematic and generative approaches discussed below.

2.4.2 Schematic approaches

One of the most prevalent models for German nominal plural formation is the "prototype-based 'schema' model" by Köpcke and colleagues (e.g., Köpcke, 1988, 1993, 1998; Köpcke et al., 2021; Köpcke and Wecker, 2017). In this schema-based model, specific plural forms signify plurality more robustly than others because certain features or cues, e.g., polysyllabicity, umlaut, the determiner *die* ("pl." or "fem. sg."), and word-final characteristics (see Köpcke, 1988 for particulars), or combinations of these cues, make a word appear more like a typical plural. Plural markers are analyzed as "abstract schemata" consisting of these cues as opposed to "individual morphemes" (Köpcke, 1988, p. 330). Arguing that language users engage in a "schema-matching process", Köpcke proposes that individual plural forms differ in how strongly they are associated with the function of plurality due to differences in the "cue strength" of each feature's salience, type frequency, cue validity, and iconicity. For instance, in the case of the plural markers introduced above, many of these markers—e.g., the suffixes <-e> /-ə/ and <-er> /-ɐ/—can also occur as word-final segments on singular forms, e.g., *Leh.rer* "teacher" or *Lö.we* "lion". It is argued that the link between form and function becomes stronger when a specific form occurs particularly often in a plural form (type frequency) while occurring rarely in a singular form (validity). To provide a concrete example, in this proposal, nouns ending in <-en> indicate plurality more strongly than nouns ending in <-e>. This is because word-final <-en> occurs more frequently in the plural than in the singular, while singular forms ending in <-en> are uncommon. This model has recently been used to describe the acquisition of plural forms in L1- and L2-speaking children in German primary schools (Binanzer and Wecker, 2020; Köpcke and Wecker, 2017; Wecker, 2016). The latest instantiations of schematic approaches now also reflect the increased attention paid to the prosodic regularity of German plurals as they incorporate the trochee as the optimal prosodic form for plurals, rather than just polysyllabicity vs. monosyllabicity or number of syllables. This is the case in a clinical case-study with an individual with primary progressive aphasia (Domahs et al., 2017) and in a study on child L2 German acquisition among elementary school students in Germany (Köpcke and Wecker, 2017). This last study led to the most recent elaboration of the schema model (Köpcke and Wecker, 2017; Wecker, 2016) with "schema-pairs" (Köpcke et al., 2021, p. 7), which are intended to capture the paradigmatic relation between the two first-order schemas for singular and plural, respectively. As an example, a second-order "super"-schema (Köpcke and Wecker, 2017, p. 85) accounts for the connection between singular schemas with the feminine article and an <-e> ending, on the one hand, and plural schemas with the plural article and the <-en> suffix, on the other hand (Köpcke and Wecker, 2017, p. 85).

2.4.3 Generative analyses

To account specifically for the prosodic pattern evident in German plural formation, several formal accounts have been proposed from a generative linguistic perspective. In particular, early morphophonological formalisms describe the prosodic shaping of plural formation in German in terms of edge effects; regardless of the singular form, the right edge of the plural must align with a disyllabic trochee. Drawing on Prosodic Morphology (McCarthy and Prince, 1996, et seq.), Smith (2004, 2020) analyzes plurals using a prosodic, foot-based template that requires plurals to end in a disyllabic trochee, i.e., (σσ)#. In Smith's analysis, the right edge of plurals is mapped to a complex prosodic template consisting of a disyllabic trochee with a final schwa-syllable (a so-called "schwallable", Booij, 2002). Thus, if a singular form is already trochaic, e.g., *ˈLeh.rer* "teacher" or *ˈLam.pe* "lamp", then the plural ending must be non-syllabic, i.e., *-Ø* or *-n*, respectively, so as to not disturb the trochee, thus *ˈLeh.rer-Ø* "teacher-s" or *ˈLam.pe-n* "lamp-s". However, if the singular form does not end in a trochee, e.g., *Tag* "day", *I.ˈdee* "idea", or *Feld* "field", then the plural ending is syllabic, i.e., *-e, -en,* or *-er*, respectively (*ˈTa.g-e* "day-s", *I.ˈde.-en* "idea-s", or *ˈFel.d-er* "field-s"), to satisfy the trochaic template. While Smith views templates as tendencies rather than absolutes, she concedes that exceptional forms such as *ˈKun.din.-nen* discussed above reveal the conflict between prosodic tendencies and morphological requirements; in these cases, morphological demands can take presedence over prosody, thereby upsetting the prosodic template.[6] Thus, in multimorphemic stems, the trochaic template can be outweighed

6 In Smith's account, the prosodic tendency would be outweighed by the demands of morphological suffixes such as *-in* and *-ung* to select specific plural suffixes.

by morphological concerns as the main stress remains on the stem syllable (e.g., ˈ*Kö.ch-in,* ˈ*Kö.ch-in.-nen* "female cook-s").

Building on this approach, Schuhmann and Putnam (2021) have also recently modeled the prosodic template for German plurals within a Distributed Morphology (DM) approach. There, a prosodic boundary for nominal plurals is established directly in the syntax through the generation of a Prosodic Word ω. Crucially, in this account, the prosodic unit for plural exponents is established only for a particular feature configuration in the syntactic structure (n[+pl]). Thus, in this first step, the syntax establishes just a prosodic boundary, devoid of any segmental information; as is typical in DM, the phonological realization of morphosyntactic features is achieved through the insertion of Vocabulary Items at Spell-Out (when morphosyntactic features are mapped onto phonological content). Similar criticisms as above can be raised here about how exceptions are modeled formally. While approaches proposed by Schuhmann and Putnam (2021) and Smith (2004, 2020) did not address this question, one option could include word-level-specific tags or features—perhaps inherent in the derivational suffixes discussed above—that would mark individual words such as multimorphemic stems as exceptional. This could also include marking some words differently which have been borrowed but not integrated into German. The feature configuration with this additional component (tag or feature) would no longer trigger the necessary prosodic boundary for trochaic plural exponents.

Other generative linguistic perspectives have analyzed the prosodic generalization for German noun plurals with Optimality Theory (OT, e.g., Prince, 1993/2004). In Wiese's (2009) account, the prosodic pattern was formalized as the markedness constraint "Trochee" interacting with other constraints. For this analysis, Wiese (2009) postulates constraints that are specific to plural or singular forms of the morphological paradigm. In particular, the central markedness constraint Trochee is specific to plural forms, requiring that they end in a trochee. Faithfulness constraints are likewise specific to certain parts of the morphological paradigm, i.e., singular vs. plural forms. Overall, at the core of this paradigmatic analysis, Trochee is ranked higher than a faithfulness constraint against segment insertion in plural forms (DEP-SegPl) but lower than a faithfulness constraint against segment insertion in singular forms (DEP-SegSg). Clearly, one central drawback of such an analysis is that the crucial constraints are assumed to be specific to either the plural forms or the singular forms of the morphological paradigm, thus reducing the explanatory factor in this model. Here, to achieve the prosodic pattern in plurals, very specific constraints have to be stipulated for specific parts of the paradigm. Since in OT, all constraints are universal and violable, the very specific formulation of constraints made to fit the German plural data makes them less promising as constraints of the universal set of OT constraints.

In another recent OT account of German plurals, Trommer (2021) argues that the prosodic shape of German plural forms arises as an Emergence-of-the-Unmarked effect (TETU, McCarthy and Prince, 1994). Trommer's model captures both the various allomorphs and the prosodic pattern of German plurals in a unified account in which he combines older phonological proposals with recent developments within OT. Specifically, his account is situated within Stratal OT (Kiparsky, 2015; Bermúdez-Otero, 2018) and Autosegmental Colored Containment Theory (Revithiadou, 2007; van Oostendorp, 2008; Zimmermann, 2014, 2017). Notably, Trommer (2021) does not assume a prosodic foot template for plurals, but argues that this pattern arises as a TETU effect (McCarthy and Prince, 1994). This is achieved in large part by resurrecting an older proposal that the representation of schwa-affixes constitutes a defective segment-sized phonological unit which, according to Trommer (2021, p. 605), corresponds to the C- or X-slot in Wiese (1986), Hall (1992), and Noske (1993). Given that faithfulness constraints protect incomplete segments less strictly, the templatic metrical effects emerge as a TETU phenomenon. By proposing that there is only one general underlying plural affix, the account also achieves an analysis in which the prosodic pattern holds for all plural forms. This single, general underlying plural affix consists of an underspecified root node and a floating coronal feature. If this general, phonologically underspecified, featureless affix is not combined with any other suffixes, Trommer suggests that the phonology will spell it out as schwa, the most unmarked segment in German. The general affix can also be combined with more restrictive plural affixes, either a feminine or neuter plural marker, thereby specifying additional segmental information, such as a nasal or a pharyngeal feature. Finally, drawing on the pre-OT literature, Trommer (2021) proposes that umlaut is triggered by affixes containing a floating feature which conditions vowel fronting, here a floating coronal feature. Overall, Trommer (2021) consternates that, unlike Wunderlich (1999), most proposals cannot achieve a natural account of the prosodic regularity, the allomorphs, and the "implicational relationships between umlaut and suffixation" (p. 650).

Thus, prosody figures prominently in many—albeit not all—formal models of German plural formation, including in Köpcke and colleagues' schema-based accounts. Here, in particular, clinical studies have assigned a more central role to the trochaic patterns within schematic accounts of plurals when discussing a case study of a patient with aphasia (Domahs et al., 2017), and child language acquisition in impaired and non-impaired language development contexts (Kauschke et al., 2013), as discussed above. This work suggests that prosody is the constraining framework within which plural morphology is built. The PTH introduced above similarly proposes that L2 prosodic-phonological representations are the basis for inflectional morphology in L2 development. Thus, we set out to test whether and how these prosodic representations might develop in adult L2 acquisition in a foreign language learning context.

2.5 Filling the gap: testing prosodic sensitivity in German plurals

In this work, we focus on whether the prosodic pattern that has been established in descriptive and theoretical analyses of German plural forms can be considered a productive part of the grammatical system of L1 German users, and how it might develop in L1 English-L2 German users across language proficiency levels. This aspect of the study aims to address if and when L2 German users develop intuitions about a *potential* prosodic condition for German nominal plural formation. In other words, we ask whether this metrical pattern discussed for German plurals is merely the result of historical developments in German (see Smith, 2022 for an overview), or whether it is now an existing but perhaps non-productive pattern in the German grammatical and/or lexical system. One way to test productivity is to assess how speakers deal with non-words. While the trochaic pattern occurs in many existing German words as well as in integrated loanwords such as *Ga.'ra.ge-n* "garage-s", to our knowledge, empirical studies of healthy adult language users have not yet focused on the role of prosody in the production of novel plural forms or in well-formedness judgments by (healthy) language users and during adult L2 acquisition (but see Vogt, 2016 for a pilot study with L2 German learners).

To this end, we developed two tasks to assess the role of prosody in productive and receptive German plural formation. The first task (L1 and L2 German participants) was a plural elicitation task to examine the role of prosody in the production of plural forms in non-words. Just for the L2 German users, we also included a plural elicitation task with existing German words.[7] The third task (L1 and L2 German participants) was then a grammatical acceptability judgment task (GAJT), which examined the role of prosody in the ratings of existing and made-up plural formations of existing German words. Note that both English and German have been characterized as languages containing the prosodic constituent "foot". Similar to German, evidence supports the claim that English "builds binary trochaic feet" (Garcia and Goad, 2021, p. 20; see also Domahs et al., 2014), an aspect to which we will return in the discussion. Thus, our tasks were designed to answer the following research questions respectively:

- Task 1: *Plural elicitation task—non-words*
  - RQ1a: To what extent do L1 German and L1 English-L2 German speakers produce trochaic plurals for non-words?
  - RQ1b: Does L2 German speakers' use of trochaic plural forms for non-words increase with increasing proficiency?
- Task 2 (L2 users only): *Plural elicitation task—existing words*
  - RQ2a: How accurately do L1 English-L2 German speakers produce plural forms for existing German nouns? Does accuracy improve with increasing proficiency?
  - RQ2b: To what extent do L1 English-L2 German speakers produce trochaic plural forms for existing German nouns? Does the number of trochaic plurals increase with increasing proficiency?
  - RQ2c: To what extent do L1 English-L2 German speakers produce trochaic plurals forms for incorrectly produced existing German nouns?
- Task 3: *Plural well-formedness rating task*
  - RQ3a: To what degree do well-formedness judgments of L1 German and L1 English-L2 German users correlate with the plural forms' prosodic structure? I.e., do users rate plural forms that are *non*-trochaic as less well-formed compared to forms that are *trochaic*, both within correct and within incorrect plurals?
  - RQ3b: Do L2 German users' ratings of the well-formedness of correct and incorrect (trochaic and non-trochaic) plural forms change with increasing proficiency? I.e., is the correlation between plural type (correct/incorrect by trochaic/non-trochaic) stronger in more proficient users?

The methodological details of the study tasks, the procedures, and the participants are described in the following sections.

3 Empirical study: probing prosodic intuitions of L1 and L2 users

3.1 Participants

As noted, data were collected from two major groups of participants: (1) L1 users of German and (2) L1 English-L2 German users. A total of 54 (21 male, 33 female) L1 German users participated, of which we will present the data from 30 (10 male, 20 female) participants below 40 years old (mean 26, range 18–38), which is closer to the age-range of the L2 participants (mean: 21.5, range 18–28). A total of 98 (48 male, 50 female) L1 English-L2 German users participated in the study (however, $n = 97$ for Task 3). All participants in this group were enrolled in German classes at a large private university in the U.S. Here, we use semester-level (plus immersion experience) as a proxy for proficiency. L2 participants were assigned to one of four proficiency group levels based on the number of semesters they had studied German. Twelve (m = 4, f = 8) students were in their 1st or 2nd semester German classes ("Year 1"), 15 (m = 2, f = 13) participants were in their 3rd or 4th semester German classes ("Year 2"); 10 (m = 5, f = 5) participants were enrolled in a 5th semester course or above and had spent less than 6 months in a German-speaking country ("Year 3"). Sixty-one participants (m = 37, f = 24) were enrolled in a 5th semester course or above and had at least 6 months of immersion experience ("Year 3 + Imm."). These participants comprised the largest group because of the tendency for German students at this university to have spent 16–22 months in a German-speaking country.

3.2 Methodology

3.2.1 Task 1: non-word plural elicitation task

The first task was a non-word plural production task, similar to a *wug*-test (Gleason, 1958). This task was intended to examine L1 and L2 German users' prosodic intuitions for plural formation by testing how frequently they would produce trochaic plural forms. The non-words were presented with an indefinite article (*ein/e* "a, masc. or neutr. / fem.") and participants were asked to provide a plural form for each word. To respond, participants were asked to both say out loud and type how they would form the plural for each given singular non-word. To help facilitate the plural formation, they were given a number to use for the plural production. For

7 Note that this is part of a larger project on the role of the trochaic pattern in German plurals, its acquisition, and how it might be taught to L2 learners (e.g., Schuhmann and Smith, 2022, accepted).

example, participants saw and heard *ein Schliemo* ("a Schliemo") and responded *3 Schliemos* (or other plural options).

In total, participants responded to a total of 76 non-word tokens. Some non-words were drawn from other studies on German plurals (Hahn and Nakisa, 2000; Kauschke et al., 2013; Köpcke, 1998), while others were created by the authors. Non-words were created so as to respect German phonotactics and co-occurrence restrictions between suffixes and stem-vowel changes (Augst, 1979; Wurzel, 1998; *see above*). Non-words were distributed across seven categories, which are presented with example non-words for each category in Table 2.

Of these, all but one category, namely the tokens ending in an unstressed syllable with a full vowel, e.g., *Schliemo,* included non-words that were expected to form their plural by applying the prosodic principle. These vowel-final tokens were expected to form plurals using *-s*, which is an ending that does not have to conform to the prosodic principle (see Section 2.2 above); however, since the singular forms are trochaic, an *-s* suffix would still lead to a trochaic plural form.

Effectively, for monosyllabic and polysyllabic words with final stress, participants were expected to add a syllabic suffix (<-en> /-ən/, <-e> /-ə/, or <-er> /-ɐ/, respectively) to achieve a word-final trochee. For the remaining words (words in various schwa-final-syllables), participants were expected to add non-syllabic suffixes (<-n> /-n/, -Ø, or <-s> /-s/).

3.2.2 Task 2: real word plural elicitation task

This task was presented only to the L2 participants as a pre-task to examine their accuracy with and prosodic intuitions for plural formation for existing words in German. This task mirrored Task 1, the plural elicitation task with non-words, except that it was conducted with existing words from the German lexicon. Here, participants responded to a total of 151 real word stimuli, with 122 critical stimuli and 29 fillers. Fillers included words with suffixes, e.g., *'Lös-ung* "solution", or words with stressed prefix-like elements, e.g., *'Ab-schied* "farewell". The critical stimulus words were distributed across eight categories, which are presented with example stimuli in Table 3.[8] Notably, each participant responded to 106 critical stimuli where the prosodic principle was expected to take effect.

Just as in Task 1, the words in all but one category were chosen because each of them forms its plural by applying the prosodic principle. Thus, tokens ending in an unstressed full vowel, e.g., *Auto* "car", were included here as well. These forms are not guided by the prosodic principle but form plurals in *-s* in SG. Even when participants do not build accurate plural forms, this task allowed us to test whether L2 users apply the prosodic principle when forming plurals. Thus, to achieve a word-final trochee, participants would need to add a syllabic suffix to form plurals for monosyllabic and polysyllabic words with final stress. For other words (words in open or closed schwa-final-syllables), participants were expected to add non-syllabic suffixes (including the null-suffix) to keep a word-final trochaic word form in the plural.

3.2.3 Task 3: grammatical acceptability judgment task (GAJT)

Task 3 was a grammatical acceptability judgment task (GAJT) to examine L1 and L2 users' intuitions of plural formation by testing whether the (lack of a) trochaic pattern contributes to the well-formedness judgments. To this end, participants were presented with various plural forms of real German words, including both correct and incorrect, i.e., non-standard German plural forms. For instance, the task included *der Schlüssel—die Schlüssel* (correct) "the key—the keys" as well as incorrect/non-standard *die Schlüssels, die Schlüsseln*, etc. Participants saw (and heard) the singular form with the definite article (*der*, *die*, *das*, "the, sing. masc., fem., neutr."), followed by the plural form twice, once with the definite plural article (*die*, "the, plural"; no gender distinctions), and once with *viele* "many", e.g., *der Schlüssel—die Schlüssels, viele Schlüssels* "the key—the keys, many keys".

After reading and hearing the singular and plural forms, participants were asked to repeat the plural form with *viele* "many", and then asked to rate the well-formedness of the provided plural forms on an 8-point Likert Scale, which was visualized with an arrow pointing in both directions. The left side of the arrow, numbers 1-4, were presented in green, while the right side, numbers 5-8, were presented in red. Thus, there was no mid-point and no neutral point on the scale; participants were asked: "Wie GUT oder SCHLECHT klingt diese Pluralform?" ("How GOOD or BAD does this plural sound?"), whereby *good* was presented in green and *bad* in red. The options on the Likert Scale were labeled with the phrases "Für mich klingt diese Pluralform____" ("For me, this plural sounds ______"). The values spanning 1-2 and 3-4 received the green labels "SEHR GUT" ("REALLY GOOD") and "EHER GUT" ("SOMEWHAT GOOD"), respectively, while 5-6 and 7-8 received the red labels "EHER SCHLECHT" ("SOMEWHAT BAD") and "SEHR SCHLECHT" ("REALLY BAD"), respectively.

This task included 179 tokens. This included the correct and various incorrect, i.e., non-standard plural forms for 42 existing lexemes of the German lexicon.[9] Crucially, incorrect forms included trochaic and non-trochaic forms. For example, for the stimulus *der Schlüssel* "the key", the correct plural form *die Schlüssel*, as well as the incorrect plural forms

8 *Steuer* occurred as feminine ("tax") and neuter ("helm, steering wheel"); these real words differ in their SG plural forms, *Steuern* and *Steuer*, respectively. The singular *Motor* "motor" occurred with two stress patterns, thus in two categories: final stress and stress shift. *Kaffee* "coffee" also occurred with two stress patterns (initial and final stress) in the singular, but was always part of the vowel-final category.

9 Note that this task was designed to focus on suffixes, and avoided, for the most part, testing plural marking by means of umlaut. Two of the stimuli form the plural with umlaut in SG (*Garten—Gärten* "garden—gardens", *Ofen—Öfen* "oven—ovens"), and all plural forms presented in the task included umlaut for both of these stimuli. Additionally, five nouns (*Mädel* "girl", *Löffel* "spoon", *Schlüssel* "key", *Schüssel* "bowl", *Tür* "door") had an umlaut in the orthography of the singular and the plural form in SG, which was kept for all plural forms presented in this task.

TABLE 2 Categories and examples of tokens used in Task 1 (*Plural elicitation task—non-words*).

Higher-level categories (singular form)	Categories (singular form)	Sub-categories (singular form)	Examples	Application of prosodic principle expected?
Final schwa-syllable ($n = 26$)	**Disyllabic, final schwa-syllable** ($n = 26$)	*-er* ($n = 7$)	*'Zolger, 'Treiker*	Yes
		-el ($n = 7$)	*'Spotel, 'Wenfel*	
		-e ($n = 7$)	*'Dalle, 'Jechte*	
		-en ($n = 5$)*	*'Tefen, 'Zaupen*	
No final schwa-syllable ($n = 50$)	**Monosyllabic** ($n = 26$)	*NA*	*'Kland, 'Jent*	
	Multisyllabic, no final schwa-syllable ($n = 24$)	Final stress ($n = 19$)	*Pintala'kor, Flasi'tar*	
		Final full vowel ($n = 5$)	*'Schliemo, 'Doscha*	No

Two words each from the *-er*, *-el*, and *-e* classes and eight from the monosyllables were presented with both *ein* "a, masc./neutr." and *eine* "a, fem." (*Words ending in *-en* are never feminine in SG.)

TABLE 3 Categories and examples of tokens used in Task 2 (*Plural elicitation task—existing words*).

Higher-level categories (singular form)	Categories (singular form)	Sub-categories (singular form)	Examples	Application of prosodic principle expected?
Final schwa-syllable ($n = 58$)	**Disyllabic, final schwa-syllable** ($n = 58$)	*-er* ($n = 17$)	*'Lehrer, 'Leiter*	Yes
		-el ($n = 15$)	*Ar'tikel, 'Kugel*	
		-e ($n = 17$)	*'Wolke, 'Name*	
		-en ($n = 9$)*	*'Kuchen, 'Wagen*	
No final schwa-syllable ($n = 64$)	**Monosyllabic** ($n = 19$)	*NA*	*'Berg, 'Lied*	
	Multisyllabic, no final schwa-syllable ($n = 45$)	Final stress ($n = 24$)	*Universi'tät, Stu'dent*	
		Nouns with stress shift ($n = 5$)	*'Autor, Di'rektor*	
		Final full vowel ($n = 16$)	*'Auto, 'Opa*	No

The stimuli also included all three genders (*der/die/das*: "the, masc./fem./neutr."), except for nouns with stress shift (all masculine). (*Words ending in *-en* are never feminine in SG.)

**Schlüssele* and **Schlüsseln* were presented. The categories used in this GAJT task, as well as example words for the lexical items and incorrect tokens, are shown in Table 4. These categories were meant to parallel those in the two plural elicitation tasks.

To summarize, this GAJT task was designed to test whether the prosodic shape, here trochaic vs. non-trochaic patterns within correct and incorrect plural forms, correlates with the well-formedness ratings of plural forms within the groups of L1 and L2 German participants.

3.3 Procedure

The three tasks were conducted using E-Prime (version 2.0). For each task, stimuli were presented one by one to participants visually on a laptop screen and auditorily via headphones at the same time. The auditory presentation of the stimuli was intended to ensure that all participants could hear the stress pattern. These tokens were recorded by a female speaker in Germany who grew up as a monolingual L1 German speaker. She pronounced the real words, incorrect plural forms, and non-words in a natural way and at a natural speaking rate.[10] The L2 participants began with Task 2, the plural elicitation task with real words. They then followed the same procedure as the L1 participants, completing Task 1 (plural elicitation task with non-words) followed by Task 3 (GAJT). At the end, each participant also completed a brief language background questionnaire.

3.4 Analyses

The research questions for all three tasks focus on the prosodic patterns in the elicited and provided plural forms. To this effect, the analyses and results presented here primarily focus on prosody, as well as a potential progression within the levels of L2 German users. For the non-word plural elicitation task (Task 1), results focus on whether the plural forms produced show a word-final trochaic form, which is analyzed as a binary categorical variable (trochaic vs. non-trochaic). For the plural elicitation task with

10 The speaker was instructed by author KS (an L1 German speaker) to pronounce all words as if they were real words. KS examined the recordings and chose those that sounded most natural in her judgment.

TABLE 4 Categories of word types used in Task 3 (*GAJT: grammatical acceptability judgment task*).

Higher-level categories (singular form; *n* lex. items; *n* tokens)	Categories (singular form; *n* lex. items; *n* tokens)	Sub-categories (singular form; *n* lex. items; *n* tokens)	Example lexical items	Example non-standard plural forms (tokens)
Final schwa-syllable ($n = 20$; 68)	**Disyllabic, final schwa-syllable** ($n = 20$; 68)	*-er* ($n = 6$; 24)	*'Fenster, 'Schwester*	*Fenstern, Fenstere*
		-el ($n = 10$; 32)	*'Löffel, 'Gabel*	*Löffeln, Löffelen*
		-en ($n = 4$; 12)	*'Kissen, 'Garten*	*Kissene, Kissenen*
No final schwa-syllable ($n = 21$; 67)	**Monosyllabic** ($n = 13$; 42)	*NA*	*'Fisch, 'Frau*	*Fisch, Fischen*
	Multisyllabic, no final schwa-syllable ($n = 8$; 25)	Final stress ($n = 3$; 11)	*Klau'sur, Sa'lat*	*Klausur, Klausure*
		Nouns with stress shift ($n = 3$; 9)	*Pro'fessor, 'Doktor*	*Professore, Professorn*
		Final full vowel ($n = 2$; 5)	*'Kino, 'Auto*	*Kino, Kinoer*

TABLE 5 Model summary for Percent Trochaic plural forms in the plural elicitation task with non-words in L2 and L1 speakers (Task 1).

	Estimate	Std. error	z-value	*p*-value
(Intercept)	7.87	1.10	7.15	<0.001
Level1: Y1 vs. Y2	0.30	0.43	0.69	0.488
Level2: Y1+2 vs. Y3	1.59	0.43	3.74	<0.001
Level3: Y1-3 vs. Y3+	0.50	0.29	1.70	0.089
Level4: L2 vs. L1	0.00	0.36	0.01	0.990
Class: *-el* vs. *-e*	−3.41	1.14	−3.00	0.003
Class: *-en* vs. *-e*	−3.80	1.14	−3.33	0.001
Class: *-er* vs. *-e*	−3.73	1.13	−3.30	0.001
Class: FinStress vs. *-e*	−4.69	1.11	−4.22	<0.001
Class: Mono vs. *-e*	−4.63	1.11	−4.17	<0.001
Class: -V vs. *-e*	−2.82	1.15	−2.47	0.014
Ending: *-e* vs. *-n*	0.13	0.20	0.67	0.504
Ending: *-en* vs. *-n*	−0.21	0.18	−1.12	0.261
Ending: *-er* vs. *-n*	−0.24	0.31	−0.76	0.445
Ending: -Ø vs. *-n*	−1.45	0.19	−7.70	<0.001
Ending: other vs. *-n*	−2.26	0.44	−5.15	<0.001
Ending: *-s* vs. *-n*	−2.55	0.24	−10.51	<0.001

real words (Task 2) completed only by the L2 users, results focus on (i) accuracy scores—a binary categorical variable (correct vs. incorrect), and (ii) just as in Task 1, whether the plural forms produced show a word-final trochaic form—a binary categorical variable (trochaic vs. non-trochaic). For Task 3, the grammatical acceptability judgment task, the data obtained by the experiment yielded ordinal ratings on an 8-point Likert scale. For the results presented here, we set aside questions of stem-vowel changes (umlaut) (but see Schuhmann and Smith, accepted).

3.4.1 Coding

3.4.1.1 Trochaic vs. non-trochaic

For both the non-word and real word plural elicitation tasks, we analyzed how often participants produced trochaic plural forms for the singular forms provided. A form was coded as "trochaic" when the plural form ended in a sequence of a stressed followed by an unstressed syllable. All other forms were coded as "non-trochaic", i.e., most notably, monosyllabic forms, forms with a final stressed syllable, or a sequence of two final reduced syllables. All items were included in the analyses.

3.4.1.2 Ending chosen

When participants produced a form in either of the plural elicitation tasks that did not differ from the singular form provided, this null-marking was coded as "NO". Thus, we conceptualized this as a zero or null-morpheme (a morpheme with no overt phonetic content). When an ending did not correspond to any of the typical plural markers for SG, i.e., *-e*, *-en*, *-n*, *-er*, *-s*, or -Ø, it was coded as "other" (e.g., <-in> or deleting final consonants, e.g., turning *Rücken* "back" into *Rücke* "backs"). Note that these are the suffixes used by the speakers in the study, so the occurrences of individual suffixes are not evenly distributed; in particular, both *-er* and "other" occurred rarely as suffix-markers. Umlaut is not considered in this analysis of suffixes chosen.

3.4.1.3 Accuracy

For the real word task, we analyzed accuracy, i.e., whether the plural forms produced by the L2 users corresponded to "correct", i.e., target-like plural forms in terms of the overall *segmental* composition of the plurals. A form was coded as correct if it corresponded to SG or frequently used forms in colloquial German, such as *Mädel* "girls" or *Mädel-s* "girls, colloq.", respectively, which are also listed in Duden-Deutsches Universalwörterbuch (2023).

3.4.2 Statistical analyses of trochaic pattern and accuracy in Tasks 1 and 2

For the binary categorical variables (trochaic vs. non-trochaic, correct vs. incorrect), the data were analyzed with a binomial generalized linear mixed-effects model in R (version 4.2.2; R Core Team, 2021) using the lme4 package (version 1.1-31) (Bates et al., 2015). We included either "Percent Trochaic" (trochaic vs. non-trochaic) or "Percent Accurate" (accurate vs. inaccurate) as the binary dependent variable. In a generalized linear mixed-effects model, the binary dependent variable is transformed into a continuous variable via a log odds link function, logit(p), i.e., the natural log of the probability. Probability here refers to the probability of providing a trochaic answer.

We included proficiency level ("Level") as a fixed effect in the model. Level had four levels of increasing German language proficiency for the real word task with just L2 users (Year 1, Year 2, Year 3, Year 3 + Imm.), and five levels of increasing German language proficiency (Year 1, Year 2, Year 3, Year 3 + Imm., L1) for the non-word task with L1 and L2 participants. This factor was reverse-Helmert-coded, where each level of a factor is compared to the mean of the previous levels. Thus, in the model outputs, Level1 represents the contrast between the mean of Year 1 and Year 2, Level2 the contrast between the mean of Years 1 and 2 vs. the mean of Year 3, Level3 the contrast between the mean of Years 1, 2, and 3 versus the mean of Year 3 + Imm., and—for the non-word task additionally—Level4 the contrast of the mean of all L2 levels (Year 1, Year 2, Year 3, Year 3 + Imm.) vs. the mean of L1.

We fitted models with by-subject and by-word random intercepts and a random slope for Level by word. Since a stimulus' membership in a word class ("Class") and—at a reviewer's suggestion—the chosen ending ("EndingChosen") could have an effect on the prosodic shape, we tested whether Class and EndingChosen should be included in the model as covariates for Percent Trochaic by means of model comparisons (cf. Baayen et al., 2008).[11] Class refers to the stimulus category of a singular word (labeled "categories" and "sub-categories" in Tables 2–4) based on features such as its prosody, stress pattern, and word-final phonology. The seven levels of Class were: (singular words ending in) *-e, -el, -en, -er*, FinStress (words ending in a stressed syllable), Mono (monosyllabic words), and -V (words ending in an unstressed full vowel). Task 2 with real words additionally included the Class category of StrChg (words with a stress change between singular and plural forms). EndingChosen refers to the specific plural suffixes that participants provided when pluralizing each of the singular stimulus forms. The seven levels of EndingChosen were: the suffixes *-e, -n, -en, -er*, NO (null morpheme/no phonetically overt suffix), *-s*, and other. Class and EndingChosen were each treatment-coded, which compares each level of Class or EndingChosen to a reference level. For Class and EndingChosen, we selected the most predictable forms as the reference level. Thus, for Class, we chose singular forms ending in *-e*, for which the plural form is unambiguously the *-n* suffix (see also Kauschke et al., 2013). Conversely, for EndingChosen, the reference level was the *-n* suffix, as this is the predictable plural suffix for a subset of stimuli, namely singulars ending in *-e* (note, however, that the *-n* suffix also occurs on feminine singulars ending in a closed schwa-syllable).[12] *Post-hoc* pairwise contrasts for Level were calculated with the *emmeans*-function (Lenth, 2021). The statistical analysis for the Likert scale-like data in the grammatical acceptability judgment task (GAJT) in Task 3 is discussed in the relevant section (Section 4.3) below.

4 Results

4.1 Results Task 1: non-word plural elicitation task

4.1.1 Percent Trochaic

We first analyzed how often participants produced trochaic plural forms in the non-word task. Figure 1 presents the mean Percent Trochaic values and 95% confidence intervals (CI) for participants' plural forms at each (proficiency) Level (L2: Year 1, Year 2, Year 3, Year 3 + Imm., and L1). With 0.92 (SD 0.27), the vast majority of plural forms produced by L1 German users were trochaic. The L2 users produced an increasing number of trochaic plural forms for non-words with increasing L2 experience. Their mean percentage scores ranged from 0.78 (SD 0.41), to 0.87 (SD 0.34), 0.94 (SD 0.24), and 0.95 (SD 0.23) trochaic plural forms in the first year, second year, third year, and third year plus immersion, respectively. Of note is also that the means for L2 users in the two Year 3 groups (Year 3 and Year 3 + Imm.) showed a slightly higher number of trochaic plural forms than the L1 German users in this task.

Model comparisons within a binomial generalized linear mixed-effects model analysis and "Percent Trochaic" (trochaic vs. non-trochaic) as the binary dependent variable revealed that model fit improved when including (word) Class as a covariate

11 A model with an interaction of Level, Class, and EndingChosen would be too complex, due to the large number of factor levels for Level ($n = 5$), Class ($n = 7$; for real words: $n = 8$), and EndingChosen ($n = 7$), as well as imbalances in the data at hand.

12 Thus, the Class and EndingChosen comparisons were as follows. For Class: "Class: *-el* vs. *-e*", "Class: *-en* vs. *-e*", "Class: *-er* vs. *-e*", "Class: FinStress vs. *-e*", "Class: Mono vs. *-e*", "Class: -V vs. *-e*", and "Class: StrChg vs. *-e*" (only included in Task 2 with real words). For EndingChosen: "Ending: *-e* vs. *-n*", "Ending: *-en* vs. *-n*", "Ending: *-er* vs. *-n*", "Ending: NO vs. *-n*", "Ending: *-s* vs. *-n*", and "Ending: other vs. *-n*".

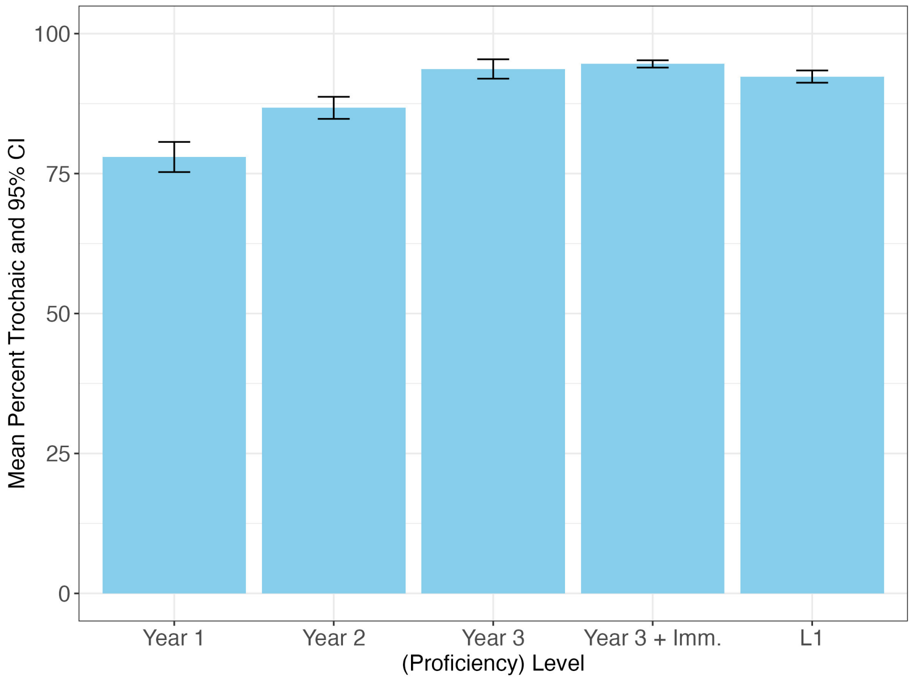


FIGURE 1
Percent Trochaic plural forms in the plural elicitation task with non-words (Task 1) in L2 and L1 speakers, split by (proficiency) Level.

($\chi^2_{(6)} = 60.5$, $p < 0.001$), and further improved when including EndingChosen ($\chi^2_{(6)} = 233.7$, $p < 0.001$). The model output (see Table 5) indicates that Level as well as Class and EndingChosen are significant predictors, as one Level contrast, as well as all of the Class contrasts and some EndingChosen contrasts were significant. *Post-hoc* pairwise comparisons for Level showed that the percentage of trochaic plural forms was significantly higher for more proficient L2-levels for the comparisons between Year 1 and Year 3 ($p < 0.005$) and Year 2 and Year 3 ($p = 0.012$), and marginal for the comparison between Year 1 and Year 3 + Imm. ($p = 0.053$), as well as Year 2 and Year 3 + Imm. ($p = 0.068$). None of the contrasts between individual L2 levels and the L1 level were significant.

These results suggest that overall, the percentage of trochaic forms in the production of plural forms for German non-words increased with increasing proficiency among L2 users, with significant differences between Year 1 and Year 3 (without immersion), and between Year 2 and Year 3 (without immersion). However, L2 users produced a similar number of trochaic plural forms as L1 users.

4.1.2 Percent Trochaic by (word) Class

The model summary from Table 5 above also indicates that (word) Class had an influence on Percent Trochaic. Due to the larger number of (word) Classes ($n = 7$) and (proficiency) Levels ($n = 5$), a Class-by-Level interaction was not included in the model.[13] Instead, we present visualizations for each (proficiency) Level, which illustrate whether the percentage of trochaic plurals varied with (word) Class. The results are shown in Figure 2.

Overall, Figure 2 illustrates the effect of Class on the percentage of trochaic plural forms in non-words across (proficiency) Levels. The reference level of Class, singular non-words ending in *-e* (schwa), is listed first in each panel. For all groups, this level showed the highest percentage of trochaic plural forms, with close to 100% trochaic forms in each case. This can be accounted for by the fact that words in *-e* unambiguously form the plural with *-n* in German (cf. Kauschke et al., 2013). All comparisons between non-words in *-e* and the other individual Class levels are significant in the model summary (in Table 5) above, which collapses across all (proficiency) Levels. After words ending in schwa, the classes of words with final stress ("FinStress"), monosyllabic singulars ("Mono"), and words ending in an unstressed full vowel in the final syllable ("V") showed the next highest levels of trochaic plural forms at Years 1 and 2 (and to some degree at Year 3). L1 speakers also showed slightly lower percentages of trochaic forms for these three classes compared to words ending in schwa.[14]

Crucially, the graph further indicates that the lower percentage of trochaic plural forms in less proficient L2 German users was driven by the lower number of trochees in one specific group of non-words, namely non-words ending in closed schwa-syllables (*-el*, *-en*, *-er*). In other words, L2 users in Year 1, who showed the lowest percentage of trochaic plural responses in the analyses

13 This also means that *post-hoc* contrasts between Classes at individual Levels could not be conducted.

14 In fact, for the L1 speakers, the (word) Class of "Mono" (monosyllabic singulars) numerically showed the lowest amount of trochaic forms (0.87), followed by singular forms with final stress ("FinStress": 0.91), and then singular forms ending in a syllable with an unstressed full vowel ("V": 0.95).

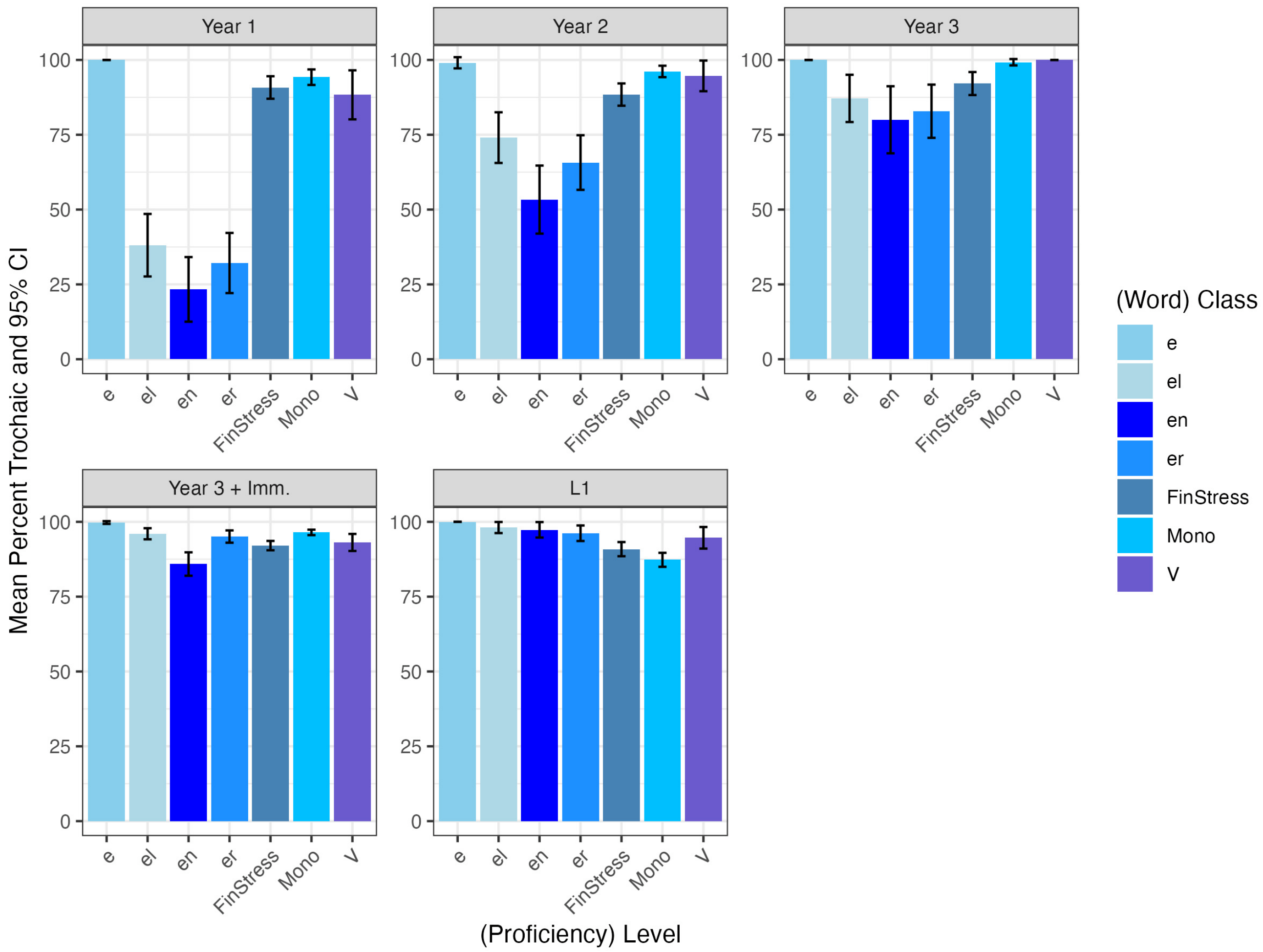


FIGURE 2
Percent Trochaic plural forms in the plural elicitation task with non-words (Task 1) in L2 and L1 speakers, split by (proficiency) Level and (word) Class.

above (see also Figure 1), produced overwhelmingly trochaic plural forms in all classes except for the three classes of singular forms ending in pseudo-suffixes, i.e., in closed schwa-syllables (*-el*, *-en*, *-er*). The panels across the four L2 (proficiency) Levels (the first four panels in Figure 2) further indicate that the gap between the lower percentage of trochaic plural forms in words ending in closed schwa-syllables and words in an open schwa-syllable gradually closed over the course of the L2 developmental sequence. In the group of L1 users (the last panel), the three classes ending in pseudo-suffixes were all produced highly trochaically (0.96–0.98).

4.1.3 Percent Trochaic by EndingChosen

The model summary in Table 5 above also indicates that EndingChosen had an influence on Percent Trochaic. Thus, we asked whether and which specific suffixes were used more often with trochaic plural forms than others. The results are shown in Figure 3.

Figure 3 shows that at Year 1, L2 speakers produced the greatest number of trochaic plural forms using *-n*, followed by trochaic forms with the suffixes *-e*, *-s*, and *-er*. Fewer trochaic forms were produced with the *-en* suffix, the null-suffix ("NO"), as well as "other" (non-standard) suffix choices. By Year 2, both *-en* and the null-suffix options were used to produce a similar number of trochaic plural forms as with the other existing suffixes (*-n, -e, -s, -er*). At Year 3 with and without immersion, L2 speakers showed a slightly higher percentage of trochaic forms for the suffixes.[15]

The L1 speakers showed trochaic plural forms with little variability when using the suffixes *-e*, *-en*, and *-n*, but fewer trochaic forms with the null-suffix, and even fewer trochaic plurals with the *-s* suffix (albeit with more variability). The suffix *-er* occurred very rarely in L1 users ($n = 447$, only 3.74% of the data), which is in line with the fact that this suffix has been frequently discussed in the literature as a non-productive plural allomorph. Both the *-er* and "other" endings ($n = 138$, only 1.15% of the data) were used with relatively trochaic plural forms but show large variability due to their very small number of occurrences.

4.1.4 EndingChosen based on (word) Class—across (proficiency) Levels

Figure 4 shows which suffixes participants chose for each (word) Class, broken down by (proficiency) Level. This graph

15 Except that the *-s* suffix (and possibly the null-suffix) occurred with slightly fewer trochaic plurals in Year 3 + Imm.

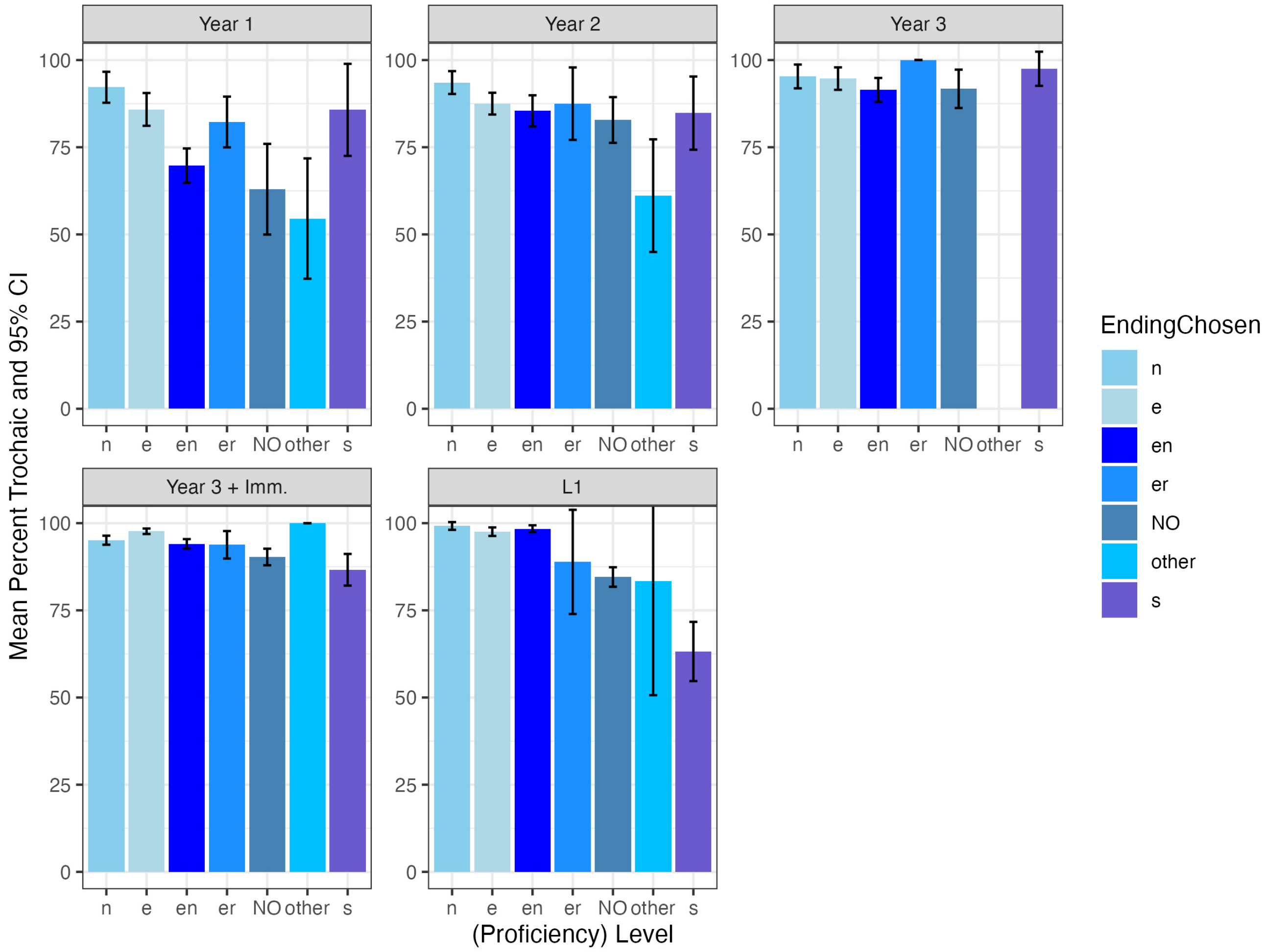


FIGURE 3
Percent Trochaic plural forms in the plural elicitation task with non-words (Task 1) in L2 and L1 speakers, split by (proficiency) Level and EndingChosen (i.e., suffixes produced).

serves to illustrate the combined information from the previous two figures. First, most of the differences between (proficiency) Levels can be seen for words ending in one of the three closed schwa-syllables. In these three (word) Classes, the choice of endings became more target-like (i.e., more similar to L1 speaker responses) with increasing proficiency. Combined with the discussion from the two previous sections, it is apparent that at Year 1, L2 learners used a fair number of *-en* suffixes—and some *-e* suffixes—for non-words ending in closed schwa-syllables, thus leading to trisyllabic forms, i.e., non-trochaic forms. This explains the substantial decrease in trochaic plural forms in these three (word) Classes at Year 1 in Figure 2, and the decrease of trochaic forms in comparison to other groups when the suffix *-en* was used at Year 1 in Figure 3.

Non-words ending in *-el* and *-er* also received more *-n* suffixes and fewer null-suffixes[16] at all L2 (proficiency) Levels compared to L1 speakers. At Year 2, the L2 users showed an increase in null-suffixes for all three closed schwa-syllables. At the same time, L2 users reduced their use of *-en* suffixes while increasing their use of *-n* suffixes for *-el* and *-er* classes from Year 1 to Year 2, resulting in more trochaic plural forms.[17]

4.2 Results Task 2: real word plural elicitation task

4.2.1 Accuracy

For the real word task, which was completed by L2 speakers only, we first present the accuracy results. In Figure 5, the darker, left-hand bar within each (proficiency) Level represents the mean accuracy and 95% confidence interval for participants producing correct plural forms. As can be seen in this figure, L2 learners produced more correct plural forms with increasing L2 experience. L2 participants produced 0.41 (SD 0.49), 0.52 (SD 0.50), 0.66 (SD 0.47), and 0.66 (SD 0.47) correct plural forms in the first year

16 Non-words ending in *-en* received a fair number of *-en* suffixes, rather than *-n* suffixes, as the latter cannot be added to *-en* for phonotactic reasons.

17 There was also a small increase in *-e* suffixes for these three (word) Classes, in particular *-en* words, as well as singular words with final stress ("FinStress") and monosyllables ("Mono") by Year 2.

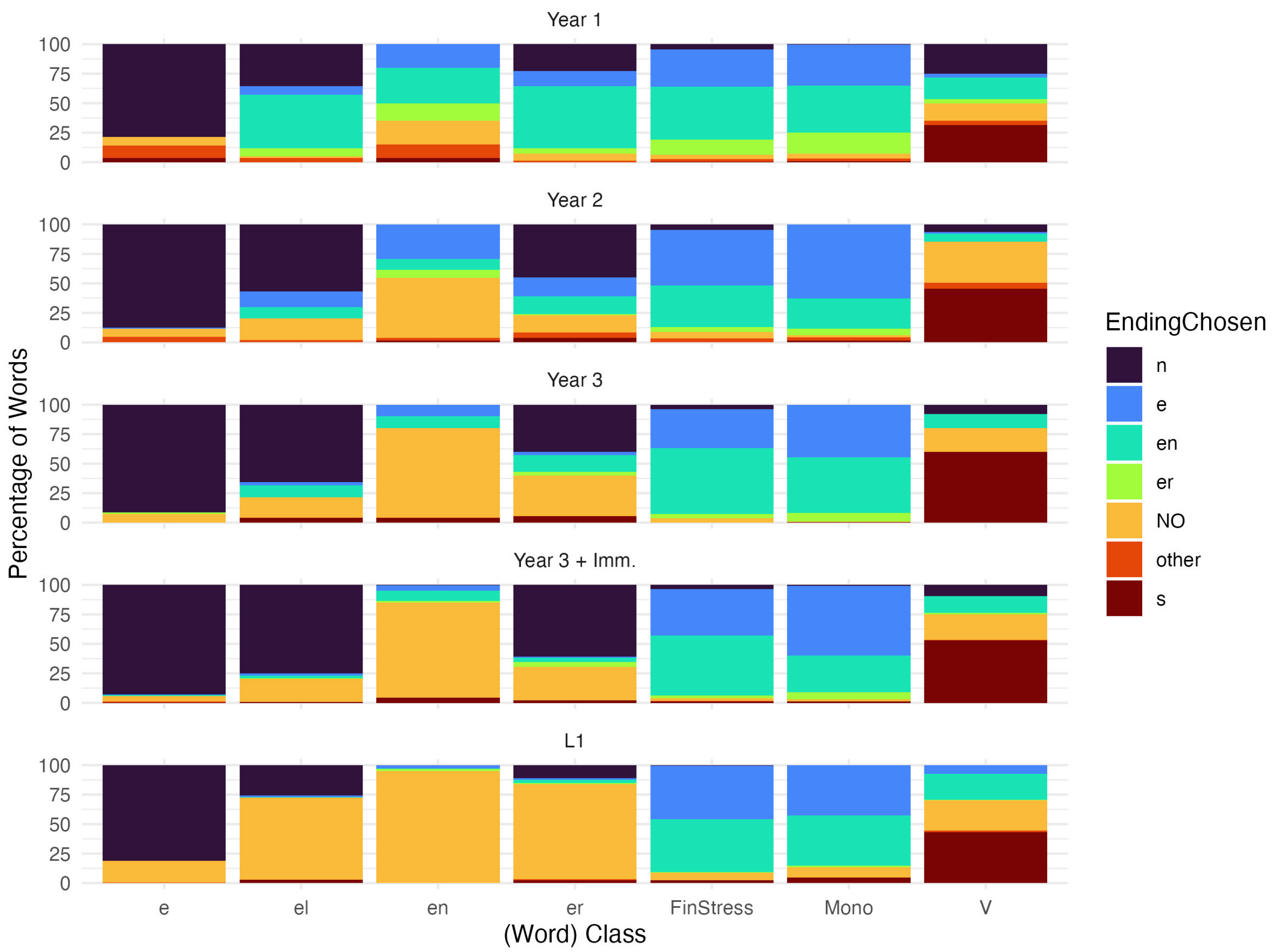


FIGURE 4
Percentage of words produced with a particular Ending (EndingChosen) in the plural elicitation task with non-words (Task 1) in L2 and L1 speakers, split by (proficiency) Level and (word) Class.

(Year 1), second year (Year 2), third year (Year 3), and third year plus immersion (Year 3 + Imm.), respectively.

Model comparisons within a binomial generalized linear mixed-effects model analysis and "Percent Accurate" (accurate vs. inaccurate) as the binary dependent variable revealed that model fit improved when including (word) Class as a covariate ($\chi^2_{[7]} = 88.6, p < 0.001$), and further when including EndingChosen as a covariate ($\chi^2_{[6]} = 386.1, p < 0.001$). The model output in Table 6 indicates that Level, Class, and EndingChosen were each significant predictors, as all Level contrasts were significant, as well as all Class and most EndingChosen contrasts. *Post-hoc* pairwise comparisons for Level showed that accuracy was always significantly higher for higher Levels ($p < 0.05$), with one exception: The contrast for Year 3 with and without immersion was not significant.

Overall, these results suggest that accuracy in the plural production of existing German words increased with course level, although the two Levels at the high end of the proficiency spectrum (Year 3 vs. Year 3 + Imm.) did not show significant differences between each other. At the same time, accuracy was far from ceiling for even the most advanced groups in our study, highlighting the difficulty in producing accurate plural forms for L2 speakers. (Figures for accuracy split by Class and EndingChosen at the different (proficiency) Levels are available as Supplementary Figures 1, 2.)

4.2.2 Percent Trochaic

We next analyzed how often participants produced plural forms that were trochaic. In Figure 5 below, the lighter, right-hand bars present the mean Percent Trochaic values and 95% confidence intervals for participants' plural forms at each (proficiency) Level. As can be seen in this figure, the L2 learners produced a larger number of trochaic plural forms with increasing L2 experience. Their mean percentage scores ranged from 0.73 (SD 0.44), to 0.82 (SD 0.38), 0.90 (SD 0.30), and 0.92 (SD 0.27) trochaic plural forms in the first year, second year, third year, and third year plus immersion, respectively.

Model comparisons within a binomial generalized linear mixed-effects model analysis and "Percent Trochaic" (trochaic vs. non-trochaic) as the binary dependent variable revealed that model fit improved when including (word) Class as a covariate, ($\chi^2_{[7]} = 95.6, p < 0.001$), and further when including EndingChosen

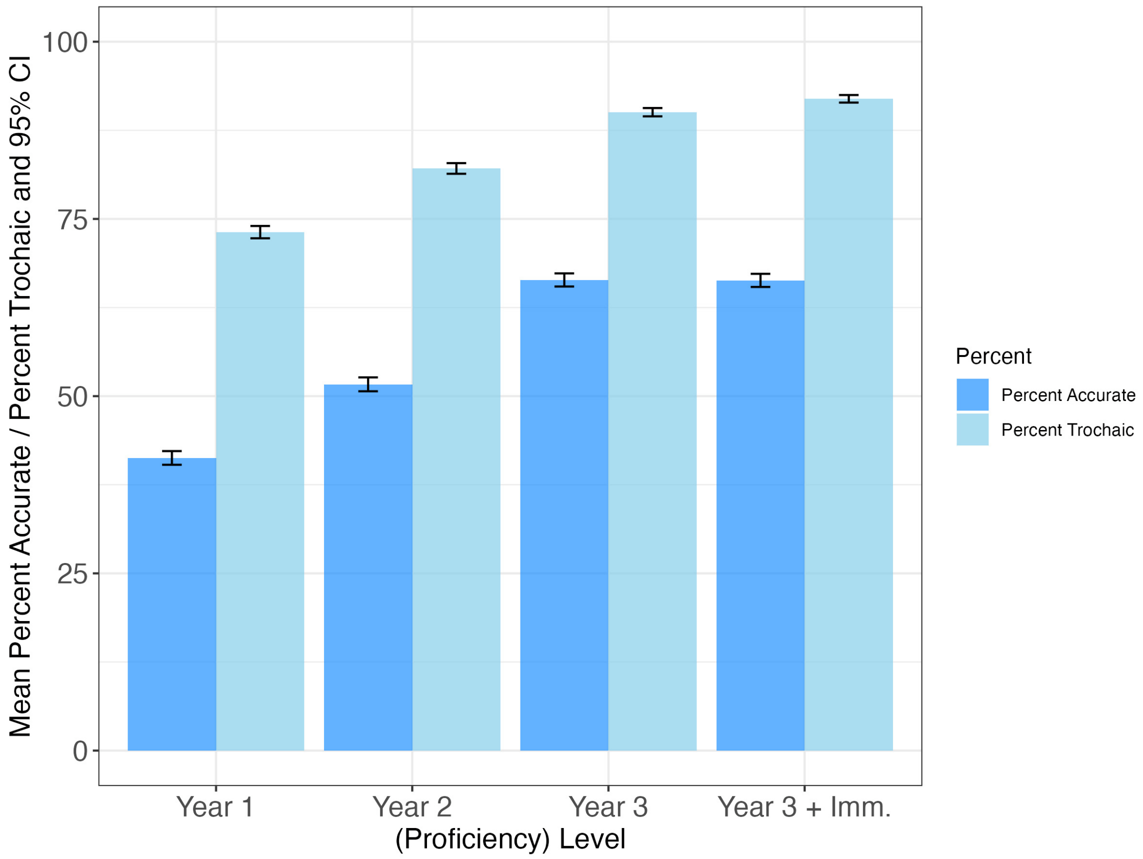


FIGURE 5
Percent Accurate (darker, left-hand bars) and Percent Trochaic (lighter, right-hand bars) plural forms in the plural elicitation task with real words (Task 2) in L2 speakers, split by (proficiency) Level.

($\chi^2_{[6]} = 215.3$, $p < 0.001$). The model output (see Table 7) indicates that Level, Class, and EndingChosen were significant predictors, as all three Level contrasts were significant, as well as many of the Class and all EndingChosen contrasts. *Post-hoc* pairwise comparisons for Level showed that the percentage of trochaic plural forms was always significantly higher for higher Levels ($p < 0.001$), with two exceptions: The contrast between Year 1 and Year 2 was only marginally significant ($p = 0.057$), and the contrast for Year 3 with and without immersion was not significant.

Overall, these results suggest that the percentage of trochaic forms in the production of plural forms for existing German words increased with course level, although the low and high end of the proficiency spectrum (Years 1–2 and Year 3–Year 3 + Imm.) did not show significant differences from each other.

4.2.3 EndingChosen by (word) Class—split by (proficiency) Level

Figure 6 illustrates the effect of EndingChosen by (word) Class at the different (proficiency) Levels. This figure illustrates the combined information from Supplementary Figures 3 and 4, which show Percent Trochaic split by Class and EndingChosen, respectively, at the different (proficiency) Levels. Paralleling the non-word task, most of the development across (proficiency) Levels took place in the Classes of words ending in closed schwa-syllables. First, all of these three Classes showed a substantial number of *-en* suffixes at Year 1, which were reduced in number by Year 2 and further reduced in number by Year 3 (without and with immersion). A small number of *-e* suffixes also appeared at Years 1 and 2. In tandem, the number of null-suffixes increased over the course of the four (proficiency) Levels, in particular for words ending in *-en* (and similarly for *-er*; yet Year 3 + Imm. had fewer null-suffixes for words ending in *-er* than Year 3). When comparing the patterns with those in the non-word task, L2 speakers at Year 1 started out using more null-suffixes for real words ending in *-el, -en, -er*, especially for words ending in *-en*.

4.2.4 Percent Trochaic of incorrect plurals

Further analyses of just the incorrect plural forms revealed that the overall grand mean percent of trochaic items decreased from 0.88 (accurate and inaccurate plurals) to 0.74 (only inaccurate plurals). When examining the percentage of trochaic forms among incorrect plurals across the (proficiency) Levels, an increase in prosodically well-formed—albeit incorrect—plural forms can be seen. In other words, when analyzing only the incorrect plural forms, the percentage of trochaic plural forms increased with increasing proficiency. Specifically, the L2 users across the four (proficiency) Levels produced incorrect but trochaic plural forms

TABLE 6 Model summary for Percent Accurate plural forms in the plural elicitation task with real words in L2 speakers (Task 2).

	Estimate	Std. error	z-value	*p*-value
(Intercept)	2.62	0.29	9.00	<0.001
Level1: Y1 vs. Y2	0.59	0.23	2.62	0.009
Level2: Y1+2 vs. Y3	1.22	0.22	5.58	<0.001
Level3: Y1-3 vs. Y3+	0.90	0.14	6.59	<0.001
Class: *-el* vs. *-e*	−4.73	0.41	−11.42	<0.001
Class: *-en* vs. *-e*	−3.21	0.46	−6.99	<0.001
Class: *-er* vs. *-e*	−3.87	0.39	−9.83	<0.001
Class: FinStress vs. *-e*	−2.82	0.37	−7.69	<0.001
Class: Mono vs. *-e*	−3.27	0.39	−8.43	<0.001
Class: StrChg vs. *-e*	−2.69	0.56	−4.83	<0.001
Class: -V vs. *-e*	−2.05	0.40	−5.17	<0.001
Ending: *-e* vs. *-n*	1.15	0.12	9.54	<0.001
Ending: *-en* vs. *-n*	0.89	0.11	8.08	<0.001
Ending: *-er* vs. *-n*	2.31	0.21	10.84	<0.001
Ending: -Ø vs. *-n*	0.75	0.09	8.27	<0.001
Ending: other vs. *-n*	−30.01	223,518	0.00	1.00
Ending: *-s* vs. *-n*	0.40	0.12	3.34	0.001

TABLE 7 Model summary for Percent Trochaic plural forms in the plural elicitation task with real words in L2 speakers (Task 2).

	Estimate	Std. error	z-value	*p*-value
(Intercept)	7.07	0.66	10.74	<0.001
Level1: Y1 vs. Y2	0.77	0.30	2.52	0.012
Level2: Y1+2 vs. Y3	1.73	0.32	5.47	<0.001
Level3: Y1-3 vs. Y3+	1.51	0.20	7.46	<0.001
Class: *-el* vs. *-e*	−5.25	0.74	−7.12	<0.001
Class: *-en* vs. *-e*	−3.48	0.78	−4.44	<0.001
Class: *-er* vs. *-e*	−4.83	0.73	−6.64	<0.001
Class: FinStress vs. *-e*	−1.73	0.72	−2.41	0.016
Class: Mono vs. *-e*	−1.21	0.75	−1.60	0.109
Class: StrChg vs. *-e*	−1.38	0.90	−1.53	0.127
Class: -V vs. *-e*	−2.28	0.75	−3.04	0.002
Ending: *-e* vs. *-n*	−1.72	0.17	−9.99	<0.001
Ending: *-en* vs. *-n*	−1.73	0.15	−11.31	<0.001
Ending: *-er* vs. *-n*	−1.94	0.34	−5.66	<0.001
Ending: -Ø vs. *-n*	−1.44	0.15	−9.37	<0.001
Ending: other vs. *-n*	−2.25	0.29	−7.85	<0.001
Ending: *-s* vs. *-n*	−2.33	0.18	−13.31	<0.001

in 0.57 (SD 0.50), 0.66 (SD 0.47), 0.75 (SD 0.43), and 0.82 (SD 0.39) of cases, respectively (cf. Supplementary Figure 5).

Performing the same binomial generalized linear mixed-effects model analysis as above with "Percent Trochaic" (trochaic vs. non-trochaic) as the binary dependent variable, a model comparison revealed that model fit improved when including (word) Class as a covariate ($\chi^2_{(7)} = 72.2$, $p < 0.001$), and further when including EndingChosen ($\chi^2_{(6)} = 317.3$, $p < 0.001$). The model output (see Supplementary Table 1) indicates that (proficiency) Level, Class, and EndingChosen were significant

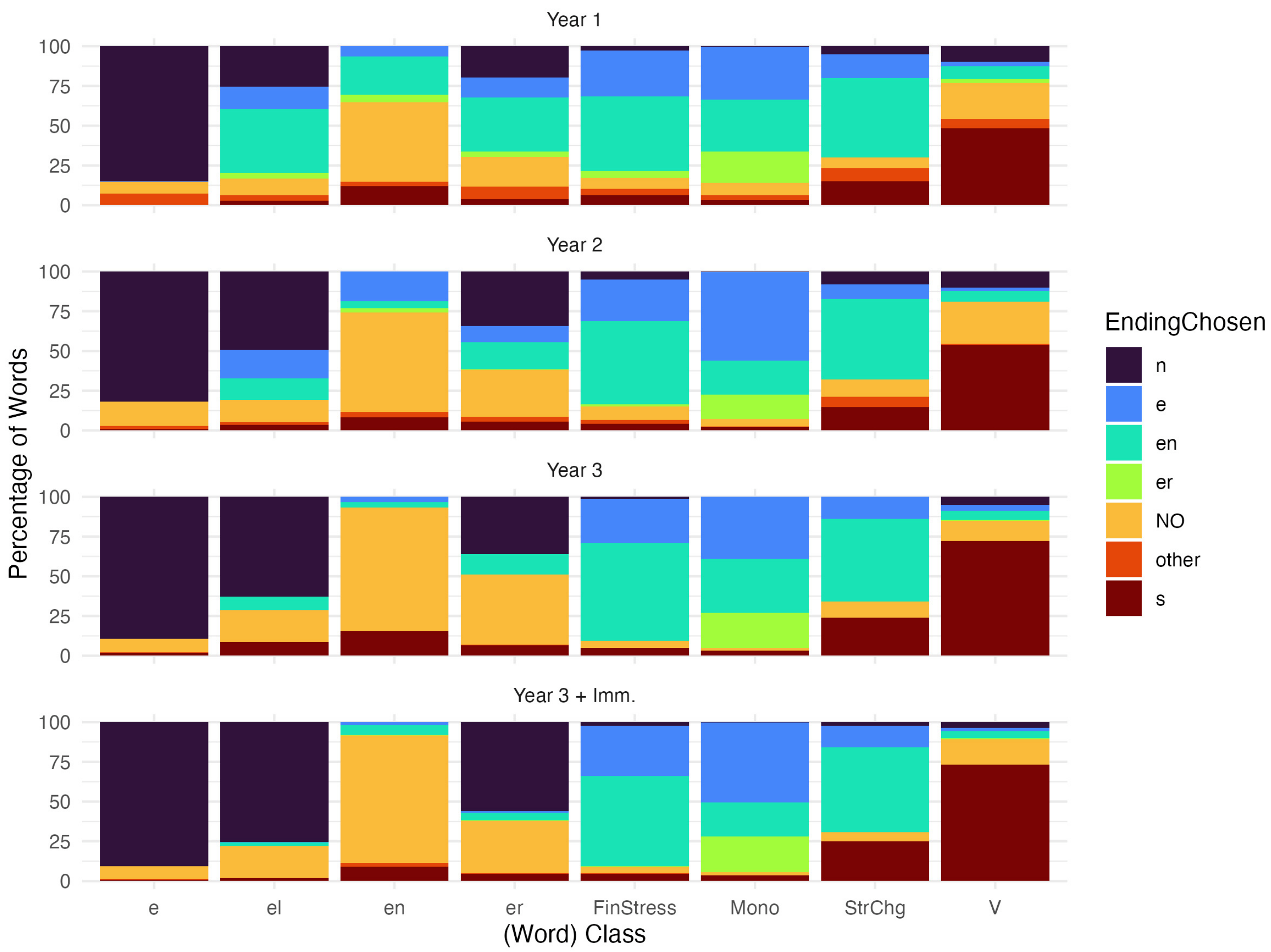


FIGURE 6
Percentage of words produced with a particular ending (EndingChosen) in the plural elicitation task with real words (Task 2) in L2 speakers, split by (proficiency) Level and (word) Class.

predictors, as several of each of the Level, Class, and EndingChosen contrasts were significant. *Post-hoc* pairwise comparisons for Level showed that the percentage of trochaic plural forms was always significantly higher for higher Levels ($p < 0.02$), with two exceptions: The contrast between Year 1 and Year 2, and the contrast for Year 3 with and without immersion were not significant.

Overall, the results of just the incorrect plural forms suggest that the percentage of trochaic forms in the production of plural forms for existing German words increased with course level, although the two Levels at the low and high end of the proficiency spectrum (Year 1–Year 2, and Year 3–Year 3 + Imm., respectively) did not show statistically significant differences.

4.3 Results Task 3: grammatical acceptability judgment task (GAJT)

4.3.1 Correlations of ratings and plural categories

This task was designed as a well-formedness rating task of existing and made-up plural forms of existing German words. First, we expected that incorrect forms would be rated as less acceptable than correct forms. The crucial question was whether the rating would reflect both the forms' accuracy and prosodic form such that within both the group of correct and the group of incorrect plurals, non-trochaic forms would be rated as less acceptable than trochaic forms. Thus, we hypothesized that sensitivities to both accuracy and the prosodic pattern in German plurals should be reflected in the ratings (as lower numbers on the scale indicate more acceptability) across the four tested accuracy-by-prosody Plural Types, in this order: Correct Trochaic < Correct Non-Trochaic < Incorrect Trochaic < Incorrect Non-Trochaic. Figure 7 presents the well-formedness rating results split by (proficiency) Level.

In the raw L1 and Year 3 (without immersion) data, it is apparent that incorrect plurals were less acceptable than correct plurals, and most importantly, that within incorrect plurals, the non-trochaic forms were less acceptable than the trochaic forms. It should be noted that the second group, the correct non-trochaic plurals, made up a very small category with just one type of plural, namely monosyllabic singular forms pluralized

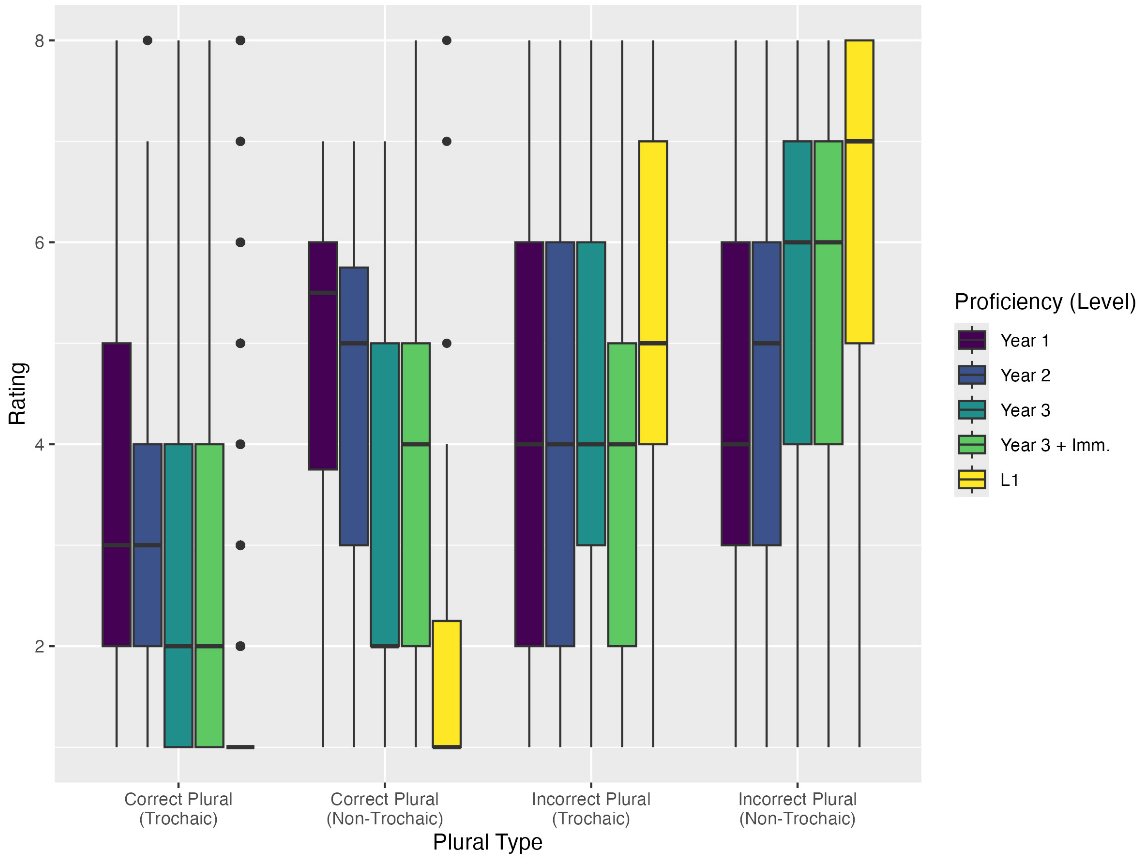


FIGURE 7
L2 and L1 users' well-formedness ratings (raw data) on the grammatical acceptability judgment task (GAJT, Task 3), split by Plural Type and (proficiency) Level.

with the *-s* suffix ($n = 6$).[18] Figure 7 presents, for each level of proficiency, a relatively steady increase in rating responses across all four Plural Types: Correct Plurals (Trochaic) < Correct Plurals (Non-Trochaic) < Incorrect Plurals (Trochaic) < Incorrect Plurals (Non-Trochaic), with the following exceptions for the second category of correct, non-trochaic plurals: The two mean values of Years 1 and 2 for the second category went against this overall trend, as they rejected non-trochaic but correct plurals strongly. Similarly, Year 3 + Imm. showed the same mean rating for the second (correct, non-trochaic plurals) and third category (incorrect, trochaic plurals). For the second Plural Type (correct, non-trochaic items), Year 1 and Year 2 users rejected these monosyllabic forms pluralized with the *-s* suffix on average more strongly than both types of incorrect plurals (Year 1), and more strongly than incorrect trochaic but approximately equally to incorrect non-trochaic plurals (Year 2).

We used the nonparametric Spearman's rank-order correlation test to analyze the ordinal Likert scale-like data. The four Plural Types were ranked in this order: Correct Trochaic—Correct Non-Trochaic—Incorrect Trochaic—Incorrect Non-Trochaic. We tested whether the participants' ratings correlated with the four ranked types of plurals in the study at each of the five (proficiency) Levels (Year 1 through Year 3 + Imm., L1). The individual participants' ratings were first turned into by-participant ranking of scores (ascending order).[19] This type of data analysis helped address differences in individual participants' use of the rating scale, as visual inspection suggested that L2 users tended to use more of the central parts of the scale. On the other hand, L1 users as a group made use of the entire scale for the categories examined.

The results showed a moderate positive correlation for Year 1 and Year 2. For Year 1: Spearman's $r_s(1996)$: 0.343, $p < 0.001$; for Year 2: $r_s(2500)$: 0.451, $p < 0.001$. The Year 3 groups without and with immersion and the L1 users showed a strong positive correlation. For Year 3: $r_s(1662)$: 0.606, $p < 0.001$; for Year 3 + Imm.: $r_s(9999)$: 0.604, $p < 0.001$; for L1: $r_s(5038)$: 0.647, $p < 0.001$.

Thus, the analysis showed decreasing acceptability ratings at each (proficiency) Level for these four ranked Plural Types: Correct

18 This class was included in the analysis at the request of a reviewer.

19 This means that we computed the rank of each rating value within each participant's ratings; the numerically lowest rating was assigned rank 1, the next lowest rating was assigned rank 2, etc.

Plurals (Trochaic) < Correct Plurals (Non-Trochaic) < Incorrect Plurals (Trochaic) < Incorrect Plurals (Non-Trochaic). However, the first two Levels of L2 users, Years 1–2, showed a moderate correlation, while the higher (proficiency) Levels of the Year 3 users (with and without immersion) and the L1 users showed strong correlations.

4.3.2 Ratings by (word) Class and Ending

As for the other tasks, we next present the data broken down by Ending (suffix provided) in Figure 8 (the data are broken down by (word) Class in Supplementary Figure 6), to visually examine whether the ratings varied by this factor.

All three figures for the GAJT (Figures 7, 8, Supplementary Figure 6) illustrate that the L2 German users in this study assigned lower acceptability ratings to incorrect plural forms, including non-trochaic plural forms, with increasing proficiency. At the same time, more proficient users also rated correct items as more acceptable. This involved, in particular, learning to accept correct non-trochaic forms (the second Plural Type), i.e., correct monosyllabic plurals in *-s*. This group of correct plural forms was strongly rejected in Years 1 and 2; even Year 3 + Imm. rejected this type of correct plurals more strongly than some incorrect trochaic plurals. Additionally, L2 acquisition involves learning to accept correct trochaic plural forms which are marked for plurality with a null-suffix. Non-overtly marked plurals had the lowest acceptance rate among all correct trochaic forms in Year 1. In fact, across all L2 (proficiency) Levels, trochaic plurals with the null-suffix consistently received the lowest or one of the lowest ratings, both within the group of correct and within the group of incorrect plurals.

Finally, among incorrect forms, Year 1 participants showed the highest acceptability ratings for both trochaic and non-trochaic plural forms with the *-en* suffix. Year 1 participants' ratings reflected a preference for incorrect trochaic plurals with the *-en* suffix over correct trochaic plurals with a null-suffix. At Year 1, trochaic forms with the suffixes *-en* and *-e* showed higher mean ratings than their non-trochaic counterparts; meanwhile, both trochaic and non-trochaic plural forms with the *-n* suffix and the null-suffix (as well as the *-s* suffix) showed comparable mean acceptability ratings. For L2 users beyond Year 1, each individual suffix received lower ratings in non-trochaic compared to trochaic plurals.

5 Discussion

5.1 Answering the research questions

5.1.1 RQ1: prosodic structure of non-words (Task 1)

The first research question examines the extent to which L1 and L2 German speakers produced trochaic plural forms in the non-word plural elicitation task, i.e., Task 1. As expected, the L1 German speakers produced a large majority of trochaic plurals (0.92, SD 0.27). Perhaps more surprising is the large majority of trochaic plurals produced by the L2 users, ranging from 0.78 (SD 0.41) by the Year 1 group to 0.95 (SD 0.23) by the Year 3 group with immersion. Overall, there was indeed a significant effect of (proficiency) Level, such that the more proficient the users, the more they produced trochaic plural forms. However, there was no significant difference between the Year 1 and Year 2 groups or between the two Year 3 groups (with and without immersion). Notably, both Year 3 groups actually produced percentages of trochaic plurals that were slightly higher—but not significantly different—than their L1 counterparts. This could point toward a slight tendency to overgeneralize the trochaic pattern in more advanced L2 users.

5.1.2 RQ2: accuracy and prosodic structure of real words (Task 2)

Recall that this set of research questions applies strictly to the L2 users of German and relates to their performance on a plural elicitation task using existing German words in terms of both accuracy (RQ2a) and creating trochaic plurals (RQ2b,c). In terms of accuracy, L2 users of German were not particularly successful at producing large numbers of accurate plurals overall, ranging from just 0.41 (SD 0.49) accuracy for participants in the Year 1 group to a high of 0.66 (SD 0.47) and 0.66 (SD 0.47) for users in the Year 3 groups, with and without immersion, respectively. Accuracy did improve significantly across proficiency Levels, although the two Year 3 groups did not differ significantly from each other.

In terms of the percentage of trochaic plurals produced during this task with real words overall (RQ1b) and on just incorrect words (RQ1c), a similar tendency was found. For all plurals produced as well as for just incorrect plural forms, the L2 users tended to produce more trochaic plural forms as proficiency increased (though without a significant difference between the Year 1 and Year 2 groups, and between the two Year 3 groups in each case). It is also worth noting that the mean percentage of trochaic forms was substantially higher than the means for each group's accuracy scores, although it has to remain acknowledged that the actual values ranged both above and below that mean (e.g., Year 1 means: trochaic 0.73 vs. accuracy 0.41—difference: 31.8 percentage points; Year 3 + Imm. means: trochaic 0.92 vs. accuracy 0.66—difference: 25.6 percentage points). This indicates—and was confirmed when examining just forms that were incorrectly pluralized—that even when L2 users formed incorrect plurals, those forms became more trochaic as proficiency increased.

5.1.3 RQ3: prosodic structure and well-formedness judgments (Task 3)

Results from the well-formedness judgment task (Task 3) revealed that there were significant correlations between the participants' ratings and the four types of plural forms which differed in accuracy (correct/incorrect) crossed with prosodic structure (trochaic/non-trochaic). This indicates that, overall, L1 and L2 German users rated non-trochaic forms as less well-formed compared to trochaic forms, both within the group of correct and within the group of incorrect plural forms provided. Furthermore, the strength of the correlations differed between (proficiency) Levels. L1 users and L2 users in Year 3 (with and without immersion) showed a strong correlation, while less proficient L2 users in Years 1 and 2 showed only a moderate correlation.

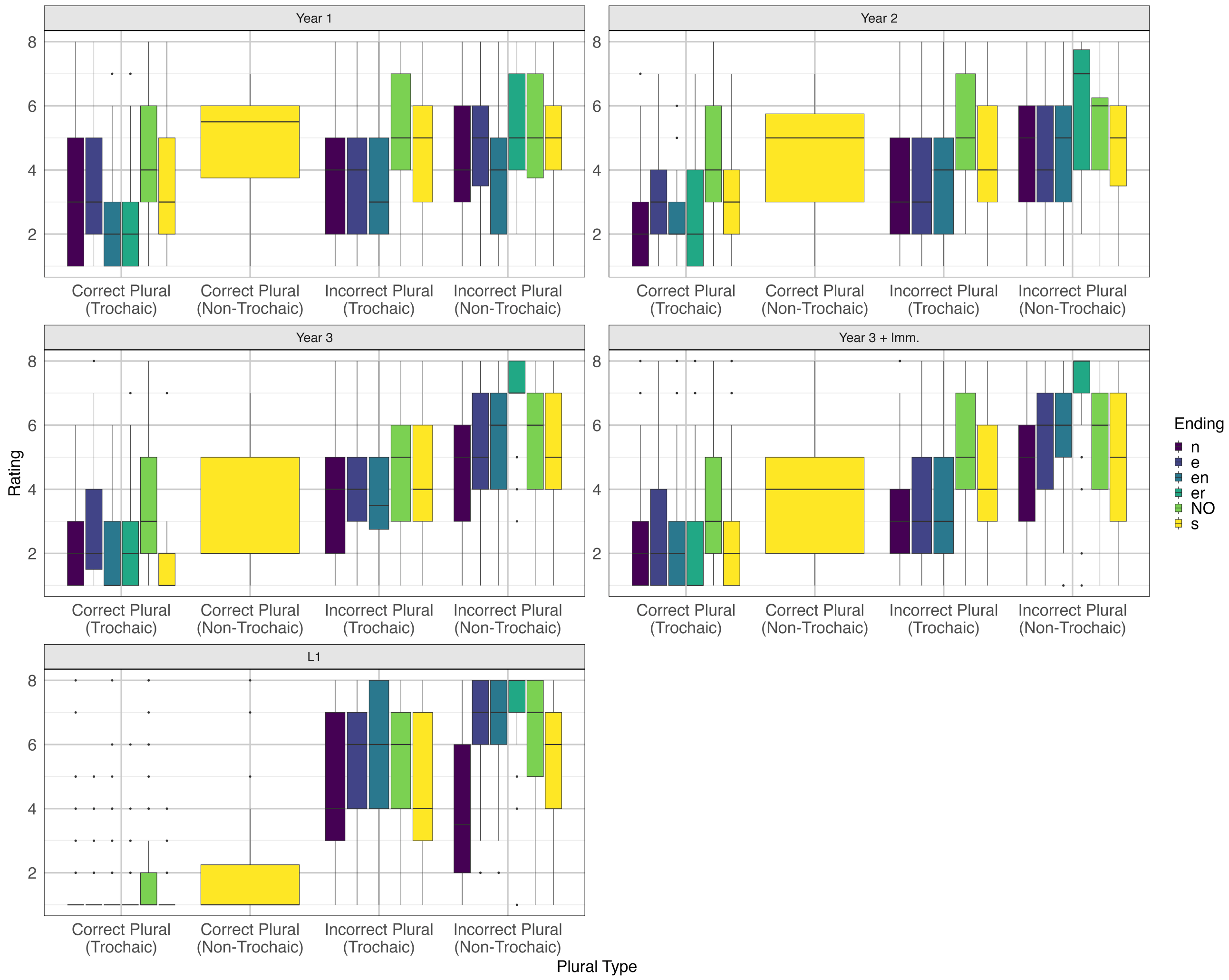


FIGURE 8
L2 and L1 users' well-formedness ratings on the grammatical acceptability judgment task (GAJT, Task 3), split by (proficiency) Level, Plural Type, and Ending (suffix provided).

5.1.4 Summary of the experimental data

One key take-away from the three tasks in this study is that L1 English-L2 German users produced more trochaic plurals and rated plural forms based on both accuracy and prosodic structure with increasing experience and proficiency. The data presented above also suggest that this group of L2 users started out with high levels of trochaic plural forms, around 0.73–0.78 for real and non-words, respectively. More advanced L2 users may have even tended to overgeneralize the trochaic pattern in the non-word data (although these differences were not statistically significant). The L2 speakers' performance in Task 2 with plural elicitations for existing German words indicated that L2 users might have been relatively more adept at the prosodic pattern compared to their overall accuracy for plural markings. (It should be acknowledged, however, that the potential for non-target-like performance on plural accuracy is naturally several times higher than for non-target-like performance on prosodic shape.)

5.2 Implications for formal models and accounts of German plural (acquisition)

5.2.1 The role of prosody in L1 and theoretical models

The results regarding the prosodic pattern from the three tasks outlined above are in line with previous experimental studies (e.g., Domahs et al., 2013, 2017; Kauschke et al., 2013) reviewed in the background section. As a reminder, Kauschke et al. (2013) noted that German-speaking children without language impairments produced more trochaic plural forms for real and non-words than their peers with impairments. Further clinical evidence in favor of the "optimal prosodic shape" in perception, i.e., in the interpretation of forms as singular or plural, was provided by a case study on a patient with aphasia—a person with impaired lexical knowledge (Domahs et al., 2017).

A number of theoretical analyses of the German plural, including those reviewed in Section 2, have drawn on the trochee as an important factor or constraint in shaping German plurals. The empirical results in this study are in line with theoretical accounts which analyze German plurals as containing a word-final disyllabic trochee. In particular, they are in line with generative accounts in which the trochee is modeled as a constraint (Wiese, 2009) or template (Smith, 2004, 2020; cf. Schuhmann and Putnam, 2021 for a phase-based account within Distributed Morphology) in the grammar for German plurals. From an OT perspective, in particular, it could be argued that during their language development, L2 learners of German come to rank a well-formedness constraint for trochees undominated in its area of application or weighted more heavily than other constraints related to the morphophonology of German plural formation (cf. Pater, 2009). Yet, detailed analyses of how the prosodic pattern interacts with morphology allow us to also evaluate other central models discussed in the background section.

5.2.2 The interplay of prosody and morphology

An exploratory look at the interplay of the prosodic patterns with the morphology, i.e., the stimuli's (word) Class and EndingChosen (in the two elicitation tasks) or Ending (suffix provided in the rating task) revealed several important aspects. First, the suffix *-s* was rarely used in the elicitation tasks (cf. also Schuhmann and Smith, accepted); when it was used, it mostly occurred on words ending in a syllable with an unstressed full vowel, both in L1 and L2 speakers across (proficiency) Levels. This finding in particular would not have been predicted by (the original version of) the dual route models (Clahsen, 1999; Marcus et al., 1995). As described in Section 2, in the dual route models, the *-s* suffix is considered by some scholars to be the default and would have been expected to occur very frequently with new or unknown words.

Secondly, the morphological analyses revealed which specific contexts correlated with lower percentages of trochaic forms in the elicitation tasks or lower ratings in the GAJT among the less proficient L2 users. Here, we found that L2 users started out with a very low percentage of trochaic plural forms for singulars ending in closed schwa-syllables. This resulted from learners in Year 1 frequently using syllabic suffixes instead of null-suffixes to mark plurality on these words (except for words ending in *-en*), most notably the suffix *-en*, less frequently *-e*, and occasionally *-er*. Null-marking gradually increased over the subsequent (proficiency) Levels but still did not reach L1 levels by Year 3. In place of the null-suffixes for words in closed schwa-syllables, L2 users produced more *-n* suffixes through Year 3 + Imm.; this led to forms which maintain a trochee but are segmentally not target-like. As a result, the percentage of trochaic plural forms increased across (proficiency) Levels for words in closed schwa-syllables, while segmentally, the plural forms still differed markedly from the L1 group by Year 3.

5.2.3 A proposal of relevant factors in the acquisitional path of German plurals

We suggest that the pattern discussed above for words ending in closed schwa-syllables among early L2 German learners can be accounted for by a few driving forces for marking the function of plurality on nouns in early L2 acquisition. These forces have already been proposed as part of Köpcke's schema model discussed above (e.g., Köpcke, 1988, *inter alia*). As a brief reminder, Köpcke's schema model proposes that certain plural forms signify plurality more strongly than other plural forms, depending on the strength of the cues, the latter of which include saliency, frequency, validity, and iconicity.

First, the patterns in the data overall can be captured with the notion of *iconicity*, i.e., a principle which marks plurality explicitly and overtly, thus rendering plurals different from singulars (cf. Eisenberg, 2020; Köpcke, 1988 for slightly different definitions). Secondly, Year 1 speakers frequently add syllabic suffixes such as *-en*, which could additionally be captured by the related strategy of *saliency* in Köpcke's schema model (e.g., Köpcke, 1988, *inter alia*). This expresses the notion that syllabic suffixes serve as a better cue to plurality than non-syllabic suffixes or umlaut. In the GAJT, the L2 users in Year 1 rated plurals formed with the suffix *-en* better than those formed with a null-suffix in all relevant Plural Types. This further corroborates the findings from the elicitation tasks where L2 learners initially overgeneralized the *-en* suffix to contexts where a null-suffix was expected.

Thirdly, in the GAJT (Task 3), learners in their first years of study strongly rejected existing monosyllabic plurals formed with the suffix *-s*. Together with the trisyllabic plurals for words ending in closed schwa-syllables just discussed, the L2 learners seemed to start out by requiring *multisyllabic* plural forms—in addition to iconicity and saliency. All three of these principles have been proposed in Köpcke's schema model (e.g., Köpcke, 1988, *inter alia*). These notions overlap partially, meaning that it is not always possible to identify which principle is the driving force in the observable behavior. Note, however, that this is very much in the spirit of Köpcke's schema model, in which prototypical plural forms unite a maximal number of cues to plurality (such as iconicity, saliency, multisyllabicity or—in later versions—trochee) on themselves.

It is also worthwhile to consider possible transfer effects from the learners' L1. As it stands, the contribution of *iconicity* might be reinforced by the learners' L1 patterns, as English plurals are—with a few exceptions—typically distinct from their singular forms. On the other hand, *multisyllabicity* is not a prevalent pattern in L1 English plural formation and may thus be less likely to be reinforced by L1 patterns, and therefore appears to be a strategy that the learners developed specifically for German plurals. Overall, the data from the early L2 learners can be meaningfully analyzed using the strategies proposed in the schema model by Köpcke and colleagues.

5.2.4 Comparison with other data on L2 German plural acquisition

The fact that German learners in Year 1 produced a substantial percentage of non-trochaic forms—albeit only for the closed schwa-syllables—is intriguing since other work on L2 German acquisition reported only a handful of cases of prosodically ill-formed plurals. Previous studies on L2 acquisition that have examined prosody were either concerned with children learning German as a second language in schools in Germany (Köpcke and Wecker, 2017; Wecker, 2016) or with adult foreign language learners of German at an Italian university (pilot study by Vogt, 2016). Children with SLI, however, were reported to form prosodically non-optimal plurals (Kauschke et al., 2011, but see Kauschke et al., 2013). While our study includes data from adult foreign language learners of German in their very early L2 development without much target-language input, both the child and adult L2 learners in the literature had received many years of input or exposure to German. This suggests that with increased target-language input, non-clinical learners will come to acquire the specific prosodic pattern of German plurals, i.e., trochaic rather than multisyllabic plural forms, as is also the case in the study reported here.

In order to account for the developmental path across proficiency levels in child L2 acquisition, an adapted schema model with a second-order schema has recently been proposed (Köpcke and Wecker, 2017; Wecker, 2016). These studies argue that second-order schemata, which are intended to capture paradigmatic relationships such as the contrast between singular and plural forms, are acquired late in child L2 German learners. For our data with adult foreign language learners, on the other hand, the paradigmatic contrast—which, as indicated above, overlaps with the drive for iconicity, saliency, and multisyllabicity—appears strongest at the very beginning of the language acquisition process. Future research should continue to investigate how the acquisitional path in the data presented here could be accounted for with a further revised schema model.[20]

5.2.5 The relevance of prosody in acquisition and theoretical models

Overall, the interplay of morphology and prosody in our exploratory analysis suggests that prosody plays a central role in the acquisitional path in our data. The role of prosody materializes in several forms: There appears to be an initial requirement for *multisyllabic* plurals (thus rejecting monosyllabic plurals in *-s*), as well as saliency, here defined as adding a *syllabic* suffix to a singular form to mark plurality. Thus, the nuanced analyses of the (word) Classes show that prosody plays a foundational role in the plural formation from the beginning, even if it might not lead to target-like trochaic productions or ratings, and even if it might not be the only factor involved in plural formation.

Yet, in the end, the available data to date do not allow us to unambiguously decide between various theoretical models and accounts of German plural formation or acquisition. Crucially, the data from Year 1 and subsequent (proficiency) Levels suggest that syllabic suffixes—notably *-en*—on words ending in closed schwa-syllables reduce with increasing proficiency, making way for more null-suffixes and, in particular, *-n* suffixes. On the one hand, this could be accounted for by a trochaic pattern for plurals becoming stronger with increasing proficiency. On the other hand, this could also be accounted for by L2 users learning the distribution of suffixes, i.e., which types of (word) Classes co-occur with which plural suffixes across (proficiency) Levels. This latter account would not necessarily require a prosodic condition to capture the same pattern of data. The trochaic pattern in L1 and especially more advanced L2 users may then be a mere by-product (e.g., Trommer, 2021) or just one of many cues to plurality (see Köpcke and colleagues' schema model, e.g., Köpcke, 1988; Köpcke and Wecker, 2017). We have sketched an initial proposal of how the acquisitional path could be captured with the strategies for plural marking in the schema model, although it appears that the relevance of specific cues and their interaction would have to be different than the proposed model for child L2 acquisition (Köpcke and Wecker, 2017; Wecker, 2016).

5.2.6 Limitations and open questions

Readers should keep in mind that the data presented in this study are cross-sectional. Ideally, further research would add a longitudinal perspective to the cross-sectional data presented here to follow the same learners as they develop their L2 language skills over time. This kind of work could then also test whether and how these prosodic patterns develop within individuals over the course of their L2 acquisition process and how this might align with the PTH (cf. also Cabrelli, 2019). Such a study could provide insights into individual acquisitional paths and individual differences related to both the production and perceptivity of prosodic patterns in plural forms, and their potential interaction with other plural markings, specifically, umlaut and suffixal choice for plurals. Additional caution needs to be taken when interpreting the results reported above due to imbalances in the data. The analyses of the interaction of prosody and morphology—(word) Class and Ending/EndingChosen—presented here should be considered exploratory and will need to be ratified in future work.

Finally, while not explicitly tested here, the similarity between participants' L1 English and the target language German may contribute to facilitating effects that may not necessarily be replicated with L2 users whose L1 is prosodically different from German. For instance, English-speaking and French-speaking adult L2 German learners differ in their preferences for lexical stress assignment based on their L1 (O'Brien and Sundberg, 2023). More cross-linguistic work in this area could be another testing ground for the role of prosody and perhaps further examine the validity of accounts with prosodic constraints or templates. English and German share how prosodic prominence is used to mark word stress, and the trochee is a prevalent pattern in German, English, and Dutch (cf. Domahs et al., 2014). Other L1 language backgrounds might include pitch-accent languages, tonal languages, or a language that might not utilize (trochaic) feet, as has been suggested for Portuguese (cf. Garcia and Goad, 2021). Garcia and Goad (2021) argue that, while metrical stress data for English align with a foot-based analysis, Portuguese metrical stress is not captured "optimally" with an analysis that assumes feet. The authors provide additional evidence from sonority effects and word minimality issues from both languages in support of their analysis that some languages may not build feet. Thus, future research could investigate L1 Portuguese learners of German to determine whether this group of L2 users would

20 In fact, words ending in closed schwa-syllables are central to the schema model and present one promising avenue for further research among our adult L2 users.

show patterns that align with a foot-based prosodic analysis of German plurals.

6 Conclusions

Although many descriptive and theoretical accounts of German plurals have drawn on the syllabic or prosodic structures of nouns, there have hitherto only been a few empirical studies testing the productivity of this prosodic aspect or the interaction of prosody and morphology for German plurals in clinical or acquisitional contexts (but see Domahs et al., 2017; Kauschke et al., 2011, 2013; Vogt, 2016). The data presented in this study arguably confirm that the word-final trochee requirement is a productive pattern in L1 users of German. Similarly, the trochaic pattern progressively developed as a productive pattern in the plural elicitation data from L1 English-L2 German speakers: As proficiency increased, L2 users produced more trochaic plural forms, both in non-words and in existing words of the German lexicon, whether pluralized accurately or not. The descriptive data from the rating study further suggest that this prosodic structure in plurals develops concomitantly as learners progress in their L2 proficiency. In fact, throughout the study tasks, by the third year of university study, L2 users produced and preferred (in their ratings) trochaic plural forms.

To this end, the results of our behavioral psycholinguistic study are consistent with an account in which the trochaic template for German nominal plural formation is part of L1 users' grammars and mental representations, and develops with increasing proficiency in L1 English-L2 German participants' grammars. Yet, based on the available data, we cannot unambiguously rule out accounts of German plurals in which the prosodic pattern is merely epiphenomenal or a by-product of morphological patterns (e.g., Trommer, 2021) rather than a prosodic constraint or principle that learners need to acquire separately from the morphology. In the latter case, learners might still produce and prefer trochaic forms—or develop these with increasing input and target language proficiency—but without the need for a separate constraint or template. Our exploratory analyses of the interplay between prosody and morphology of German plurals suggest that iconicity, saliency, and multisyllabicity—factors from Köpcke's schema model (i.e., Köpcke, 1988, *inter alia*)—could explain the early phases of the acquisitional path in the adult foreign language learning data presented here.

We leave it open for future psycholinguistic research, and perhaps computational modeling, to further examine whether L1 and L2 language users develop sensitivities for the prosodic patterning itself, which would be in line with generative accounts (e.g., Schuhmann and Putnam, 2021; Smith, 2004, 2020; Wiese, 1996, 2009), or whether users primarily develop sensitivities for the distribution of the plural allomorphs. In the end, we hope that the findings presented here invite further cross-linguistic inquiries into the development of prosodic patterns in the acquisition of grammar in L1 and L2 users and into how prosodic cues and morphology or other higher-level linguistic structures interact during L2 development in various language learning contexts.

Data availability statement

The datasets analyzed for this study can be found in the Open Science Framework (osf.io): https://osf.io/nfpb5/. Further inquiries can be directed to the corresponding author KS.

Ethics statement

The studies involving humans were approved by Institutional Review Board (IRB) at Brigham Young University. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

KS: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Software, Writing – review & editing. LS: Conceptualization, Data curation, Investigation, Methodology, Project administration, Writing – review & editing, Funding acquisition, Software, Writing – original draft.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Funding was provided by KS's BMBF funds (for research assistants), NiedersachsenOPEN (for publication costs), and LS's research funds (for research assistants and study software).

Acknowledgments

We are grateful to all the L1 and L2 German users who took the time to participate in our study. Sincere thanks go to the students and student assistants at BYU and at UOL who helped with data collection and data coding. This study and the larger project surrounding the study reported here have also benefited tremendously from discussions with colleagues, including at several conferences. We would also like to thank Ankelien Schippers for statistical consultation on some of the analyses presented in this manuscript. Finally, we are thankful for the feedback from the reviewers and editors. Any errors are our own.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1338625/full#supplementary-material

References

Archibald, J. (2022). "Feature dependency and the poverty of the stimulus in the acquisition of L2 German plural allomorphy," in *Generative SLA in the Age of Minimalism: Features, interfaces, and beyond. Selected Proceedings of the 15th Generative Approaches to Second Language Acquisition Conference*, eds. T. Leal, E. Shimanskaya, and C. A. Isabelli (Amsterdam: John Benjamins), 117–136.

Augst, G. (1979). Neuere Forschungen zur Substantivflexion. *Zeitschrift für Germanistische Linguistik* 7, 220–232.

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

Baayen, R. H., and Schreuder, R. (1999). War and peace: morphemes and full forms in a non-interactive activation parallel route model. *Brain Lang.* 68, 27–32. doi: 10.1006/brln.1999.2069

Bates, D. M, Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Beretta, A., Campbell, C., Carr, T. H., Huang, J., Schmitt, L. M., Christianson, K., et al. (2003). An ER-fMRI investigation of morphological inflection in German reveals that the brain makes a distinction between regular and irregular forms. *Brain Lang.* 85, 67–92. doi: 10.1016/S0093-934X(02)00560-6

Bermúdez-Otero, R. (2018). "Stratal phonology," in *The Routledge Handbook of Phonological Theory*, eds. S. J. Hannahs and A. R. K. Bosch (Abingdon: Routledge), 100–134.

Binanzer, A., and Wecker, V. (2020). "Lernergrammatik und Grammatikunterricht – Überlegungen zu natürlichen Erwerbssequenzen und didaktischer Progression," in eds. L. Miriam *Grammatikdidaktik – theoretische und empirische Zugänge zu sprachlicher Heterogenität* (Baltmannsweiler: Schneider), 61–85.

Bittner, D. (1991). *Von starken Feminina und schwachen Maskulina: die neuhochdeutsche Substantivflexion; eine Systemanalyse im Rahmen der natürlichen Morphologie* (PhD diss.). Universität Jena, Jena, Thuringia, Germany.

Booij, G. (1998). "Phonological output constraints in morphology," in *Phonology and Morphology of the Germanic Languages*, eds. W. Kehrein, and R. Wiese (Tübingen: Max Niemeyer Verlag), 143–163.

Booij, G. (2002). "Prosodic restrictions on affixation in Dutch," in *Yearbook of Morphology 2001*, eds. G. Booij and J. van Marle (Dordrecht: Springer), 183–201.

Cabrelli, J. (2019). Plotting individual learning trajectories in the acquisition of L2 prosodic constraints. *Linguist. Approaches Biling.* 9, 822–826. doi: 10.1075/lab.19065.cab

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of German inflection. *Behav. Brain Sci.* 22, 991–1013. doi: 10.1017/S0140525X99002228

Clahsen, H., Eisenbeiss, S., and Sonnenstuhl-Henning, I. (1997). Morphological structure and the processing of inflected words. *Theoret. Linguist.* 23, 201–250. doi: 10.1515/thli.1997.23.3.201

Clahsen, H., Rothweiler, M., Woest, A., and Marcus, G. F. (1992). Regular and irregular inflection in the acquisition of German noun plurals. *Cognition* 45, 225–255. doi: 10.1016/0010-0277(92)90018-D

Clahsen, H., Sonnenstuhl, I., and Blevins, J. (2003). "Derivational morphology in the German mental lexicon: A dual mechanism account," in *Morphological Structure in Language Processing*, eds. R. Baayen and R. Schreuder (Berlin, New York: De Gruyter Mouton), 125–156.

Demuth, K. (1994). "On the 'underspecification' of functional categories in early grammars," in *Syntactic Theory and First Language Acquisition: Cross-Linguistic Perspectives*, eds. B. Lust, M. Suñer, and J. Whitman (Hillsdale, NJ: Lawrence Erlbaum Associates), 119–34.

Demuth, K. (2009). "The prosody of syllables, words and morphemes," in *Cambridge Handbook on Child Language,* ed. E. Bavin (Cambridge: Cambridge University Press), 183–198.

Demuth, K., and McCullough, E. (2009). The prosodic (re)organization of children's early English articles, *J. Child Lang.* 36, 173–200. doi: 10.1017/S0305000908008921

Domahs, F., Bartha-Doering, L., Domahs, U., and Delazer, M. (2017). "Wie muss ein guter deutscher Plural klingen?," in *Sichtbare und hörbare Morphologie,* eds. N. Fuhrhop, R. Szczepaniak, and K. Schmidt (Berlin, Boston: De Gruyter), 205–236.

Domahs, U., Lohmann, K., Moritz, N., and Kauschke, C. (2013). The acquisition of prosodic constraints on derivational morphology in typically developing children and children with SLI. *Clini. Linguist. Phonet.* 27, 555–573. doi: 10.3109/02699206.2013.796405

Domahs, U., Plag, I., and Carroll, R. (2014). Word stress assignment in German, English and Dutch: quantity-sensitivity and extrametricality revisited. *J. Comparat. Germa. Linguist.* 17, 59–96. doi: 10.1007/s10828-014-9063-9

Duden-Deutsches Universalwörterbuch (2023). Dudenredaktion (Ed.). *Duden. Deutsches Universalwörterbuch. 10., völlig überarbeitete und erweiterte Auflage.* Berlin: Dudenverlag.

Duden-Grammatik (1995). Drosdowski, Günther (Ed.). *Duden. Grammatik der deutschen Gegenwartssprache. 5., völlig neu bearbeitete und erweiterte Auflage.* Mannheim: Dudenverlag.

Duden-Grammatik (1998). Dudenredaktion (Ed.). *Duden. Grammatik der deutschen Gegenwartssprache. 6., neu bearbeitete Auflage.* Mannheim: Dudenverlag.

Eisenberg, P. (2020). *Grundriss der deutschen Grammatik. Band 1: Das Wort. 5. Aufl.* Stuttgart: Metzler.

Féry, C. (1994). Umlaut and inflection in German. *ROA-34, Rutgers Optimality Archive*. Available at: http://roa.rutgers.edu/files/34-1094/34-1094-F%E9RY-0-0.PDF (accessed August 31, 2024).

Garcia, G. D., and Goad, H. (2021). Weight Effects and the Parametrization of the Foot: English vs. Portuguese. Available at: https://osf.io/pc6hg/download (accessed August 31, 2024).

Gerken, L. (1996). Prosodic Structure in Young Children's Language Production. *Language* 72, 683–712. doi: 10.2307/416099

Gleason, J. B. (1958). *The Wug Test.* Cambridge: Larchwood Press.

Goad, H., and White, L. (2019). Prosodic effects on L2 grammars. *Linguist. Approaches Biling.* 9, 769–808. doi: 10.1075/lab.19043.goa

Hahn, U., and Nakisa, R. C. (2000). German inflection: single route or dual route? *Cognit. Psychol.* 41, 313–360. doi: 10.1006/cogp.2000.0737

Hall, T. A. (1992). *Syllable structure and syllable related processes in German.* Tübingen: Niemeyer.

Itô, J., and Mester, A. (1997). "Sympathy Theory and German truncations," in *Proceedings of the Hopkins Optimality Workshop/Maryland Mayfest 1997*, eds. by V. Miglio and B. Moren (College Park, MD: University of Maryland), 117–138.

Kauschke, C., Kurth, A., and Domahs, U. (2011). Acquisition of German noun plurals in typically developing children and children with specific language impairment. *Child Dev. Res.* 2011:718925.

Kauschke, C., Renner, L., and Domahs, U. (2013). Prosodic constraints on inflected words: an area of difficulty for German-speaking children with specific language impairment? *Clini. Linguist. Phonet.* 27, 574–593. doi: 10.3109/02699206.2013.798033

Kiparsky, P. (2015). "Stratal OT: A synopsis and FAQs," in *Capturing Phonological Shades*, eds. E. Y. Hsiao and L. Wee (Cambridge: Cambridge Scholars Publishing), 2–44.

Köpcke, K.-M. (1988). Schemas in German plural formation. *Lingua* 74, 303–335. doi: 10.1016/0024-3841(88)90064-2

Köpcke, K.-M. (1993). *Schemata bei der Pluralbildung im Deutschen.* Tübingen: Narr.

Köpcke, K.-M. (1998). The acquisition of plural marking in English and German revisited: schemata versus rules. *J. Child Lang.* 25, 293–319. doi: 10.1017/S0305000998003407

Köpcke, K. M., Schimke, S., and Wecker, V. (2021). Processing of German noun plurals: Evidence for first-and second-order schemata. *Word Struct.* 14, 1–24. doi: 10.3366/word.2021.0173

Köpcke, K. M., and Wecker, V. (2017). Source-and product-oriented strategies in L2 acquisition of plural marking in German. *Morphology* 27, 77–103. doi: 10.1007/s11525-016-9292-5

Laaha, S. (2011). Sonority, gender and the impact of suffix predictability on the acquisition of German noun plurals. *Lang. Interact. Acquisit.* 2, 82–100. doi: 10.1075/lia.2.1.04laa

Laaha, S., Ravid, D., Korecky-Kröll, K., Laaha, G., and Dressler, W. U. (2006). Early noun plurals in German: regularity, productivity or default? *J. Child Lang.* 33, 271–302. doi: 10.1017/S0305000906007379

Lenth, R. V. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.1-1*. Available at: https://CRAN.R-project.org/package=emmeans (accessed August 31, 2024).

Lleó, C., and Demuth, K. (1999). "Prosodic constraints on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages," in *Proceedings of the 23rd Annual Boston University Conference on Language Development, Vol. 2*, eds. A. Greenhill, H. Littlefield, and C. Tano (Cascadilla Press), 407–418.

Marcus, G. F., Brinkmann, U., Clahsen, H., Wiese, R., and Pinker, S. (1995). German inflection: the exception that proves the rule. *Cogn. Psychol.* 29, 189–256. doi: 10.1006/cogp.1995.1015

McCarthy, J., and Prince, A. (1994). "The emergence of the unmarked: Optimality in Prosodic Morphology," in *Proceedings of the North East Linguistics Society,* ed. M. Gonzàlez (Amherst, MA: Graduate Linguistic Student Association), 333–379. doi: 10.7282/T3Z03663

McCarthy, J. J., and Prince, A. (1996). "Prosodic morphology 1986," in *Linguistics Department Faculty Publication Series*, 13. Available at: http://ruccs.rutgers.edu/roa.html (accessed August 31, 2024).

Mugdan, J. (1977). *Flexionsmorphologie und Psycholinguistik. Untersuchungen zu sprachlichen Regeln und ihrer Beherrschung durch Aphatiker, Kinder und Ausländer, am Beispiel der deutschen Substantivdeklination.* Tübingen: Gunter Narr.

Neef, M. (1998). "The reduced syllable plural in German," in *Models of Inflection*, eds. R. Fabri, A. Ortmann, and T. Parodi (Tübingen: Niemeyer), 244–265.

Noske, R. (1993). *A theory of syllabification and segmental alternation: With studies on the phonology of French, German, Tonkawa, and Yawelmani.* Tübingen: Niemeyer.

O'Brien, M., and Sundberg, R. (2023). Lexical stress assignment preferences in L2 German. *Int. Rev. Appl. Linguist. Lang. Teach.* 61, 449–478. doi: 10.1515/iral-2020-0104

Pater, J. (2009). Weighted constraints in generative linguistics. *Cognit. Sci.* 33, 999–1035. doi: 10.1111/j.1551-6709.2009.01047.x

Penke, M., and Krause, M. (2002). German noun plurals: a challenge to the dual-mechanism model. *Brain Lang.* 81, 303–311. doi: 10.1006/brln.2001.2526

Pinker, S., and Ullman, M. T. (2002). The past and future of the past tense. *Trends Cogn. Sci.* 6, 456–463. doi: 10.1016/S1364-6613(02)01990-3

Prince, A. and Smolensky, P. (1993/2004). *Optimality Theory: Constraint Interaction in Generative Grammar.* Hoboken, NJ: John Wiley.

R Core Team (2021). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/ (accessed August 31, 2024).

Revithiadou, A. (2007). "Colored turbid accents and containment: a case study from lexical stress," in *Freedom of Analysis?*, eds. S. Blaho, P. Bye, and M. Krämer (Berlin: De Gruyter), 149–174.

Schreuder, R., and Baayen, R. H. (1995). "Modeling morphological processing," in *Morphological Aspects of Language Processing,* L. B. Feldman (Hillsdale, NJ: Erlbaum), 131–154.

Schuhmann, K. (2015). "Degrees of (un)acceptability in syllable contact," in *Rutgers Optimality Archive*, 1–30. Available at: http://roa.rutgers.edu/article/view/1496 (accessed August 31, 2024).

Schuhmann, K. S., and Putnam, M. (2021). Relativized prosodic domains: a late-insertion account of German plurals. *Languages* 6, 1–20. doi: 10.3390/languages6030142

Schuhmann, K. S. and Smith, L. C. (accepted). "Interindividuelle Unterschiede in der Pluralallomorphie von Kunstwörtern in L1 und L2 Deutsch-Sprachverwendenden," in *Beiträge zur Fremdsprachenforschung (BFF) 18. (Dokumentation zum 30. Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF), Freiburg 2023)*, eds. N. Kulovics, O. Mentz, and T. Raith (Bielefeld: Schneider bei wbv).

Schuhmann, K. S., and Smith, L. C. (2022). Practical prosody: a new hope for teaching German plurals. *Die Unterrichtspraxis/Teaching German* 55, 1–24. doi: 10.1111/tger.12192

Smith, L. C. (2004). *Cross-Level Interactions in West Germanic Phonology and Morphology* (Dissertation). University of Wisconsin–Madison, Madison, WI, United States.

Smith, L. C. (2020). "The role of foot structure in Germanic," in *Cambridge Handbook of Germanic Linguistics*, eds. R. Page, and M. Putnam (Cambridge: Cambridge University Press), 49–72.

Smith, L. C. (2022). "Moving from syllables to feet in the prosodic hierarchy: how foot-based templates reflect prosodic preferences," in *Linguistic Preferences*, eds. P. Noel Aziz Hanna and L. Smith (Berlin, Boston: De Gruyter Mouton), 109–140.

Smith, L. C. (2024). A case for Old Frisian vowel balance as prosodically-driven vowel reduction. *Amsterdamer Beiträge zur älteren Germanistik*. 84, 372–394. doi: 10.1163/18756719-12340328

Sonnenstuhl, I., Eisenbeiss, S., and Clahsen, H. (1999). Morphological priming in the German mental lexicon. *Cognition* 72, 203–236. doi: 10.1016/S0010-0277(99)00033-5

Sonnenstuhl-Henning, I. (2010). *Deutsche Plurale im mentalen Lexikon: Experimentelle Untersuchungen zum Verhältnis von Speicherung und Dekomposition*. Tübingen: Max Niemeyer Verlag.

Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German speaking children. *First Lang.* 21, 109–141. doi: 10.1177/014272370102106201

Trommer, J. (2021). The subsegmental structure of German plural allomorphy. *NLLT* 39, 601–656. doi: 10.1007/s11049-020-09479-7

van der Hulst, H., and Kooij, J. (1998). "Prosodic choices and the Dutch nominal plural," in *Phonology and Morphology of the Germanic Languages*, eds. W. Kehrein, and R. Wiese (Tübingen: Max Niemeyer Verlag), 187–197.

van Oostendorp, M. (2008). Incomplete devoicing in formal phonology. *Lingua* 118, 1362–1374. doi: 10.1016/j.lingua.2007.09.009

Vogt, B. (2016). Prosodiegeleitete Grammatik: Zum Erwerb der Pluralflexion im Deutschen. *German Foreign Language* 2, 67–80. Available at: http://gfl-journal.de/wp-content/uploads/2024/06/20160206-vogt.pdf (accessed August 31, 2024).

Wecker, V. (2016). "Strategien bei der Pluralbildung im DaZ-Erwerb," in *Eine Untersuchung mit russisch und türkischsprachigen Lernern*. Berlin: De Gruyter.

Wegener, H. (1995). *Die Nominalflexion des Deutschen, verstanden als Lerngegenstand*. Tübingen: Max Niemeyer Verlag.

Wiese, R. (1986). Schwa and the structure of words in German. *Linguistics* 24, 697–724.

Wiese, R. (1996). *The Phonology of German*. Oxford: Oxford University Press.

Wiese, R. (2009). The grammar and typology of plural noun inflection in varieties of German. *J. Comparat. German. Linguist.* 12, 137–173. doi: 10.1007/s10828-009-9030-z

Wunderlich, D. (1999). *German Noun Plural Reconsidered*. Düsseldorf: University of Düsseldorf. Available at: https://user.phil.hhu.de/~wdl/Nplural.pdf (accessed August 31, 2024).

Wurzel, W. (1984). *Flexionsmorphologie und Natürlichkeit*. Berlin: Akademie-Verlag.

Wurzel, W. U. (1998). "Drei Ebenen der Struktur von Flexionsparadigmen," in *Models of Inflection*, eds. R. Fabri, A. Ortmann, and T. Parodi (Tübingen: Niemeyer), 225–243.

Zaretsky, E., Neumann, K., Euler, H., and Lange, B. (2013). Pluralerwerb im Deutschen bei russisch- und türkischsprachigen Kindern im Vergleich mit anderen Migranten und monolingualen Muttersprachlern. *Zeitschrift für Slawistik* 58, 43–71. doi: 10.1524/slaw.2013.0003

Zimmermann, E. (2014). *A Phonological Account of Morphological Length* (PhD diss.) Universität Leipzig, Leipzig, Germany.

Zimmermann, E. (2017). *Morphological Length and Prosodically Defective Morphemes*. Oxford: Oxford University Press.

frontiers | Frontiers in Language Sciences

TYPE Original Research
PUBLISHED 14 February 2025
DOI 10.3389/flang.2024.1254956

OPEN ACCESS

EDITED BY
John Archibald,
University of Victoria, Canada

REVIEWED BY
Mariko Nakayama,
Tohoku University, Japan
Christine Shea,
The University of Iowa, United States

*CORRESPONDENCE
John H. G. Scott
✉ jhgscott@umd.edu

RECEIVED 07 July 2023
ACCEPTED 02 September 2024
PUBLISHED 14 February 2025

CITATION
Scott JHG (2025) Detargeting the target in phoneme detection: aiming the task at phonological representations rather than backgrounds. *Front. Lang. Sci.* 3:1254956. doi: 10.3389/flang.2024.1254956

Detargeting the target in phoneme detection: aiming the task at phonological representations rather than backgrounds

John H. G. Scott[1,2]*

[1]Department of German Studies, School of Languages, Literatures and Cultures, University of Maryland, College Park, MD, United States, [2]L2+ Sound Learning Lab, Division of German, Russian, Arabic Language and Muslim Cultures, School of Languages, Linguistics, Literatures and Cultures, University of Calgary, Calgary, AB, Canada

One challenge of learning a second or additional language (L2+) is learning to perceive and interpret its sounds. This includes acquiring the target language (TL) contrastive phonemic inventory, the sounds' systematic behavior in the TL phonology, and novel relationships between spelling and sound (GPCs; grapheme-phoneme correspondences). Many perception tasks require stipulation of written labels for target speech sounds (e.g., phoneme detection). Listening for this target is not necessarily, or even frequently, an equivalent cognitive task between participant groups. The incongruence of phonological and orthographic domains and their GPCs poses a methodological challenge for L2+ research. The author argues that phoneme detection tasks should avoid the phone of investigative interest (*x*) as the direct target of listener attention and redirect focus to an adjacent listening target (*y*). Ideally, this target should not trigger or otherwise be implicated in the phonological process or phonotactic constraint under investigation. The careful choice of listening target (*y*) with both a familiar sound and a congruent orthographic label for both (or all) language groups of the experiment yields an equivalent task and better indicates implicit knowledge of the phenomenon under study. This approach opens up potential choices of phonological objects of interest (*x*). The two phoneme detection experiments reported here employ this novel adjacent-congruent listening target approach, which the author calls the Persean approach. Experiment 1 establishes baseline performance in two assimilation types and replicates processing inhibition in first-language (L1) German speakers in response to violations of regressive nasal assimilation. It also uses [t] as the Persean listening target to test sensitivity to preceding violations of progressive dorsal fricative assimilation (DFA). Experiment 2 investigates sensitivity to violations of DFA in both L1 German speakers and L1 English L2+ German learners. Experiment 2 also uses the Persean method for the first phoneme detection investigation demonstrating sensitivity to violation of a prosodic/phonotactic constraint banning /h/ in syllable codas. The study demonstrates that phoneme detection with Persean listening targets is a viable instrument for investigating regressive and progressive assimilation, prosodic/phonotactic constraints, and prelexical perceptual repair strategies in different language background groups and proposes statistical best practices for future phoneme detection research.

KEYWORDS

phoneme detection, L2+ perception, reaction times, phonological representation, grapheme–phoneme correspondence, underspecification, assimilation, German

1 Introduction

Language users have different first-language (L1) and prior-language experience profiles, which poses an inherent methodological challenge for intergroup task parity when investigating cross-dialect or cross-language perception and second or additional language (L2+) phonology. Speakers of different languages may perceive a sound differently, have different familiarity with the sound (a familiar phone or allophone for one group is a novel phone for another), or may label it with different sets of letters (single sound-to-letter correspondence for one group and multiple spellings allowed for the sound in another group). When learners seek to acquire the phonology of L2+, one aspect that they must learn is the contrastive phonemic inventory (as well as predictable allophonic variants). Determining whether a particular learner has acquired a particular phoneme presents certain challenges to the psycholinguist. Because we cannot look directly at the phonological grammar, we must turn to a range of experimental tasks and then interpret the behavioral results to infer the relevant properties of the grammar. For example, imagine that we wonder if a learner has acquired a phonemic representation for front rounded vowels /yː ʏ/ in their L2+ German (represented orthographically as <ü>[1]), particularly if the L1 inventory lacks the /yː ʏ/ pair. We commonly present tasks to see if they can reliably *identify* or *discriminate* sounds, such as the corresponding back rounded [uː ʊ] and front unrounded [iː ɪ] pairs, from front rounded [yː ʏ] in German words. Alternatively, we may want to see if participants can simply *detect* an [yː] or [ʏ] in a word or a phrase. Detecting the sound of interest (call it x) tells us something about the representation of x in the learner's interlanguage (IL). However, many factors can influence the behavioral results of this sort of *phoneme detection* task. Is x absent from or frequent in the L1? If present, is x a predictable allophone or a full phoneme? How is it represented featurally? Is it frequent or rare in the L2+ lexicon and usage? Is the phone reliably encoded in the orthography? All these factors have been shown to influence phonological identification and detection tasks (Bassetti et al., 2015b; Connine, 1994, p. 115–116; Connine and Titone, 1996, p. 639; Cutler and Otake, 1994; Darcy et al., 2007; Frauenfelder and Seguí, 1989; Otake et al., 1996; Scott and Darcy, 2023; Scott et al., 2022; Seguí and Frauenfelder, 1986). The literature presents a complex and, at times, contradictory picture of what the phoneme detection task can tell us.

The phoneme detection task measures accuracy and reaction time (RT) in response to detecting a specified listening target in the stimulus. As with many RT methods, behavioral responses to phoneme detection (accuracy, systematic changes in processing speed) are employed as proxy measures representing underlying grammatical knowledge (Hui and Jia, 2024). The phoneme detection task has the advantage that it does not require high target-language (TL) proficiency or lexical knowledge. As such, it is useful for investigating prelexical processing, even with pre-learner and early L2+ learner groups, as long as the listening target is viable and congruent between languages. In this article, I introduce a new variant of the phoneme detection task to shed light on some phonological representations in L1 and L2+. In this variant of the phoneme detection task, participants do not focus on detecting the novel L2+ sound of interest x but rather attend to a sound that occurs *adjacent* to the object of interest (call it y). When y is not implicated in the phonological phenomenon of interest, I call this the Persean approach, in reference to how Perseus required the reflection of his shield to look on the Gorgon Medusa's face without being turned to stone by her direct gaze. I explain why having participants detect x directly can be as fatal an error as looking directly at Medusa, particularly when investigating multiple language-background groups or if the aim is to investigate *implicit*, or what may be called *optimum* or *automatized explicit*, knowledge in cross-language or IL phonological perception, all of which have theoretical and practical relevance for L2+ acquisition research (Bordag et al., 2021; Rebuschat, 2013; Strange, 2011; Suzuki, 2017). The experimental results of this Persean approach reveal that the detection of y can tell us something about the nature of the representation of x, adding an important tool to our methodological toolbox.

In Section 2, I highlight the difficulties that arise in task design for phoneme detection experiments. I focus on the problems found in choosing listening targets for L2+ learner experiments, especially regarding task parity for intergroup comparison. In Section 3, I review the sparse literature using phoneme detection tasks to investigate two place assimilation phenomena, right-to-left regressive nasal assimilation (RNA) and progressive (left-to-right) dorsal fricative assimilation (DFA), in German, and critically examine their choices of listening targets with this task. In Section 4, I briefly summarize the prosodic ban on /h/ in syllables codas in English and German to lay the groundwork for experiment 2, which conducts the first phoneme detection investigation of syllable structure constraints governing segment distribution. Then, in Section 5, I outline a strategic innovation to the phoneme detection task designed to avoid the potential methodological pitfalls described (adapted from Otake et al., 1996). The aim of this innovation is to thread the methodological needle of listening target labels in L2+ perception studies by focusing the listener's attention not on the actual object of interest (x) but rather on an *adjacent* listening target (y): a Persean approach to steal a glimpse of the Gorgon. This adjacent target should be (a) familiar to both L1 and TL phoneme inventories and (b) not directly implicated in the phonological process or phonotactic constraint under investigation (i.e., neither the trigger of the phonological process nor the phone to which the phonological process or constraint applies).

I present the research questions in Section 6, and in Sections 7, 8, I report on two experiments that serve as test cases for the modified phoneme detection method, based on studies originally reported by Scott (2019a,b). The first tests the modified phoneme detection method in L1, investigating German RNA and DFA in L1 German speakers. This experiment is a replication and expansion of studies by Otake et al. (1996) and Weber (2001a,b, 2002). The nasal data may offer insight into theories of phonological feature (under-)specification and variation as they relate to place assimilation. The second experiment investigates German DFA with L1 German and L1 American English L2+ German learner groups, following Weber (2001a,b, 2002) and Lindsey (2013; unpublished thesis, Indiana University, Bloomington), whose studies investigated L1 Dutch and L1 American English groups, respectively. It additionally investigates the phonotactic/prosodic

1 ... or rarely <y>. Angle brackets < > indicate an orthographic representation.

ban on /h/ in syllable codas. Crucially, both experiments in this study avoid listening targets with unfamiliar or incongruent orthographic representations, unfamiliar phonetic transcriptions (e.g., Thomson, 2018), or other symbol types (e.g., Thomson, 2011) for listening targets that may be subphonemic (intra-category) variants or that lack graphemic or phonemic congruence between L1 and L2+ phoneme inventories.

2 The problems of labeling listening targets: facing a Gorgon

2.1 Phonological knowledge: more than phones

Phonemes are multifaceted knowledge structures that include the categorization and distribution of phones and, for most adults, orthographic labels. The *categorization* of phones sorts acoustically similar speech sounds into discrete categories according to articulatory features or acoustic cues with various manifestations along several continua (e.g., place of articulation and voice onset time). The *distribution* of phones describes where in a word a language permits a particular phone or allophonic variant to occur. This is phonotactics, a statistical type of well-formedness knowledge that may derive from categorical constraints or probabilistic knowledge based on the lexical (in)frequency of a particular form (Steinberg, 2014, p. 11–17). For reading populations, *orthographic* labels are conventionally used to denote a particular phone or phoneme. These three types of representational knowledge are necessarily connected by mappings that function to associate a specific label to a specific phonetic category in a specific context. Such context may be determined by phonotactic distribution, morphological structure, lexical-semantic content, and other factors. We know very little about how L2+ learners acquire these aspects of phonemes in relation to each other (see Ontogenesis Model: Bordag et al., 2021).

2.2 The underacknowledged problem of orthography

In designing experiments to reveal the properties of phonological representation, one must also take certain orthographic factors into account. Alphabetic literacy—that is, the knowledge of how labels are applied to individual speech sounds by orthographic convention—influences phonological awareness in undeniable but still poorly understood ways. Adult non-readers without alphabetic literacy exhibit a reduced capacity for phonological tasks that require manipulation at the segmental level (e.g., phoneme deletion or detection) relative to former non-readers who have later learned to read. This effect is less pronounced when syllables or rhymes are the unit of focus (Morais et al., 1986). For those with literacy of a script that encodes syllables rather than segments, the development of L1 phonemic awareness may proceed along different paths (Mann, 1986). Investigating the connection between alphabetical literacy and L1 phonemic awareness has a long tradition in reading and cognition research (see Bertelson, 1986b, special issue articles in Bertelson, 1986a, and Castro-Caldas, 2004 for helpful reviews). More recently, research connecting this vein with L2+ phonology is rapidly emerging (e.g., Bassetti et al., 2015a, special edition of *Applied Psycholinguistics*, including a state-of-the-art review by Bassetti et al., 2015b). In addition, qualitative evidence suggests that groups from different L1 orthography backgrounds, despite similar quantitative performance results on the same phonemic awareness tasks, may employ different phonological processing procedures in L2+ scenarios (e.g., Korean vs. Chinese; Koda, 1998).

Just as phonology and lexical items are language-specific, so is orthography. Not only must contrastive phoneme distinctions of the language be represented (*chip* vs. *ship*), but each grapheme–phoneme correspondence (GPC) also has its own phonotactic and morphological distribution in the lexicon (e.g., <sh> and <ch> vs. <ci> and <ti> as labels for English /ʃ/ in *ship*, *fish*, *shanty/chantey*, *chute* vs. *commercial*, *navigation*). Invented L2+ spellings based on L1 GPCs illustrate the specificity of orthography nicely. Consider examples such as <JUELLULIB> and <GUARIYUSEI> ("Where do you live?" and "Waddayasay?"), both attempts by Mexican migrant workers to write down helpful phrases of English spoken in rural southern Illinois with Spanish spellings (Kalmar, 2015, p. 19, 51). On the Ontogenesis Model of lexical representations in L2+ (Bordag et al., 2021), such examples illustrate *fuzziness* in the phonolexical and lexico-semantic representations of the migrant workers' L2+ English (Cook et al., 2016; Darcy et al., 2013). For these speakers, the phonological and orthographic domains, and the mappings (GPCs) between them, lag behind the semantic domain that allows them to use these phrases communicatively.

The problems of incongruent GPCs are similar for phoneme detection tasks, where the label of the listening target may represent different phonological information between groups (e.g., <N> with L1 Japanese vs. L1 Dutch; Otake et al., 1996; <CH> or <G> for L1 German vs. L1 Dutch or L1 English; Lindsey, 2013; Weber, 2001a,b, 2002; [u] for L1 French vs. [u] in [.Cu.] but not [.u.] for L1 Japanese; Dupoux et al., 1999, p. 1,570). In such cases, the intended label represents a phoneme or allophonic variant in one language but not the other (e.g., [x] and [ç] in German vs. American English).

2.3 Why labels are a problem for L2+ learners

As L2+ learners learn the sound system of a new language, they gradually acquire what the sounds are, where the sounds go, how they are written, how they are combined to label lexemes, and the relationships between these components. Just as we do not expect early, intermediate, and even advanced learners to have native-like production, vocabulary, and semantics, we should also not expect learners to have native-like, also called *optimal*, phonological perception and representations in the TL (*optimal encoding*, *optimum range*; Bordag et al., 2021). It is crucial for the design of laboratory phonology studies of L2+ acquisition to take into account that the components of phonological representations may not all be fully optimized in cross-language and subsequent L2+ perception. Indeed, they most likely are neither optimized nor closely yoked. Learners' IL phonological categories, orthographic knowledge, and phonolexical representations both are unevenly optimized and may remain divergent from and less precise than the representations of L1 speakers of the TL (Best and Tyler, 2007; Cook et al., 2016; Darcy et al., 2013). If we drop any assumption that L2+ learners' phonological, phonotactic, and orthographic knowledge is fully optimized, then labels for speech sounds become a problem, as we cannot assume congruent meaning for the label between languages or stages of IL development. The two experiments of this study serve to demonstrate the benefits of employing a phoneme detection task in cross-language and L2+ studies if one can avoid certain methodological concerns that may arise from stipulating listening targets by means of labels that differ in their phonological status between L1 and L2+ groups.

2.4 Why labels are a problem for intergroup comparisons

A necessary condition for experimental research in cross-language and L2+ phonology is that we investigate groups with different profiles of L1 and prior language experience. Yet this also poses a serious methodological challenge for ensuring task parity between groups. We routinely control for such factors as age of acquisition, proficiency, and literacy, among others, but we should also control for the comparability of task demands for the different groups. Many perception tasks require using labels for speech sounds, such as those motivating the Perceptual Assimilation Model (Best, 1995; Best and Tyler, 2007) or a phoneme detection task (aka phoneme monitoring; Foss, 1969). An example of the former would be something along the lines of "When you hear the sound [ɾ] does it sound more like a type of /t/ or a type of /r/ ?" For an English speaker, a [ɾ] might be thought of as a kind of /t/ or <t> (as in words like *city*), while for a Spanish speaker, a [ɾ] might be thought of as a kind of /r/ or <r> (as in words like *pero*). An example of the latter would be "Press the red button if you hear a [θ] in the following sentence." The nature and cognitive load of such tasks, due to the representation of the listening target itself ([ɾ] and [θ] in the preceding examples), may differ between language groups (Otake et al., 1996, p. 3,838–3,840). This depends on factors such as the label's status in the listeners' phonemic inventory, exposure to subphonemic variants, phonological or orthographic nativeness, or the phonological status of certain features of the stimuli for each group (e.g., cue weightings). A specific listening target might be an L1 phoneme to one participant group, an allophonic variant to another, and a novel non-phoneme to a third. For example, to an L1 Japanese speaker, [ɾ] is the straightforward realization of the phoneme /ɾ/, whereas an L1 English speaker may perceive it as a positional allophonic variant of the phoneme /t/ , and an L1 Mandarin speaker may perceive it as a novel non-phoneme (or perceptually assimilate it to /l/ or /t/). Thus, listening for a target can constitute cognitively distinct tasks between groups. Similarly, listening for the target <u> likely would not be equivalent for L1 French listeners and L1 Japanese listeners. In French, the letter <u> typically represents the front rounded vowel phoneme /y/. The back rounded vowel [u] (phonemic /u/) is typically represented in French orthography by a digraph (e.g., <ou>). In contrast, L1 Japanese listeners may interpret <u> as either the syllable "u" represented in *rōmaji* (Roman script) as <u> and in *hirigana* by its own glyph <う>, or as the nuclear constituent of another canonical syllable (e.g., <bu>/<ぶ>, <pu>/<ぷ>, <mu>/<む>; Dupoux et al., 1999, p. 1,570). For them, <u> represents a close back unrounded [ɯ̟] or compressed vowel [ɯ̟$^{\beta}$]. Such examples are common in cross-language and L2+ phonology, with methodological implications for perception research on many language pairings. How should we determine if Japanese or French speakers hear an [u]? Should we ask the French speakers if they hear <ou> but ask the Japanese speakers if they hear <う>? Would this allow us to compare the results across groups? Would such a comparison still be confounded due to fundamentally different vowel qualities or because a French speaker cued to <ou> has one phonological unit to consider, whereas a Japanese speaker has both a full vowel (V) syllable and the nuclei of several consonant vowel (CV) syllables to listen for? Most likely, French and Japanese speakers face tasks with different cognitive processes and different cognitive loads in this case. In addition to creating congruency problems between languages, orthography also can be misleading due to inconsistency within a language when one label does not reliably indicate one sound (e.g., <CH> represents [t͡ʃ] or [ʃ] in English *chant*, *chute*). This sort of confound can be compounded when using an L2+ label to focus attention on an L2+ category listening target that has no analog in the L1 (e.g., using <CH> for [ç] or [x] in German). For such reasons, it is important for ensuring study validity that perception tasks do not rely on listening targets with divergent phonological status between participant groups. To ensure parity between groups for phonological awareness tasks such as phoneme detection, listening targets and the GPCs used to stipulate them to participants should be selected for their congruence, as much as possible, given each group's language background. This study adopts this standard.

3 Phoneme detection for investigating place assimilation

3.1 The classic phoneme detection paradigm

The phoneme detection task was introduced to psycholinguistics by Foss (1969). As characterized by Weber (2001b, p. 12), it is a dual task, entailing the detection of a predetermined target sound in speech presented aurally and then a timed response. Participants indicate their detection (or not) of the listening target in the stimulus by pressing a response key as quickly as possible after the target sound is heard. Like other reaction time methods, accuracy and reaction time (RT) are the dependent variables of interest (Hui and Jia, 2024). To avoid response bias, the target items and distractors are counterbalanced by *fillers* that do not contain the listening target. If real words are used, then semantic priming (and other lexical factors) may affect RT (Frauenfelder and Seguí, 1989; Seguí and Frauenfelder, 1986). If non-words are used, it has been shown that for items that phonologically resemble a real word, RT is relatively faster than for those differing more from real words. Such similarity to extant words or sequences of phones may motivate facilitation even for non-words. This is a potential source of variation in RT for phoneme detection and similar tasks (Connine, 1994, p. 115–116; Connine and Titone, 1996, p. 639). Despite some observed effects of the lexicon on tasks using non-word stimuli, numerous studies support some degree of abstraction from the input before accessing the lexicon, thus placing phonotactic processing (e.g., reinterpretation of raw percepts according to well-formedness conditions; Selkirk, 1984, p. 114) at a prelexical stage of processing, as early as 200 ms after stimulus onset (Dehaene-Lambertz et al., 2000; Steinberg, 2014; Steinberg et al., 2010a,b, 2011; Whalen, 1991). The influence of real-word phonolexical representations—even on non-words—and the early (prelexical) timing of phonotactic processing will be especially relevant for interpreting the results of experiment 2, in which L1 German speakers selectively compensate for the illicit occurrence of [h] in syllable codas but only in vocalic contexts where [x] would be licit.

One way to determine if a participant has acquired a particular process or constraint is to see if they react differently when a given string violates that TL pattern. Like other reaction time procedures such as phonetic decision, repetition, and lexical decision, phoneme detection typically indicates a phoneme's goodness of fit to its context through overall *slower* RT for a mismatching context (Whalen, 1984) and, specifically, violations of obligatory assimilation (Weber, 2001a, p. 96). Following established usage for this experimental paradigm, I refer to these slower RTs as processing *inhibition* (Marslen-Wilson and Warren, 1994; Martin and Bunnell, 1981, 1982; Otake et al., 1996; Streeter and Nigro, 1979; Whalen, 1984, 1991). However, violation of listener expectations under certain conditions may also yield a *faster* RT, which I refer to as processing *facilitation* (Cutler et al., 1987; Mills, 1980; Swinney and Prather, 1980). Because phoneme detection can yield either inhibitory or facilitative RT effects, it may be necessary to analyze different phonological conditions in separate statistical models rather than as levels of the same factor in a unified model, as experiments 1 and 2 demonstrate.

3.2 Place assimilation: nasals and dorsal fricatives

Two of the three phonological phenomena investigated here are types of place assimilation. Place assimilation often yields allophonic variants by context. For example, in English, underlying velar stops typically surface as velar before back vowels but palatal before front vowels (*cougar* [ˈkuː.gɹ] vs. *keener* [ˈciː.nɹ]). This phonetic effect results from regressive place assimilation, where the place of the vowel influences the place of the preceding stop. See Winters (2003) for a historical review of developments in theoretical, typological, and experimental research on place assimilation and an account of its articulatory motivation and Hura et al. (1992) regarding perceptual motivations for assimilation.

Feature geometric approaches to autosegmental phonology have used hierarchical relations between distinctive features to describe rules of place assimilation.[2] Figure 1 displays example assimilation and default rules that specify place for specified and underspecified segments. Feature geometry approaches typically analyze place assimilation as the application of a single *spreading rule*. This might spread Place wholesale, like the example rule in Figure 1A, by which RNA supplies Place specification to an underspecified nasal. Or it might spread a lower-tier feature, such as the example rule in Figure 1C, which adds [CORONAL] to a specified [DORSAL] fricative, a possible analysis of Standard German DFA. Such analyses may describe both regressive/leftward (Figure 1A; e.g., English [k]*ougar* ∼ [c]*eener*) or progressive/rightward assimilation (Figure 1C). Figure 1B depicts a typical *default rule* to supply [CORONAL] to a Place-unspecified segment. Feature underspecification theories avoid delinking and favor *structure-filling* rules like those in Figures 1A, B, or *structure-changing* rules like the one in Figure 1C. Feature underspecification arguments are not uncommon in the German phonology literature (e.g., Glover, 2014; Hall, 1995, 2010). Experiment 1 probes whether specified mismatches behave differently from underspecified mismatches in L2+ learners.

3.3 Regressive Nasal Assimilation (RNA)

3.3.1 RNA in English and German

RNA is typologically widespread but manifests differently between languages (Speeter Beddor and Evans-Romaine, 1995). In English and German, examples of tautomorphemic homorganic nasal-obstruent sequences, commonly argued to arise through place assimilation (e.g., Wiese, 1996), are plentiful (e.g., *ra**mp*** [ɹæ**mp**], *ra**nt*** [ɹæ**nt**], and *ra**nk*** [ɹæ**ŋk**] and German *Ka**mpf*** [ka**mp**͡f] "struggle, combat," *La**nd*** [la**nt**] "country," *Ba**nk*** [ba**ŋk**] "bank"). In English, RNA does not apply to morphologically derived nasal-obstruent sequences (e.g., *dreamt* [dɹɛm-t], *ashamed* [ə.ˈʃe͡ɪm-d]) or word-internally across morpheme boundaries (e.g., *co**nf**ess* [kən.ˈfɛs], *i**nf**inite* [ˈɪn.fɪ.nɪt], *ki**ngp**in* [ˈkɪŋ.pɪn]). The application of RNA is further limited in German, which allows labial nasals before alveolar stops (e.g., *A**mt*** [a**m**t] "office, agency," *He**md*** [hɛ**m**t] "shirt," *Sa**ms**tag* [ˈza**m**s.tak] "Saturday"), as well as rarely before velar stops across syllable boundaries (e.g., *Le**mg**o* [ˈlɛ**m**.goː] "city of Lemgo, North Rhine-Westphalia," *I**mk**er* [ˈɪ**m**.kɐ] "beekeeper," *Ir**mg**ard* [ˈiɐ̯**m**.gaɐ̯d] "proper name (fem.);" examples from Wiese, 1996, p. 218; Wiese, 2011, p. 105). These facts, and a phoneme detection experiment by Weber (2001a,b), suggest that some nasals of English and German are specified for Place and thus resist RNA. Assuming a theory of underspecification and a rule like Figure 1A, English and German nasals may only undergo RNA when not blocked by a prior place specification (e.g., labial /m/). Any nasals still lacking Place undergo repair by a default rule like in Figure 1B. For an additional discussion of the susceptibility of nasals to place assimilation, see Winters (2003).

3.3.2 Phoneme detection investigations of RNA

Otake et al.'s (1996) and Weber's (2001a) phoneme detection studies of RNA crucially inform the approach to listening target stipulation in the present task design. Following Otake et al.'s (1996) investigation of Japanese RNA, Weber (2001a) used a similar task to confirm that violations of RNA yield a similar RT effect in L1 German listeners. Together, these two studies provide the basis for the prediction that violation of RNA will yield slower RT (processing inhibition). Additionally, Otake et al. set

2 See Hura et al. (1992) for relevant theoretical discussion.

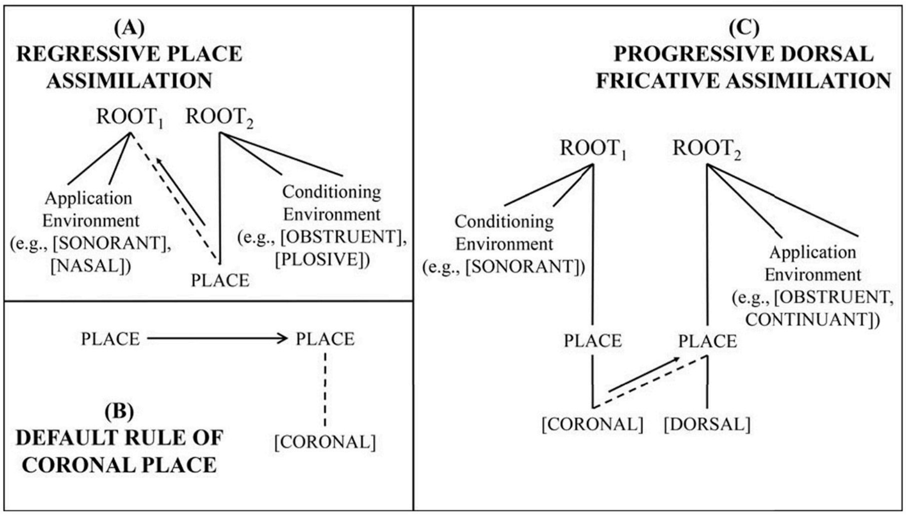


FIGURE 1
Example feature geometry rules: **(A)** Regressive nasal assimilation, **(B)** coronal default, and **(C)** progressive dorsal fricative assimilation.

the stage for using phoneme detection methodology to investigate underlying phonological representation using processing data from a behavioral task.

Otake et al. (1996) report six experiments (summarized in Table 1) investigating the phonemic representation of Japanese moraic nasals in real words based on responses to aurally presented phonetic realizations. In Japanese, moraic nasals—that is, consonant vowel nasal (CVN) syllable codas represented by the final nasal monograph <ん>—undergo complete RNA to the following consonant (including sonorants) obligatorily (Vance, 1987), suggesting a lack of underlying Place specification (Hura et al., 1992, p. 68), which they gain only through RNA. This yields homorganic clusters in the medial position (e.g., *kanpa* [m.p], *tento* [n̪.t̪], *konro* [n.ɾ], *denki* [ŋ.k]). Their experiment 1 (L1 Japanese) and experiment 2 (L1 Dutch comparison) employed naturally produced stimuli. Based on previous research supporting the use of *rōmaji* for phoneme detection with L1 Japanese speakers (Cutler and Otake, 1994), these experiments used the letter <N> to label the listening target (i.e., the nasal undergoing assimilation) for both groups (Otake et al., 1996, p. 3,832). Their experiment 4 (L1 Japanese) investigated sensitivity to the moraic nasal under valid application of RNA (i.e., homorganic clusters) and RT inhibition in response to violations of RNA causing invalid (place mismatched) clusters, for example, **to*[ŋ]*bo*, **ko*[ŋ]*to*, **ko*[m]*to* **ro*[m]*go*, **ro*[n]*go* (p. 3,836), still using <N> as the listening target. Their experiment 5 was a replication of experiment 4 (L1 Japanese, cross-spliced stimuli) with an important design modification: Rather than using the nasal as the listening target, it used the following consonants (<P B D T R K G>) as the listening targets (see Figure 1A, conditioning environment). They presented each listening target visually before each sequence (p. 3,838). Their experiment 6 employed the same procedure for an L1 Dutch comparison group. Overall, Otake et al. (1996) showed three key findings. First, L1 Japanese speakers can rapidly recover an abstract, unitary archiphonemic representation of the moraic nasal from its wide variety of phonetic realizations (see also Darcy et al., 2007, on the recoverability of the underlying phonolexical representations from assimilated stimuli). Second, the phonetic realization of the moraic nasal creates a high expectation that allows L1 Japanese speakers to anticipate the following consonant that conditions RNA (cf. Key, 2014). Third, this language-specific knowledge is not shared by L1 Dutch speakers, who have a fundamentally different phonemic representation of nasals and for whom all stimuli were non-words. Their study also instructively highlights the methodological importance of the listening target for phoneme detection in two aspects—namely, (a) the nature of the label itself and (b) the choice of focus on either the application environment (object of interest *x*) or the conditioning environment of assimilation, as depicted in Figure 1A. I take up both again in Section 5.1. Otake et al.'s (1996) methodological change in experiment 5—to use the following obstruent as the listening target rather than the nasal target of place assimilation—crucially informs the innovation of the [t]-detection condition introduced here in experiment 1 and extended in experiment 2 (see Table 1).

3.4 Dorsal Fricative Assimilation (DFA)

3.4.1 Phonological accounts of DFA in German

Like RNA and the English *cougar–keener* example, DFA manifests as coarticulation of adjacent obstruents and sonorants. In contrast to these, Standard German DFA is progressive rather than regressive. The German dorsal fricatives, commonly called the *ich-* and *ach-* sounds or "front and back *ch*," are the voiceless palatal [ç] and velar [x], respectively (Dollenmayer et al., 2014, p. 192; Valaczkai, 1998, p. 112–114). Supplementary material A summarizes the phonetic characteristics of voiceless fricatives to contextualize these phones and their acoustic cues in the contrastive inventory of German. Spectral quality will be relevant for analyzing the L1 German results in experiment 2.

In Standard German and many regional dialects, palatal [ç] and velar [x] comprise a non-contrastive front–back allophonic pair

TABLE 1 Listening target alignments summary of phoneme detection task designs.

References	Location and identity of listening target					
	Nasal assimilation (regressive)			Dorsal fricative assimilation (progressive)		
	V*N*C	V*N*C Target = *x*	V*N*C	X*F*C	X*F*C Target = *x*	X*F*C Target = *y*
Otake et al. (1996)	–	Exp 1, 2, 4	Exp 5, 6	–	–	–
Weber (2001a)	–	–	Exp 4	–	Exp 1, 2, 3	–
Weber (2001b)	–	–	Exp 4	–	Exp 1, 2, 3, 5	–
Weber (2002)	–	–	–	–	Exp 1	–
Lindsey (2013)	–	–	–	–	Exp 1, 2	–
Present study	–	–	Exp 1	–	–	Exp 1, 2

Adapted from Scott (2019a, p. 286, Table 5.1). C represents an obstruent consonant. For nasal assimilation, N represents the nasal (application environment) and V represents the preceding vowel (conditioning environment). For dorsal fricative assimilation, F represents the fricative (application environment) and X represents the preceding consonant or vowel (conditioning environment). Application environments are in italics. Conditioning environments are in bold. Listening targets, whether the object of interest *x* or the subsequent phone *y*, are underlined. Weber (2001a) reports Weber (2001b) experiments 1–4; Weber (2002) reports Weber (2001b) experiment 5.

analogous to English [k]–[c] in *cougar–keener*. They are typically described as standing in complementary distribution (e.g., Hall, 1989, 2022, p. 680; Wiese, 1996). The word-internal environment preceding any dorsal fricative is argued to determine whether it surfaces as [x] or [ç], conditioned by a preceding back or front vowel (e.g., ***Buch*** [**bu**ː**x**], "book, Sg.," vs. ***Büch**er* [ˈ**by**ː.**ç**ɐ] "book, Pl.," ***koch**en* [ˈkɔ. **x**n] "cook, Inf" vs. *wei**ch*** [v**a͡ɪç**] "soft, weak") or coronal consonant (e.g., *Mi**lch*** [mɪ**lç**] "milk," *Do**lch*** [dɔ**lç**] "dagger," *Mö**nch*** [mœ**nç**] "monk"), with a few morphologically derived exceptions (e.g., *Kuh* [kuː] "cow" vs. ***Kuhch**en* [ˈkuː.çən] "cow, Dim"). In dialects that lack DFA or palatal [ç] (e.g., in Switzerland), [x] may surface even after front sonorants (e.g., *echt* [ɛçt] ∼ [ɛxt] "real(ly), actual(ly)"), and [k] or [ʃ] may substitute for [ç] in loan words (e.g., ***Ch**ina* [ˈ**k**iː.na]/[ˈ**ʃ**iː.na] "China," ***Ch**emie* [**k**ɛ.ˈmiː]/[**ʃ**ɛ.ˈmiː] "chemistry;" Hall, 2022, p. 767–772; cf. invariant <ch> in *Chemnitz* [ˈkɛm.nɪts] "city of Chemnitz, Saxony"; see Hall, 2014, 2022, for more on dialectal variation). As DFA applies only to the dorsal subset of German fricatives (not /f v s z ʃ ʒ h/), feature spreading must be at a lower tier. See, for example, Figure 1C, which adds [CORONAL] to the [DORSAL]-specified Place node (cf. Hall, 1997; Iverson and Salmons, 1992). There has been a lack of clear consensus in the German phonological literature about precisely which feature triggers DFA for nearly a century (Hall, 2022, p. 1–6); oft-cited approaches employ [CORONAL] (e.g., Robinson, 2001), [+back] (e.g., Hall, 1989, 1992), and [front] (e.g., Wiese, 1996). Most recently, Hall (2022) presents a comprehensive review of the problem as reflected in the literature and weaves together historical, dialectal, and synchronic data for a coherent analysis of DFA (so-called because it is not assimilatory in all varieties). Hall characterizes DFA as a specific case of a more general velar fronting process (recall English *cougar–keener*), versions of which arise in many German dialects as well as national standard varieties.[3] For thorough reviews of phonological accounts of DFA, see Steinberg (2014, p. 27–35) and Hall (2022).

3 Hall further argues that [ç] and [x] have different phonological statuses—phonemic, quasi-phonemic, or allophonic variants—in different regional dialects of German.

3.4.2 Phoneme detection investigations of German DFA

Weber (2001a,b, 2002) conducted a series of phoneme detection experiments to investigate the distribution of German dorsal fricatives [x] and [ç] and the psychological reality of DFA as an obligatory assimilation for L1 German speakers, compared to L1 Dutch speakers (see Table 1). Taken together, in the subjective experience of Weber's participant groups, listening targets varied between experiments in terms of whether they were orthographic or phonemic and whether they indicated phonemic, subphonemic, or non-native sounds regarding the L1 phonology. Weber's DFA experiments focused listener attention directly on a specific allophonic variant of the dorsal fricative (*x*, the object of interest) rather than a subsequent phone (*y*) as the Persean approach does.

I also note that Weber (2001a) collected German data in Regensburg, Bavaria, a mainly velar-fronting dialect area with some enclaves that lack DFA and the fronted [ç] (Hall, 2022, p. 104, 427), while Weber (2002) collected data in Hannover, which lies squarely in a velar-fronting dialect area (Hall, 2022, p. 137–143). Given the potential for regional dialect variation in German DFA, dialect background, including regional exposure profiles of L1 German listeners, should be regarded as an important factor in future studies. For the present study, a time-limited opportunity for *in situ* L1 group data collection arose in Stuttgart, Baden-Württemberg. For this reason, the dialect background for the present study's L1 German group more closely resembles Lipski's (2006) Stuttgart sample than Weber's.

Based on the RT data elicited from the non-word DFA experiments, Weber (2001a) detected a small facilitation effect for L1 German listeners when presented with a front–back mismatch sequence (e.g., *[ɛx]) while listening for [x] in mono- and disyllables, whether the non-words were described as German or Dutch. In contrast, Weber (2002) found no RT effect—neither facilitation nor inhibition—when L1 Germans listened for [ç] in disyllables, regardless of whether it occurred in a licit front–front sequence (e.g., [iːç]) or illicit back–front sequence (e.g., *[aːç]). Weber (2001a,b, 2002) argues that this result, which departs from the inhibition reported for RNA by Otake et al. (1996), arises due to the interaction between the direction of the assimilation (progressive) and a reaction to novelty. In short, regressive assimilation such as RNA creates a strong expectation for which few phones may follow the first. For instance, once velar [ŋ] is perceived, internalized knowledge of obligatory RNA narrows down the possibilities significantly. In German, either [g] or [k] must follow; violating this strong expectation hinders the recovery of the underlying phonolexical representations (Darcy et al., 2007; Key, 2014) and causes processing inhibition (e.g., Otake et al., 1996; cf. visual attention experiments; Posner et al., 1978). In contrast, the earlier phone in DFA creates a much less restrictive expectation for the following phone. For example, a front vowel [iː] gives rise to the weak expectation that the next phone will not violate DFA (i.e., merely not *[x]); whether it conforms to DFA ([ç]) or is irrelevant to DFA (e.g., [p b m t d n g k ŋ]). Only an illegal sequence (e.g., *[iːx]) violates this weak expectation. If the sequence is nonetheless attested in the language (e.g., [uːç] in *Kuhchen* /kuː-çən/ → [ˈkuː.çən] "cow, Dɪᴍ"), then no RT effect arises. But if the sequence is both in violation of DFA and novel (e.g., *[bɪxt], *[blɪn.xən]), then *novel popout* may yield a small facilitation effect, as Weber (2001a) argues (Christie and Klein, 1996; Johnston and Schwarting, 1996, 1997; Weber, 2001b, p. 40–41, 53).

Lindsey's (2013) replication and expansion of Weber's (2001a, 2002) study included both L1 (Southern) German speakers and advanced L1 American English L2+ German learners. For this group, German dorsal fricatives were novel TL phones (Plag et al., 2009, p. 5–6). Contra Weber, Lindsey found processing inhibition in both groups for all DFA violations, not only front–back sequences. Additionally, Lindsey's study included both front and back listening target conditions (i.e., <ch> = [ç] in *Eiche* [ɑiçə] "oak" and *euch* [ɔyç] "you, 2P.Pʟ." vs. <ch> = [x] in *Bach* [bax] "brook" and *Bauch* [baʊx] "stomach") for all participants. These were presented as two separate blocks, completed in alternating orders, whereas Weber's participants were tested on just one listening target, depending on the location of data collection (Regensburg vs. Hannover). Weber's and Lindsey's studies offer independent evidence for the psychological reality of DFA to L1 German speakers, albeit with differing response patterns that suggest variable representation of DFA among different groups of German speakers (e.g., regional dialect differences).

This review of Weber's and Lindsey's studies highlights four important insights. Weber posits a potential explanation for processing facilitation effects with violations of progressive assimilation to highlight nuanced differences in strong or weak expectations that may arise due to the direction of assimilation.[4] Lindsey's study demonstrates that advanced L2+ learners (may) acquire sensitivity to assimilation violations in a TL. Like Otake et al.'s (1996) experiments 1, 2, and 4 with nasals, Weber's and Lindsey's experiments use the application environment—namely, the subphonemic variants of the dorsal fricative—as the listening target (Table 1). Finally, it is important to recall the implications of using the <ch> label for listening targets in these studies. For L1 German speakers, <ch> represents an L1 phoneme with multiple context-dependent surface forms, [x] and [ç]. Thus, Weber's and Lindsey's tasks required L1 German listeners to focus their attention on subphonemic variants with optimal (or highly overlearned) phonological and orthographic representations (Bordag et al., 2021; Strange, 2011). In contrast, Lindsey's L2 learners focused their attention on German <ch>, a *foreign* phone without L1 analog, which has variable surface forms in the TL and context conditions that must be acquired. Depending on learners' exposure to the target language and current IL representations, <ch> may not have optimal phonological or orthographic representations, and L1 representations (e.g., *cheese* [t͡ʃiːz]) may be activated and interfere. Thus, the label <ch> likely differs in meaning between groups in the phonological and orthographic domains, as well as the mappings between them. In this light, the same behavioral task could be a *phoneme* or *allophone* detection task for an L1 group, but a *phone* or "fuzzy" *phoneme* detection task for cross-language listeners or L2+ learners. To ensure task parity between groups, these crucial factors have been addressed in the experiments of this study by stipulating a Persean listening target—that is, an adjacent obstruent (*y*) that is uncontroversially a phoneme of both L1 and TL for all participants. The results of this task will tell us something about the representation of the neighboring fricatives.

4 Expectations about upcoming phones in the speech stream may be considered a phonological/phonotactic prediction during speech perception with implications for processing and reaction times (McMurray and Jongman, 2011; see Kaan and Grüter, 2021, for examples in other linguistic domains).

4 Beyond place assimilation: syllable phonotactic constraint of /h/

In addition to adjacency effects (i.e., place assimilation and associated transition cues), I aimed to test phoneme detection on a fundamentally different type of phonological knowledge: a prosodic/phonotactic constraint governing a segment's distribution in syllabic structure. The voiceless glottal fricative /h/ is uncontroversially phonemic in both English and German. A phonotactic well-formedness condition (Selkirk, 1984, p. 114) bans /h/ from syllable codas in both languages (e.g., *ahead* [ə.ˈhɛd], *heat* [hiːt] and German *Ahorn* [ˈa.hɔɐ̯n] "maple," *Hut* [huːt] "hat" vs. *[tiːh]). This is henceforth referred to as the *Coda-/h/ constraint (Scott, 2019a, p. 338–339). This restricted distribution is language-specific rather than universal (e.g., Turkish *tahta* [ˈtah.ta] "(wooden) board," Persian *šāh* [ʃɒːh] "king"). See Davis and Cho (2003), Jessen (1998, p. 152–153), and Scott (2019a, p. 83–100) for thorough discussion of /h/ and [h], including in English and German. The *Coda-/h/ constraint informs the design of experiment 2, which undertakes the first ever phoneme detection investigation of a non-linear prosodic structure constraint on segment distribution.

5 Detargeting the target: design motivations for adjacent listening targets

5.1 Where to direct attention in phoneme detection tasks: giving Perseus a shield

The selection of labels to focus listener attention for phoneme detection experiments must navigate three non-exclusive non-equivalence types in L1–L2+ scenarios. Each of these non-equivalence scenarios bears on GPC activation and impacts potential interpretations of different groups' responses in phoneme detection tasks. (a) The underlying phonological categories may not be equivalent. The chosen label may point to non-equivalent categories between languages. For example, the letter <N> does not represent the same phonological information to L1 Japanese speakers as it does to L1 Dutch speakers; similarly, <CH> predominantly indicates a different phonological category in English than it does in German (Lindsey, 2013; Otake et al., 1996). (b) A label may represent a phonological category of one language that does not exist in the other. For example, the letter <G> represents /g/ in German, which is a foreign phoneme to L1 Dutch speakers (Weber, 2001a,b), and German <CH> points to a phoneme without a correspondent in American English (Lindsey, 2013; Plag et al., 2009, p. 5–6). (c) A third type of phoneme detection task non-equivalence arises from which phone in a sequence is designated the listening target. In the case of assimilation, three types of listening targets are possible. Otake et al.'s (1996) experiments 1, 2, and 4 with nasals; Weber's (2001a,b, 2002) dorsal fricative experiments; and Lindsey's (2013) replication experiment all direct listener attention to the *application* environment of RNA or DFA, respectively (Figure 1)—that is, the object of interest (*x*). In contrast, Otake et al.'s experiments 5 and 6 and Weber's nasal experiment target the *conditioning* environment of RNA. This latter focus has the benefit of equivalence of the [p] and [k] listening targets as phonemic /p/ and /k/ between languages, at least in Weber's case of L1 German and L1 Dutch groups. For them, these phones are represented discretely on the segmental level in L1 orthography, whereas L1 Japanese learn these segmental representations through their secondary *rōmaji* script. In my study, a third approach (the Persean approach) is introduced, in which the listener is directed to focus on an adjacent phone that is neither an application environment nor a conditioning environment. Table 1 places the present study in methodological context by summarizing these listening target alignments.

5.2 What does Persean phoneme detection gain us?

The aim of the Persean approach to target selection for the phoneme detection task is to employ the phoneme detection paradigm while preventing the task and procedure from constituting different cognitive tasks to different groups due to how the listening target is stipulated. By targeting a subsequent adjacent phone, the Persean approach affords L2+ researchers four key advantages for intergroup task parity. First, it allows researchers to study objects of interest *x* that do not make good listening targets themselves because the languages do not spell the sound congruently, such as the difference between what orthographic <u> represents in French vs. Japanese *rōmaji,* or consistently, such as the numerous sounds that orthographic <ch> may represent in English. Instead, researchers can choose to focus listener attention on a target with a transparent one-to-one GPC common to multiple orthographies. This broadens the pool of potential stimulus designs substantially. Second, researchers can choose listening targets that are phonemically equivalent (or much closer) between language background groups. This is important for studies involving mixed L1 or previous L2 backgrounds (e.g., Canadian adults with extensive French and English exposure) and crucial for comparisons between L1–L2+ pairings in phonological acquisition research. A third advantage may benefit studies that investigate multiple levels of TL proficiency using cross-sectional or longitudinal designs. Group (or time-point) differences may include different degrees of orthographic or phonological optimization of fuzzy TL categories (Ontogenesis Model; Bordag et al., 2021), which we may associate with learners' gains in overlearning and automaticity of selective perception routines while processing TL input (Automatic Selective Perception Model; Strange, 2011). Fourth, the Persean phoneme detection variant draws listeners' explicit attention away from the object of interest *x* (i.e., detargets the target of research interest), allowing investigation of implicit phonological knowledge. For investigations of implicit phonological knowledge, it is cleanest to direct listener attention away from both the application environment and conditioning environment, both of which are crucial to the phonological phenomenon of interest. The [t]-detection conditions used in this study benefit from these advantages.

6 Research questions

The central motivation for this study is to expand on Otake et al.'s (1996) crucial insight that phoneme detection aimed at an adjacent segment for the listening target equips us to investigate questions of phonological representation in a language-neutral manner. First, I use Otake's method to investigate RNA in a new language (L1 German) to both calibrate the method with relatively clear predictions for an assimilation process and explore the potential to investigate the nuances of phonological representation using phoneme detection (experiment 1). Second, I investigate the potential of this Persean variant of the phoneme detection method to explore broader questions in L2+ phonological acquisition. I apply this design variant for the first time to investigate German DFA in both L1 and L2+ German groups, using a novel [t]-detection condition. Finally, I use the Persean [t]-detection task in the first phoneme detection investigation of a non-assimilatory phonological constraint (*Coda-[h]) with L1 and L2+ German groups in parallel (Experiment 2).

6.1 Research questions for RNA

Experiment 1 investigates the first two research questions.

RQ1: Does violation of RNA in an obligatory context in German yield RT patterns of consistent inhibition, variable inhibition, or no inhibition?

It is hypothesized that RNA violation will inhibit RT (i.e., slower reactions) in L1 German speakers. Experiment 1 undertakes replication of the findings for RNA with L1 German speakers by Weber (2001a,b) with [k]- and [p]-detection. A secondary motivation of experiment 1 is theoretical. The [p]- and [k]-detection blocks of experiment 1 investigate the impact of feature specification on mismatched Place violations of RNA on processing.

RQ2: What manner of feature specifications do RT data from violation of German RNA support?

An unlikely null effect would suggest tolerance of RNA violation, undermining its obligatoriness. A single distinction between match vs. mismatch would support the obligatoriness of RNA but not

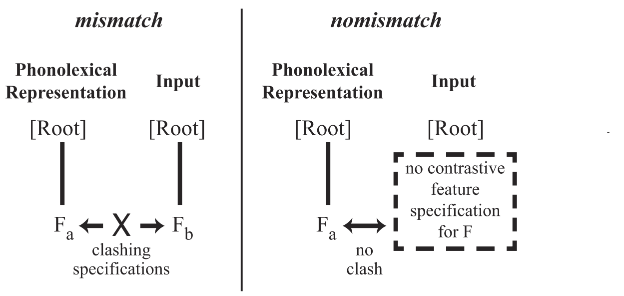


FIGURE 2
Schematic representation of a *mismatch* **(left)** between features specified in both the listener's phonolexical representation and a phone encountered in the input with an indication of a different specification of that feature vs. a *nomismatch* **(right)** between a phonolexical representation with specified feature and a phone encountered in the input that does not clearly indicate another specification of that feature.

assumptions of a psychologically real distinction between the underapplication of RNA in cases of underspecification (e.g., ?[np nk]) and true place-clash sequences, such as *[mk] or *[ŋp] that involve fully specified Place features.

The third possibility of a ternary distinction between place-match [mp ŋk], underspecified mismatch ?[np nk], and place-specified mismatch *[mk] or *[ŋp], even with non-word stimuli, might lend support to underspecification theories. One example is the ternary logic of the Featurally Underspecified Lexicon (FUL; Lahiri and Reetz, 2010, p. 50–51). This model posits a processing distinction between input that clashes with a phonolexical representation, for example, an explicit *mismatch* of specified Place features that cannot be interpreted as the output of an overlearned rule for which an underlying form is recoverable (Darcy et al., 2007; Strange, 2011) vs. input that merely fails to match a phonolexical representation, for example, *nomismatch* of Place features, due to underapplication of a rule to specify an underlyingly underspecified Place. These scenarios are schematized in Figure 2.

6.2 Research questions for DFA

RQ3: Does the [t]-detection task show processing facilitation or processing inhibition in response to violation of DFA in an obligatory context in German?

Previous phoneme detection studies report contradictory RT results for violation of DFA (Lindsey, 2013; Weber, 2001a,b, 2002), which limits the basis for predictions. Weber found processing facilitation (i.e., faster RT) for violation of DFA with L1 German listeners but only under certain conditions or with certain regional populations, whereas Lindsey found processing inhibition (i.e., slower RT) in all conditions for both L1 German and L2+ learner groups. This study aims at replication with similar groups, but without directing the attention of either group to attend to <CH> directly, to ensure task parity between language background groups and add to our empirical knowledge of phonological processing of DFA violations. In the novel [t]-detection condition of experiment 1 and in the similar DFA condition of experiment 2, the listening target is (a) familiar to both language groups as an L1 phoneme with similar acoustic realizations, and (b) irrelevant to the progressive assimilation that precedes it. The novel [t]-detection approach in experiments 1 and 2 pursues independent replication of findings with DFA in L1 German speakers or L2+ German learners, but with listener focus on a different target.

RQ4a: Do L1 American English L2+ German learners exhibit sensitivity to violations of German DFA?

RQ4b: Do L1 German speakers exhibit sensitivity to violations of German DFA?

RQ4c: Is the adjacent [t]-detection task able to detect sensitivity to violations of progressive assimilation that precede the listening target?

Experiment 2 undertakes replication of Lindsey (2013) and extends the adjacent target technique in two ways. First, it includes an L1 English L2+ German group as a further replication of Lindsey (2013), to investigate the fourth group of research questions. A demonstrated sensitivity to DFA violation in either experiment could independently confirm either Weber's or Lindsey's findings, at least with the population sampled here. It would also demonstrate that the instrument is sufficiently sensitive to detect RT effects for violation of progressive assimilation. Failure to find any significant RT effects for DFA violation could indicate that the instrument is susceptible to Type II error for this type of assimilation. Experiment 2 also adds a novel non-assimilation condition with illicit /h/ in syllable codas.

6.3 Research questions for ban of [h] in syllable codas

The final research questions expand on previous research by undertaking the first phoneme detection investigation of the prosodic constraint banning /h/ in codas. This ban is exceptionless in both languages, despite cross-language perceptual assimilation patterns for dorsal fricatives (see Section 4; Scott and Darcy, 2023), so violations should yield strong RT inhibition (cf. RNA).

RQ5a: Do L1 American English L2+ German learners exhibit sensitivity to violations of *Coda-/h/ ?

RQ5b: Do L1 German speakers exhibit sensitivity to violations of *Coda-/h/ ?

RQ5c: Is the adjacent [t]-detection task suitable for investigating types of phonotactic knowledge other than assimilation processes?

Experiment 2 also includes a novel investigation of the phonotactic/prosodic constraint *Coda-/h/ with both language background groups, to explore the task's utility with other types of phonological constraints. Demonstrated sensitivity to violation of *Coda-/h/ by either population would demonstrate that the instrument is sufficiently sensitive to detect RT effects for violation of prosodic/phonotactic well-formedness constraints that create strong expectations for upcoming phones. Failure to find any significant RT effects for *Coda-/h/ violation could indicate that the instrument is susceptible to Type II error for this type of phonotactic/prosodic constraint.

7 Experiment 1: detection of conditioning environment or Persean target in L1

Experiment 1 focuses on nasals and fricatives in L1 German. The phoneme detection task in this experiment includes two types of listening targets, represented by the following obstruent: (a) focus on the conditioning environment of RNA ([p] or [k]) and (b) focus on an adjacent phone unrelated to DFA ([t]). Two experiment blocks, [p]- and [k]-detection tasks, focus on obligatory RNA in German, with the aim to replicate previous findings of RT inhibition when RNA is violated in monosyllable codas and listener focus is directed to the obstruent that conditions assimilation of the preceding nasal. Another block employs Persean [t]-detection to investigate German DFA. It diverges from previous phoneme detection studies of DFA by directing listener attention to a following obstruent [t], which plays no role in the rule. It also has the advantage of avoiding the use of orthographic <CH> as the listening target for multiple allophonic variants (cf. Lindsey, 2013; Weber, 2001a,b, 2002). The aim is to test whether violation of DFA in an obligatory monosyllabic context yields RT inhibition (cf. Lindsey) or facilitation (cf. Weber) for L1 German speakers. On the assumption that RT effects for place mismatches carry over to influence listening targets that immediately follow the phones involved in assimilation, experiment 1 should yield shifts in RT in response to violation of RNA (RQ1 and RQ2) and DFA (RQ3), if these are psychologically real to L1 German speakers.[5]

7.1 Method

7.1.1 Participants

Seventeen L1 German speakers (13 female; ages 18–35; $M = 25.2$, $SD = 4.764$) received €5 for completing experiment 1. Eleven completed the task in Stuttgart, Baden-Württemberg, Germany, seven of whom had previously completed experiment 2 because experiment 1 was offered as an optional additional experiment in a session. Six Germans attending universities in the Midwestern United States were recruited for supplementary data collection. Additional participant details are provided in Supplementary material B.

7.1.2 Stimuli

A tabular summary of experiment 1 trial types by condition is provided in Supplementary Table C1. Non-word stimuli ($N = 304$) were prepared for three assimilation types (Nasal-[p]-Detection vs. Nasal-[k]-Detection vs. Fricative-[t]-Detection). The nasal condition was balanced for three conditions of match type ($n = 18$; 3 each for Match [mp], [ŋk] vs. Underspecified Mismatch ?[np], ?[nk] vs. Specified Mismatch *[mk], *[ŋp]). The Fricative condition was balanced for four conditions of Match type ($n = 20$; 5 each for Front Match [ɛç] vs. Back Match [ax] vs. Back–Front Mismatch *[aç] vs. Front–Back Mismatch *[ɛx]). All target stimuli included the listening target as the final obstruent of a monosyllable coda (e.g., [p] in [zɪmp], [k] in [zɔŋk], [t] in [glɛçt], [glaxt]). The 38 critical trials were balanced by 114 distractors with the listening target in non-final positions (27 with [p], 27 with [k], 60 with [t]) and 152 fillers without the listening target (36 in [p]- and [k]-detection blocks, 80 in [t]-detection) so that [p]- and [k]-detection blocks totaled 72 trials and [t]-detection totaled 160, yielding a 1:3 ratio of critical trials to distractors and 1:1 ratio of trials with the listening target (critical trials + distractors) to fillers without it (Keating and Jegerski, 2015; p. 16). Supplementary material F provides a complete list of stimuli.

At least three tokens of each item were digitally recorded in a sound-attenuated booth (sampling rate 44,100 Hz) by a phonetically trained L1 German female talker from Saxony who spoke and taught Standard German professionally in the United States. The researcher selected one token of each item for recording quality. Six training non-words were also recorded: *Tiesel*, *gamisch*, *frettig*, *Skirm*, *Prasen*, and *Schloft*. Training trials for the [t]-detection block and all distractors and fillers were drawn from experiment 2 stimuli. Files were manually cut and normalized for volume by a Praat (Boersma and Weenink, 2014; Version 5.4) script; the task was presented with OpenSesame (Mathôt et al., 2012; Version 2.9).

Listening targets in the generalized phoneme monitoring procedure (Frauenfelder and Seguí, 1989) may occur in different parts of the stimuli (distractors vs. critical trials); furthermore, each individual token of a listening target varies in duration. To compensate for varying duration, it was more accurate to combine RT measurements collected by the software (see Section 7.2.2) with the duration from listening target onset to the end of the audio file for each trial to derive an *augmented RT* measurement that reflects participants' processing time. The calculation of augmented RT for each trial is depicted in Figure 3. Segment boundaries were marked in Praat (Boersma and Weenink, 2014; Version 6.0.19), and the onset and duration of each listening target and the phone preceding it were extracted. Supplementary Tables C2, C3 describe these durations by condition in aggregate; see Supplementary material F for the extracted durations of each stimulus and Scott (2019a, p. 299–302) for additional analyses of the stimuli. Because the listening target for critical condition trials always appears at the end of the stimulus, the sum of the listening target duration and the automatically recorded RT yielded the *augmented RT*, which serves as the dependent variable for analysis.

7.1.3 Procedure

Stuttgart data were collected in a quiet computer lab of six identically configured computers running Windows 7 (Professional 9 Service Pack 1, 64-bit) with a 3.2 GHz processor, 4 GB RAM, and 1680 × 1050-pixel screen resolution. Mobile data collection in the United States used a single Dell XPS 12 two-in-one laptop (laptop and tablet modes) running Windows 8 or Windows 10 Pro (64-bit) with an Intel i7-4510U 2.6 GHz processor, 8 GB RAM, and 1080 × 1920-pixel screen resolution. Stimuli were presented through high-quality circumaural headphones.

Participants completed a language background questionnaire (see the Open Science Framework resources). The researcher briefly explained the phoneme detection task in German, and participants

5 This section presents a reanalysis of a data set originally collected in 2015 and reported by Scott (2019a, Chapter 5).

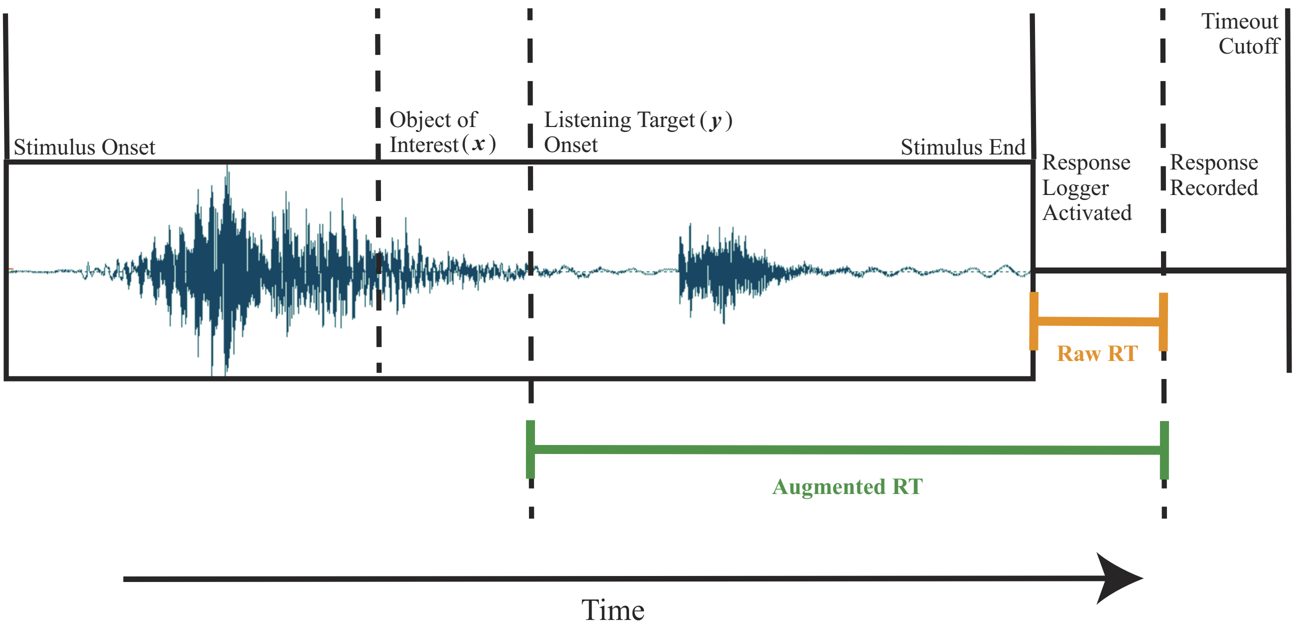


FIGURE 3
Depiction of the calculation of *augmented reaction time (RT)* from the sum of *raw RT* logged by the experiment and the duration from listening target onset to the end of audio playback. For nasal [p]- and [k]-detection conditions, the listening target was not the object of interest (*x*), but the adjacent obstruent was the conditioning environment of regressive nasal assimilation (RNA). For the fricative [t]-detection condition, the listening target was a Persean listening target: neither the object of interest (*x*) nor the conditioning environment, but an adjacent and unrelated obstruent.

TABLE 2 Signal detection rates for experiment 1 after participant exclusion (*N* = 12).

Condition	Signal detection			
	Hit	False alarm	Correct rejection	Miss
<P>/<K>				
Match	0.972	–	–	0.028
Mismatch (underspecified)[a]	0.958	–	–	0.042
Mismatch (specified)	0.958	–	–	0.042
<T>				
Match	0.942	–	–	0.058
Mismatch	0.892	–	–	0.117
Distractors[b]	0.908	–	–	0.092
Fillers	–	0.016	0.984	–

Dashes indicate fields for which no rate is possible with the "go"/"no-go" format.
[a]Nasal Specified and Underspecified Mismatch conditions exhibited similar accuracy in aggregate, with individual variation in different subconditions.
[b]Distractor Miss rate may include responses entered before the experiment response logger initiated.

read instructions on the screen that explained that they would hear invented words in three blocks. In each block, they were to listen for "T," "K," or "P" somewhere in the word and indicate when they heard the listening target by pressing the space bar as quickly as possible. If the target was absent, participants waited for the next trial without responding. Text examples of the listening targets present in various positions (or absent) were displayed with explanations of appropriate responses, and then participants completed a six-trial training phase in blocks of two for [t]- (***T****iesel*, *gamisch*), [k]- (*frettig*, ***S****kirm*), and [p]-detection (***P****rasen*, *Schloft*). The practice trials alternated with training instructions explaining the block-specific listening targets and the potential for the target to occur anywhere in the word. The order of the three blocks was random, and the trials were randomized within blocks. Every 16 trials, participants had the option to pause and resume when ready. Each trial began with a fixation point on the screen, followed by audio playback. The experiment recorded responses and RTs. The OpenSesame response logger was located immediately after playback so that a recorded RT of 0 ms corresponded to the end of the stimulus just after the release of the syllable-final listening target in critical trials (see Figure 3). Experiment 1 lasted ∼25 min.

7.2 Results

7.2.1 Exclusion criterion

This study employed a "go"/"no-go" phoneme detection response format (i.e., was the phoneme detected? Affirmative responses only). In terms of signal detection theory as used in perception research, this format only records hits (e.g., accurate detection of the target) and false alarms (e.g., spurious indication of target presence when it is absent). Correct rejections (e.g., correct indication of target absence) and misses (i.e., failure to indicate target presence when it is present) are not recorded. In this format, some correct rejections could additionally result from non-response bias, while misses could result from lack of sensitivity, non-response

TABLE 3 Descriptive statistics and tests of normality for experiment 1 by condition.

Condition	n	μ (ms)	σ	Skewness		Kurtosis		Shapiro–Wilk	
					SE		*SE*	*W*	*p*
Nasal	207	628	122.4	0.315	0.169	0.250	0.337	0.989	0.132
Match	70	564	115.3	0.272	0.287	0.121	0.566	0.987	0.681
Underspecified	69	632	96.1	0.247	0.289	−0.177	0.570	0.989	0.807
Specified	68	689	122.1	0.529	0.291	−0.181	0.574	0.968	0.079
Fricative	220	559	105.6	0.648	0.164	0.187	0.327	0.964	<0.001
Match	113	565	99.0	0.464	0.227	−0.022	0.451	0.980	0.095
Mismatch	107	554	112.3	0.820	0.234	0.393	0.463	0.940	<0.001

RT values rounded to the nearest millisecond. Observed positive skewness is not an artifact of data trimming, which excluded just one non-filler fast response (<100 ms), a *[ŋp] Nasal Mismatch.

bias, or both.[6] A minimum threshold of five hit responses was set for each Match Type condition according to Assimilation Type to ensure that no single response could influence the mean of any condition too much. This excluded five participants with fewer than five hit responses in any of the five conditions, Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch, Fricative Match, or Fricative Mismatch, while retaining 428 responses from the remaining 12 participants (i.e., 183 Match and 245 Mismatch trials; 208 Nasal and 220 Fricative trials). Table 2 displays the remaining 12 participants' signal detection rates.

7.2.2 Data trimming and preparation

Data trimming removed the upper and lower ends as follows. The response logger timed out at 700 ms, so trials with raw RT of 700 ms were excluded. Extremely short RTs might indicate responses prior to presentation of the listening target in playback, so one non-filler *[ŋp] trial with raw RT below 100 ms was excluded. The remaining non-filler trials were included for analysis of their augmented RT. Fillers were excluded.

Scott (2019a) examined these data regarding assumptions of normality, ultimately abstaining from log-transformation. It has been common practice to log-transform data to satisfy assumptions of normality that raw behavioral RT data often violate. However, log-transformation of data has recently come under much criticism (e.g., O'Hara and Kotze, 2010). For example, as data may not approximate a log-normal distribution, there is no guarantee of reduced skewness and, indeed, some risk of increased skewness; furthermore, log-transformation can often *increase* variability (Feng et al., 2014, p. 106). Some type of transformation is always available to increase or reduce the variability of original data, making their value questionable (p. 107). Finally, hypothesis testing on the log-transformed data may not address the hypothesis for the original data (p. 108). For these reasons, I apply no data transformation. Instead, descriptive statistics transparently include skewness, kurtosis, and Shapiro–Wilk tests of normality.

6 The Gardner Lab at Stanford University offers a concise introduction to the logic and terminology of signal detection theory for human perception research online at: https://gru.stanford.edu/doku.php/tutorials/sdt. For more detailed introductions in a psychology framework, see Heeger's (1997) handout or online summary at: https://www.cns.nyu.edu/ david/handouts/sdt/sdt.html.

7.2.3 Combined analysis

Table 3 displays central tendency statistics and normality tests for experiment 1 by condition. Equivalent mean RT of Match conditions across Assimilation Types (Nasal: $M = 564$ ms; Fricative: $M = 565$ ms) establishes the baseline performance for experiment 1 for both conditions against which to compare other conditions. Both Match conditions satisfy the assumption of normality, and an independent samples t-test detects no significant difference between the two Match conditions, $t_{(181)} = -0.067$, $p = 0.946$. Mean RT between Nasal conditions shows an ordinal pattern: Underspecified Mismatch ($M = 632$ ms) RT is inhibited with respect to Match, and Specified Mismatch is slower still ($M = 689$ ms). This pattern of processing inhibition does not hold for the Dorsal Fricative conditions, where Mismatch appears to be slightly faster ($M = 554$) than Match ($M = 565$).

Scott (2019a) coded both types of Nasal Mismatch together for a 2 × 2 model (Match condition vs. Assimilation type) but noted a regular difference in RT between the Specified Mismatch and Underspecified Mismatch conditions (p. 319). To investigate this, the present analysis distinguishes all three levels. This yields three levels of condition for nasals (Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch) and two levels for fricatives (Fricative Match, Fricative Mismatch), which precludes a meaningful comparison within a single model. Separate analyses for nasals and fricatives follow.

7.2.4 Nasals analysis

See Table 3 for descriptive statistics and normality tests for nasal conditions. A linear mixed-effects model was run on the nasal RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (3 levels: Nasal Match, Nasal Underspecified Mismatch, Nasal Specified Mismatch) and Target (2 levels: K, P) were declared as fixed effects. To construct a maximal initial model (Barr et al., 2013), participants and items were declared as random effects grouping factors with Condition, Target, and Condition*Target as random effects. The data set could not support the maximal random effects structure, so the model was incrementally simplified to what the data can support (Matuschek et al., 2017). The final model included Condition, Target, and Condition*Target as fixed effects with random intercept. Table 4 reports the estimated marginal means and parameter estimates. Figure 4 depicts the distribution of

TABLE 4 Estimated marginal means (ms, hits only) of experiment 1 (nasals), *SE*, and 95% confidence interval (top), with parameter estimate, variability, *SE*, *df*, *t*-value, and *p*-value (bottom).

Condition	Mean RT	*SE*	95% confidence interval	
			Lower bound	Upper bound
Match condition				
Match				
K-detection	551	23.9	504	598
P-detection	578	23.9	531	625
Mismatch				
Underspecified				
K-detection	614	23.9	567	661
P-detection	655	24.1	608	702
Specified				
K-detection	634	23.8	587	681
P-detection	752	24.5	704	800
Target condition				
K-detection				
Match	551	23.9	504	598
Mismatch				
Underspecified	614	23.9	567	661
Specified	634	23.8	587	681
P-detection				
Match	578	23.9	531	625
Mismatch				
Underspecified	655	24.1	608	702
Specified	752	24.5	704	800

Fixed effects	Estimate	*SE*	*df*	*t*	*p*
Intercept	631	16.04	11.97	39.33	<0.001
[Condition (1) = Underspecified Mismatch]	−66	11.28	12.16	−5.85	<0.001
[Condition (2) = Specified Mismatch]	4	11	12.28	0.33	0.748
[Target (1) = P]	−31	8	12.28	−3.87	0.002
[Condition = Underspecified Mismatch]*[Target = P]	17	11	12.17	1.55	0.148
[Condition = Specified Mismatch]*[Target = P]	10	11	12.29	0.93	0.372
Parameter	Variance	σ			
Residual	8,958	95			
Participant	2,318	48			
Item	369	19			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Target reference level is K. Significance calculation includes Holm-Bonferroni correction for multiple comparisons.

responses by condition, showing an apparent ordinal trend overall toward slower RT from Match to Underspecified Mismatch to Specified Mismatch (Figure 4A), although this trend appears to be less robust in [k]-detection conditions (Figure 4B, left).

With the type III tests of mixed effects, the *F*-tests show a main effect of match condition, $F_{(2,12.27)} = 21.474$, $p < 0.001$, driven by the observed difference in mean RT between match and underspecified mismatch conditions. Although the mean RT for the specified mismatch is relatively slower, its contribution to this model does not achieve significance. There is also a main effect of the listening target, $F_{(1,12.27)} = 14.940$, $p = 0.002$, confirming that [p]-detection trials have consistently slower RT than [k]-detection. The interaction of these factors is marginal, not significant, $F_{(2,12.28)} = 3.089$, $p = 0.082$. An independent samples *t*-test comparing underspecified and specified mismatch conditions reveals a significant difference, $t_{(135)} = 3.022$, $p = 0.003$, a medium effect, $d = 0.516$, $SE = 0.177$. Rerunning the model with only underspecified and specified mismatch trials shows the same main effects, but again there is not a significant interaction effect. In summary, experiment 1 replicates a robust inhibition effect for violation of RNA (cf. Otake et al., 1996; Weber, 2001a,b) and shows a significant difference between [k]-detection and the relatively slower [p]-detection conditions. There is a significant difference between underspecified mismatch ?[nk np] trials and slower specified *[ŋp mk] trials, and the differences of estimated marginal mean RTs between [p]-detection and [k]-detection are relatively wider for specified than for underspecified trials (118 vs. 41). Nonetheless, this interaction is marginal, so it does not conclusively indicate a greater inhibition effect for *[ŋp] than *[mk].

7.2.5 Fricatives analysis

Descriptive statistics and normality tests for fricative conditions are displayed in Table 5. A linear mixed-effects model was run on the fricative RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Match vs. Mismatch) and Vowel Context ([a] vs. [ɛ]) were declared as fixed effects. Participants and items were declared random effects grouping factors. As before, the maximal model was incrementally simplified, resulting in a final model that includes Condition, Context, and Condition*Context as fixed effects with random intercept. Table 6 reports the estimated marginal means and parameter estimates. Figure 5 depicts the distribution of responses by condition: licit [ɛç], illicit *[ɛx], licit [ax], illicit *[aç].

With the type III tests of mixed effects, the *F*-tests show no main effect of condition, $F_{(1,15.69)} = 0.682$, $p = 0.421$, nor is there a main effect of vowel context, $F_{(1,15.66)} = 1.531$, $p = 0.234$. The interaction is also not significant, $F_{(1,15.74)} = 1.728$, $p = 0.207$. In summary, experiment 1 replicates neither a facilitation effect (cf. Weber, 2001a,b) nor an inhibition effect (cf. Lindsey, 2013) for L1 German speakers. Subsequent analyses of these fricative assimilation data by Scott (2019a, p. 322–326) revealed a high degree of variation between participants, including individuals with strong inhibition or strong facilitation but mostly neither.

Recall from Section 3.4.2 Weber's (2001a; 2001b, p. 40–41, 53) claim that facilitation may be a novel popout reaction to DFA-violating sequences that are truly novel in the language, such as *[i:x], and not those that are merely rare (e.g., [u:ç] in *Kuhchen*

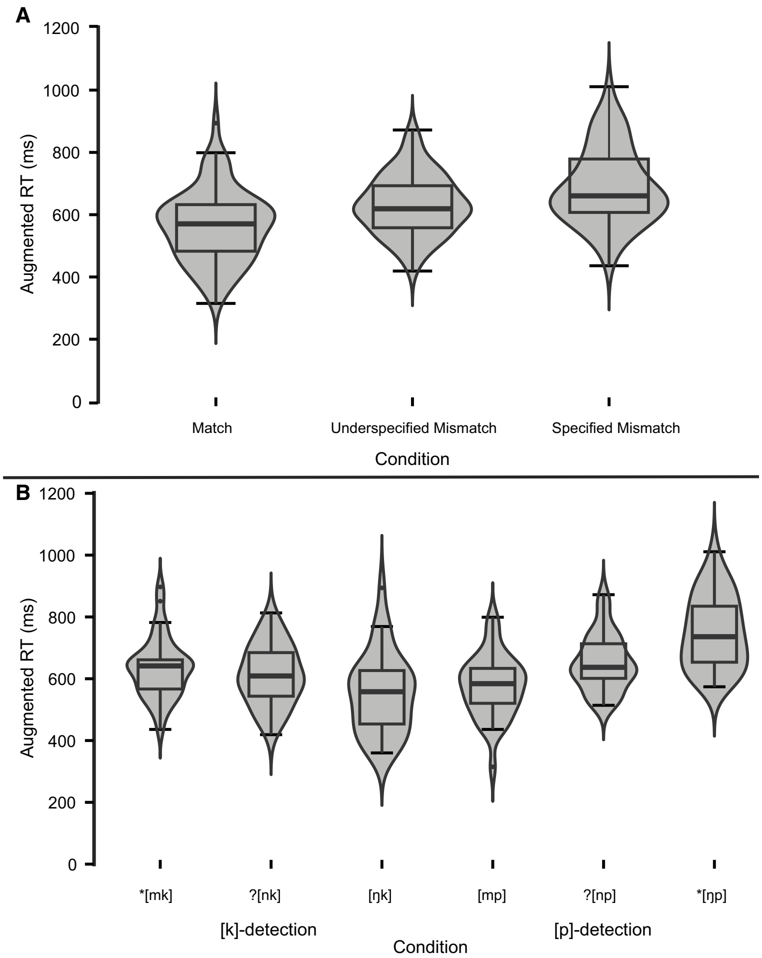


FIGURE 4
Experiment 1 (nasals): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data **(A)** by condition (match, underspecified mismatch, specified mismatch) and **(B)** additionally by target (K vs. P).

TABLE 5 Descriptive statistics and tests of normality for experiment 1 fricatives by condition.

				Skewness		Kurtosis		Shapiro–Wilk	
Condition	*n*	μ(ms)	σ		*SE*		*SE*	*W*	*p*
Match	113	565	99.0	0.464	0.227	−0.022	0.451	0.980	0.095
[ax] + t	56	581	105.6	0.343	0.319	−0.173	0.628	0.982	0.556
[ɛç] + t	57	549	90.2	0.508	0.316	0.154	0.623	0.973	0.226
Mismatch	107	554	112.3	0.820	0.234	0.393	0.463	0.940	<0.001
*[aç] + t	54	552	117.0	1.117	0.325	1.040	0.639	0.897	<0.001
*[ɛx] + t	53	556	108.4	0.469	0.327	−0.298	0.644	0.963	0.102

Reaction time values rounded to the nearest millisecond. Observed positive skewness is not an artifact of data trimming, which excluded no non-filler fast responses (<100 ms) in [t]-detection conditions.

/kuː-çən/→ [ˈkuː.çən] "cow, Dɪᴍ."). To explore this claim, an independent samples *t*-test between Back–Front Mismatch (*[aç]; $M = 552$, $SD = 117$) and Front–Back Mismatch (*[ɛx]; $M = 556$, $SD = 108$) conditions was run. No significant difference was revealed, $t_{(105)} = 0.163$, $p = 0.871$. In contrast, a comparison of the corresponding Match conditions showed a marginal difference, $t_{(111)} = 1.721$, $p = 0.088$, suggesting a trend toward a slower baseline RT for matching back [ax] sequences ($M = 581$, $SD = 106$) than matching front [ɛç] ($M = 549$, $SD = 90$). Thus, despite the slightly faster mean (by 11 ms) observed for the Mismatch condition, the statistical model of experiment 1 results does not conclusively support Weber's novel popout argument.

TABLE 6 Estimated marginal means (ms, hits only) of experiment 1 (fricatives), *SE*, and 95% confidence interval (top), with parameter estimate, variability, *SE*, *df*, *t*-value, and *p*-value (bottom).

Condition	Mean RT	*SE*	95% confidence interval	
			Lower bound	Upper bound
Match condition				
Match				
[ax] + t	584	22.0	541	627
[εç] + t	552	21.9	509	595
Mismatch				
*[aç] + t	557	22.1	514	600
*[εx] + t	558	22.2	515	601
Vowel Context				
Back [a]				
[ax] + t	584	22.0	541	627
*[aç] + t	557	22.1	514	600
Front [ε]				
[εç] + t	552	21.9	509	595
*[εx] + t	558	22.2	515	601

Fixed effects	Estimate	*SE*	*df*	*t*	*p*
Intercept	563	19.04	11.25	29.55	<0.001
[Condition (1) = Mismatch]	5	6.39	15.69	0.83	0.421
[Context (1) = ε]	8	6.38	15.66	1.24	0.234
[Condition = Mismatch]* [Context = ε]	8	6.39	15.74	1.32	0.207
Parameter	Variance	σ			
Residual	7,490	87			
Participant	3,852	62			
Item	132	11			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Vowel Context reference level is back [a]. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

7.2.6 Discussion of experiment 1

The combined analysis (Section 7.2.3) establishes baseline performance on the task for Match conditions across assimilation types. The nasals analysis (Section 7.2.4) addresses the first two research questions. As expected, violation of RNA results in consistent, pronounced RT inhibition (RQ1). Regarding the second research question, the results show that the type of mismatch, whether underspecified due to underapplication of RNA ([n]) or specified with a clash of non-coronal place features (*[mk ŋp]), makes a significant difference in RT by degrees. This distinction is primarily driven by strong processing inhibition for illicit *[ŋp] sequences, which never occur in German, whereas [mk] sequences, rare yet possible (e.g., *Imker*), manifest less delay. This may be due to the uncontroversial phonemic status of /m/ in German, which does not require derivation via RNA. In contrast, [ŋ] derives exclusively from RNA in pre-velar context. This demonstrates processing differences between Place-assimilated, Place-mismatched underspecified, and Place-mismatched specified nasals, constituting psycholinguistic evidence for the incremental ungrammaticality of phonotactic violations (RQ2). This ordinal differentiation suggests that there may be a continuum of processing inhibition for assimilation of this type according to the intensity of the RT effect:

> * (phonotactic constraint) > ? (phonological underapplication) > lexical rarity[7]

The possibility remains that the phonemic status of /m/ may also play a role, as argued by Otake et al. (1996). Note the difference in Table 4 between the quick baseline RT for homorganic labial [mp] and the much slower RT for heterorganic *[mk], which violates the Coda condition (Itô, 1989; p. 224). Because [mk] sequences occur rarely in German, labeling them questionable ?[mk] may be more appropriate. Their mean RT here also aligns better with underspecified ?[nk np]. Inhibition with ?[nk np] seems to reflect that they do not occur in German. However, following Darcy et al. (2007), unassimilated sequences ought to be recoverable as cases of the underspecified nasal due to highly overlearned (automatic) perception routines (Strange, 2011). This may play a role in the reduced degree of processing inhibition observed. In contrast, Specified Mismatch *[ŋp], which lacks any phonological or lexical motivation, exhibits more severe processing inhibition than all other conditions. This partly aligns with the ternary logic of the FUL model (Lahiri and Reetz, 2010), which describes an algorithm for comparing features extracted from the acoustic signal with features encoded in the mental lexicon to recover the speaker's intended lexical meaning. By the FUL algorithm, we might expect a *mismatch* for *[ŋp], which lacks motivation via RNA or lexical precedent, to inhibit processing more than a *nomismatch* for unassimilated /Nk/ or /Np/ surfacing with coronal [n] (see Figure 2). However, the present nasal model's marginal factor interaction does not conclusively support a difference between [k]-detection mismatches and relatively slower [p]-detection mismatches, so experiment 2 cannot address how incorrect application of RNA might interact differently with the labial feature of phonemic /m/ vs. the RNA-acquired velar feature of [ŋ]. Alternatively, this pattern may partly owe to relatively weaker invariant cues for [p] Ohala (1996, p. 1720; Weber, 2001a, p. 111). Future research with a larger data set designed to investigate match, underspecified mismatch, and specified mismatch with labials and velars, in words and non-words, is needed to address this conclusively.

Finally, the fricative analysis (Section 7.2.5) yielded no significant results for violation of DFA. It is possible that subtle RT effects may be cloaked here by the slower baseline for trials with [ax] (Table 5). Alternatively, the [t]-detection task may not be sensitive enough to detect small facilitation effects in the preceding context, or greater individual variation in sensitivity to DFA violations may

7 A fourth category, *lexical unprecedentedness*, could logically arise as the systematic result of a phonotactic constraint or as the accidental result of a lexical gap, where novel potential words may yet comply with well-formedness principles. Whether a complete lexical gap would affect processing similarly to a phonotactic ban is an open empirical question.

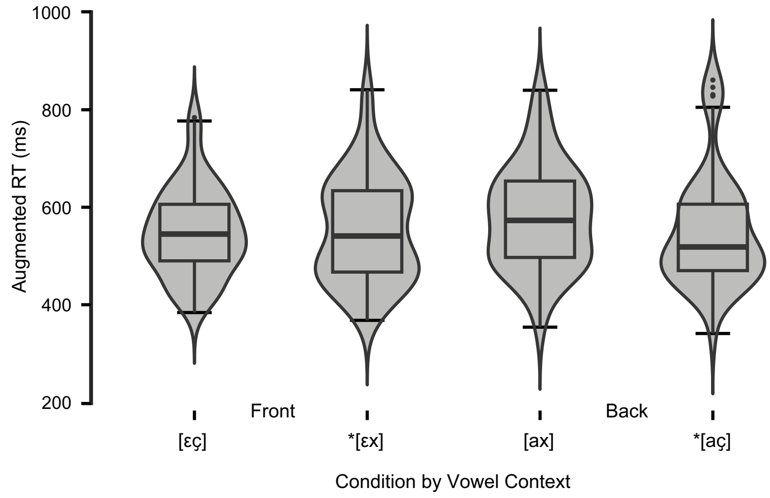


FIGURE 5
Experiment 1 (fricatives): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data by condition (match vs. mismatch).

necessitate more statistical power (Type II error). Finally, DFA may not exert much influence on perception for this L1 German sample. Thus, experiment 2 takes up the third research question again.

8 Experiment 2: detection of Persean targets in L1 and L2+

Experiment 1 validated the phoneme detection task generally by replication of an RT inhibition effect in response to violation of RNA with an L1 German group. However, the regressive (left-to-right) directionality of the RNA condition only shifts the listening target one step away from the object of interest *x* onto the following [p] and [k], both of which still play a role in the RNA process as the triggers of place assimilation. To take the next step away from *x* itself and stipulate a true Persean listening target *y*, the [t]-detection task investigated nearby DFA, a place assimilation in which the [t] plays no role. This condition was inconclusive in experiment 1 with the L1 German group, so in experiment 2, this is attempted again and with an additional L2+ learner group. Experiment 2 also novelly tests this Persean listening target on the prosodic/phonotactic constraint governing /h/ in codas.[8]

Experiment 1 also established a baseline performance pattern between phoneme detection tasks that use an uncontroversial phonemic listening target, whether that target is part of the conditioning environment (i.e., [k] and [p] in RNA context; Otake et al., 1996, experiments 5 and 6; Weber, 2001a,b, experiment 4) or it is an adjacent Persean target, unrelated to the assimilation of interest in DFA context. The aim of experiment 2 is to investigate sensitivity to phonotactic violations in both L1 speakers and adult L2+ learners, without drawing participant attention to segments that are directly involved in the phonological principles of interest, because such segments are likely attached to differing mental representations between groups (e.g., application or conditioning environments). To achieve this, experiment 2 uses [t], which is uncontroversially phonemic /t/ in both German and English. Furthermore, the [t]-detection task uses only released stops, avoiding language-specific alternatives such as optionally unreleased final stops in English. Thus, the task employs phonetic realizations that unambiguously instantiate the phoneme /t/ in both languages.

Experiment 2 draws additional motivation from research on the perception of German dorsal fricatives by L1 American English speakers. Scott (2019a) and Scott and Darcy (2023) report that prosodic and phonotactic contexts modulate perceptual assimilation (Best, 1995; Best and Tyler, 2007) of dorsal fricatives [x] and [ç] and confusability with [k], [ʃ], and [h]. This includes perceptual assimilation to [h] in coda positions, despite the phonotactic/prosodic constraint *Coda-/h/ (§4). For L1 English speakers' perception of German, such a perceptual assimilation mapping in cross-language and IL perception violates syllable well-formedness principles of both the L1 and the TL. This indicates attention to phonetic detail over phonological patterning during early exposure. To complement the explicit attention to phonetic detail inherent to perceptual assimilation or phoneme detection focused on subphonemic variants (Lindsey, 2013; Weber, 2001a,b, 2002), experiment 2 investigates implicit processing reactions while listener attention is directed elsewhere as a proxy for *automatic* or *optimal* phonological knowledge (Bordag et al., 2021; Hui and Jia, 2024; Strange, 2011).

Following the [t]-detection block of experiment 1, experiment 2 aims to replicate the findings on German DFA reported by Weber (2001a,b, 2002) or Lindsey (2013). Acknowledging the negative DFA result of experiment 1, it also remains to be shown whether this task design is sensitive enough to detect RT effects on processing caused by violations of a preceding progressive assimilation. I then undertake the first investigation of the prosodic/phonotactic constraint *Coda-/h/. Experiment 2 should yield shifts in RT in response to violation of DFA and

8 This section presents a reanalysis of a data set originally collected in 2015 and reported by Scott (2019a, Chapter 6; Scott, 2019b).

TABLE 7 Experiment 2 response tallies by group after exclusion criteria.

Group	Dorsal fricative		Glottal fricative		Totals
	Licit	Illicit	Licit	Illicit	
	[axt εçt]	[açt εxt]	$_{\sigma}$[hVCt]	[...Vht]$_{\sigma}$	
L1 ($n = 9$)	135	123	60	56	374
L2+ ($n = 14$)	194	195	100	98	587
Totals	329	318	160	154	$N = 961$

Fillers are excluded from the table and analysis. All critical trials were monosyllables with the listening target [t] in final position. In licit glottal fricative trials, the penultimate consonant was always licit in the position in both English and German.

*Coda-/h/ in each participant group for whom these phonological patterns are mentally represented.

8.1 Method

8.1.1 Participants

Two participant groups were recruited for experiment 2. Data collection for L1 German speakers (12 females, 2 males; ages 20–29, $M = 22.9$, $SD = 2.492$), took place in Stuttgart, Baden-Württemberg, Germany during a data collection period of ∼3 weeks. Seven of these also completed experiment 1 more than 30 min later after intervening experiments to avoid back-to-back presentation of two phoneme detection experiments. Supplementary material D presents additional participant details for this group. A one-semester data collection period with L1 American English adult learners of German was conducted at a Midwestern university. Participants were recruited via advertisement to all students enrolled in the university's second-semester German course during that term (150+ students) and voluntarily scheduled with the researcher to attend a data collection session at a campus computer lab during two data collection periods of ∼2 weeks. Ten learners completed the task at midterm and 19 at finals. Low enrollment in the study made a cross-sectional analysis unfeasible, so both time points were collapsed for analysis, selecting the second time point for those who completed both ($n = 6$). This yielded 22 unique L2+ participants (10 females, 12 males; ages 18–23, $M = 19.6$, $SD = 1.170$). One L2+ participant at the first time point reported simultaneous bilingualism in English and Latvian. Another reported Spanish exposure since birth, first use at age 3. Two reported initial exposure to German at age 3 or "very young" but no use until late teens. One reported birth and residence in Australia until age 4.

8.1.2 Stimuli

Experiment 2 investigates phonotactic awareness of both dorsal fricatives and [h], which is phonemic /h/ in English and German. Supplementary Table E1 summarizes conditions and trial types. The single block of items included 48 critical trials balanced for the Licit and Illicit contexts in six conditions ($n = 8$ trials per condition). As long as a dorsal fricative matches the place of the preceding vowel (DFA), German phonotactics permits both [ç] and [x] in coda clusters followed by [t]. In contrast, /h/ is never allowed in syllable codas, simple or complex. Licit conditions were Front Match [εç] (e.g., [glεçt]), Back Match [ax] (e.g., [glaxt]), and Onset-[h] (e.g., [hamt]). Illicit counterparts were Back–Front Mismatch *[aç] (e.g., *[glaçt]), Front–Back Mismatch *[εx] (e.g., *[glεxt]), and *Coda-[h] (e.g., *[gaht]). Similar to experiment 1, these were counterbalanced by 144 distractors with [t] in other positions and 192 fillers without [t] for a total of 384 trials with 1:3 critical:distractor ratio and 50% fillers. Supplementary material F provides a complete list of stimuli. Stimuli were recorded as in experiment 1.

Segment boundary marking and extraction of stimuli onset and duration were conducted as in experiment 1. Supplementary Table E2 describes these by condition in aggregate; see Supplementary material F for the extracted durations of each stimulus and Scott (2019a, p. 351–354) for additional analyses of the stimuli. Augmented RT for analysis was derived as in experiment 1.

TABLE 8 Signal detection rates for experiment 2 after participant exclusion ($n_{L1} = 9$, $n_{L2+} = 14$).

Group × condition	Signal detection			
	Hit	False alarm	Correct rejection	Miss
L1				
Dorsal licit [ax εç]	0.938	–	–	0.063
Dorsal illicit *[aç εx]	0.854	–	–	0.146
Glottal licit Onset-[h]	0.833	–	–	0.167
Glottal illicit *Coda-[h]	0.778	–	–	0.222
Distractors[a]	0.803	–	–	0.197
Fillers	–	0.012	0.988	–
L2+				
Dorsal licit [ax εç]	0.866	–	–	0.134
Dorsal illicit *[aç εx]	0.871	–	–	0.129
Glottal licit Onset-[h]	0.893	–	–	0.107
Glottal illicit *Coda-[h]	0.875	–	–	0.125
Distractors[a]	0.772	–	–	0.228
Fillers	–	0.021	0.979	–

Dashes indicate fields for which no rate is possible with the "go"/"no-go" format.
[a]Distractor Miss rate may include responses entered before the experiment response logger initiated.

8.1.3 Procedure

The same computer lab in Stuttgart was used for experiment 2, which was always administered before experiment 1 for those who completed both. Sessions lasted 100–120 min and included a language background questionnaire and two additional experiments of a different type. This group received €15 payment.

U.S. data collection took place in a university language laboratory, starting with a language background questionnaire and then testing on desktop computers running Windows 7 Service Pack 1 (64-bit). Additional specifications varied by computer: 2.6, 3.4, or 3.6 GHz processor; 4, 8, or 16 GB RAM; screen resolutions of 1024 × 768, 1440 × 900, or 1680 × 1050. Stimuli were presented with high-quality Sanako (Tandberg Educational) SLH-07 circumaural headphones; participants could adjust the volume themselves.

TABLE 9 Descriptive statistics and tests of normality for experiment 2 by condition (L2+).

Condition	*n*	μ(ms)	σ	Skewness		Kurtosis		Shapiro-Wilk	
					SE		*SE*	*W*	*p*
Licit	291	543	128.2	0.816	0.143	0.081	0.285	0.941	<0.001
DFA match	192	543	130.5	0.877	0.175	0.075	0.349	0.926	<0.001
Onset-[h]	99	543	124.3	0.689	0.243	0.140	0.481	0.961	0.005
Illicit	291	555	126.8	0.476	0.143	−0.258	0.285	0.977	<0.001
DFA mismatch	194	519	116.0	0.614	0.175	−0.164	0.347	0.960	<0.001
*Coda-[h]	97	628	116.1	0.469	0.245	−0.191	0.485	0.972	0.035

Reaction time values rounded to the nearest millisecond. Observed positive skewness is most likely not an artifact of data trimming, which excluded only three DFA trials (2 × [εç] + t Match, 1 *[εx] + t Mismatch) and one Glottal trial (Onset-[hε]) as non-filler fast responses (<100 ms).

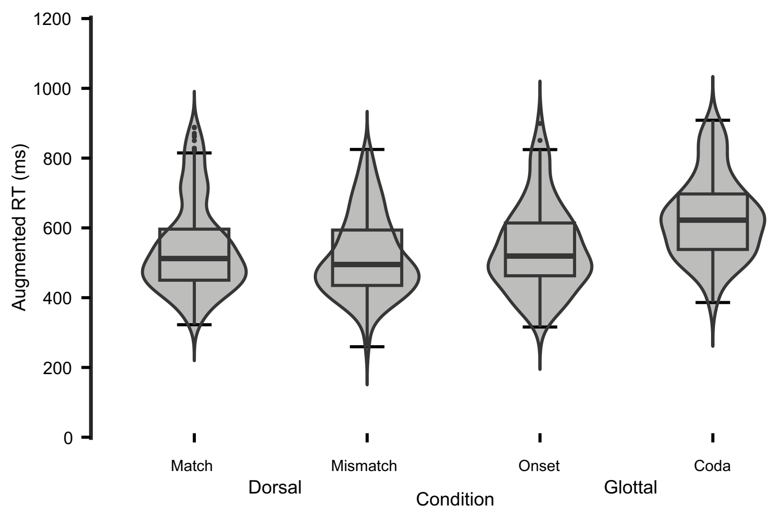


FIGURE 6
Experiment 2 (L2+): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data by fricative (dorsal vs. glottal) and condition (licit vs. illicit).

Task presentation, training phase, within-block randomization, and trial structure were the same as for the [t]-detection portion of experiment 1 but with OpenSesame (Mathôt et al., 2012; Version 2.8) and opportunity for self-paced breaks every 32 trials. With breaks in between, the experiments lasted ∼90 min. Participants in midterm data collection received US$10; those at the finals session received a 1% bonus German course credit. Returning participants were entered in drawings for one US$50 cash prize per 10 returning participants.

8.2 Results

8.2.1 Exclusion criterion

The same minimum threshold of five or more hit responses was applied to the four conditions of experiment 2: Dorsal Licit [ax εç], Dorsal Illicit *[aç εx], Onset-[h] Licit, and *Coda-[h] Illicit. This criterion retained 9 of the L1 German group and 14 of the L1 English group. I acknowledge that these are small numbers. This is due to difficulty in recruiting participants to voluntarily schedule a long laboratory session during limited data collection periods at each location (e.g., <20% of the course enrollment of L2+ learners during the semester) as well as a deliberate effort to avoid recruiting linguistics majors in Stuttgart (the most available and willing group), which would have skewed the L1 German data through their atypical metalinguistic awareness. Nonetheless, the statistical techniques employed in this study should be able to handle the small data set and provide useful insights that may be tested with replication. Table 7 describes the remaining data set, and Table 8 displays signal detection rates by group and condition. The L2+ group exhibits little variation between critical conditions, whereas the L1 German group shows consistently lower hit rates for illicit contexts than for licit contexts within each fricative type condition. They also exhibit lower hit rates for the glottal [h] fricative type conditions than for the dorsal fricative type.

8.2.2 Data trimming and preparation

The data trimming procedure was as in experiment 1. The fast trial cutoff (<100 ms) excluded four trials from the L1 German data set (2 × *[aç], 1 × [εç], 1 × *[ah]) and four trials from the L2+

TABLE 10 Estimated marginal means (ms, hits only) of experiment 2 (L2+), *SE*, and 95% confidence interval (top), with parameter estimate, variability, *SE*, *df*, *t*-value, and *p*-value (bottom).

Condition	Mean RT	*SE*	95% confidence interval	
			Lower bound	Upper bound
Licitness condition				
Licit				
Dorsal [ax ɛç] + t	550	19.3	512	588
Glottal Onset-[h]	547	19.3	509	585
Illicit				
Dorsal *[aç ɛx] + t	522	19.3	484	560
Glottal *Coda-[h]	623	19.4	590	666
Fricative				
Dorsal				
Licit [ax ɛç] + t	550	19.3	512	588
Illicit *[aç ɛx] + t	522	19.3	484	560
Glottal				
Licit Onset-[h]	547	19.3	509	585
Illicit *Coda-[h]	623	19.4	590	666

Fixed effects	Estimate	*SE*	*df*	*t*	*p*
Intercept	562	15.27	15.64	36.79	<0.001
[Condition (1) = Mismatch]	−13	6.63	44.17	−2.02	0.050
[Fricative (1) = Glottal]	−26	7.31	24.18	−3.51	0.002
[Condition = Mismatch]* [Fricative = Glottal]	27	6.62	44.09	4.11	<0.001
Parameter	Variance	σ			
Residual	11,366	107			
Participant	2,648	51			
Fricative	133	12			
Item	931	31			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Fricative reference level is Dorsal. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

German data set (2 × [ɛç], 1 × *[ɛx], 1 × [hɛ]). As in experiment 1, augmented RT data are not log-transformed. Skewness, kurtosis, and Shapiro–Wilk tests of normality are reported. See Scott (2019a) for further discussion of normality conformity and violation in these data.

8.2.3 Intergroup comparison

Supplementary material G summarizes the descriptive statistics and tests of normality for L2+ (Supplementary Table G1) and L1 (Supplementary Table G2) groups. Note that, except for two vowel–consonant pairing conditions (licit [ɛç] and illicit *[ah]), the L1 German group exhibits a slower mean RT across the board. Furthermore, while the L1 group data satisfy the test of normality for licit trials and all licit subconditions, the L2+ learner group data do not. With RT experiments, such intergroup differences may reflect real behavioral differences rooted in the fundamental difference between L1 speakers and early L2+ learners. Alternatively, as the groups were tested in different laboratory settings, this may reflect hardware or software latency differences or a combination of these factors. The present research questions do not entail intergroup comparisons, and retaining Group as an additional fixed effect might hinder statistical modeling with this small data set. Therefore, separate models for L2+ and L1 groups are reported here.

8.2.4 L2+ German analysis

Descriptive statistics and normality tests for the L2+ group in DFA and [h] conditions (licit and illicit for each) are displayed in Table 9. Figure 6 depicts the distribution of responses by condition, with generally longer RT for illicit coda-[h] than all other conditions. Supplementary Table G1 shows that, for the L1 English L2+ German group, mean RT was similar overall between Licit conditions, most of which violate the assumption of normality. This establishes baseline task performance for the L2+ group across fricative conditions. A linear mixed-effects model was run on the L2+ RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Licit vs. Illicit) and Fricative (Dorsal vs. Glottal) were declared as fixed effects. Participants and items were declared as random effects grouping factors. The maximal model was incrementally simplified, resulting in a final model that included Condition, Fricative, and Condition*Fricative as fixed effects with random intercept and Fricative as a random effect under participant. Table 10 reports the estimated marginal means and parameter estimates. The parity of licit conditions between fricative types is evident in Table 9 and Figure 6.

With the type III tests of mixed effects, the *F*-tests show the main effects for Condition, $F_{(1,44.17)} = 4.061$, $p = 0.050$, and for Fricative, $F_{(1,24.18)} = 12.288$, $p = 0.002$. The interaction of these is also significant, $F_{(1,44.09)} = 16.861$, $p < 0.001$. In summary, the model finds the RT for illicit trials is significantly different from the licit trials, driven largely by pronounced processing inhibition in response to violation of *Coda-/h/ . To check whether the relatively fast RT trend with illicit *[aç ɛx] trials indicates a facilitation effect arising from sensitivity to violation of DFA, dorsal fricative match and mismatch RTs were compared directly. An independent samples *t*-test reveals a marginal trend, $t_{(384)} = 1.936$, $p = 0.054$ (small effect size, $d = 0.197$, $SE = 0.102$). The model confirms that the reliable inhibition effect observed for the glottal fricative in the illicit *Coda-[h] context vs. the licit Onset-[h] context is not shared by the dorsal fricative type, which instead may show a slight facilitation for illicit *[aç ɛx] contexts as compared to licit [ax ɛç]. In short, robust inhibition for violations of *Coda-[h] drives the significant effects of the mixed-effects model, although this group also shows a marginal trend of facilitation for violations of DFA (cf. Weber, 2001a,b).

8.2.5 L1 German analysis

Descriptive statistics and normality tests for the L1 group in DFA and [h] conditions (licit and illicit for each) are displayed in Table 11.

TABLE 11 Descriptive statistics and tests of normality for experiment 2 by condition (L1).

				Skewness		Kurtosis		Shapiro–Wilk	
Condition	*n*	μ(ms)	σ		*SE*		*SE*	*W*	*p*
Licit	194	573	122.4	0.209	0.175	−0.260	0.347	0.990	0.203
DFA match	134	554	118.4	0.436	0.209	0.123	0.416	0.980	0.049
Onset-[h]	60	615	121.3	−0.290	0.309	−0.062	0.608	0.986	0.716
Illicit	176	588	135.6	0.472	0.183	−0.418	0.364	0.971	0.001
DFA mismatch	121	552	118.5	0.607	0.220	0.236	0.437	0.974	0.018
*Coda-[h]	55	666	135.8	0.039	0.322	−1.017	0.634	0.962	0.083

Reaction time values rounded to the nearest millisecond. Observed positive skewness is most likely not an artifact of data trimming, which excluded only three DFA trials [2 × *[aç] + t Mismatch, 1 × [ɛç] + t Match] and one Glottal trial [*Coda-[ah]] as non-filler fast responses (<100 ms).

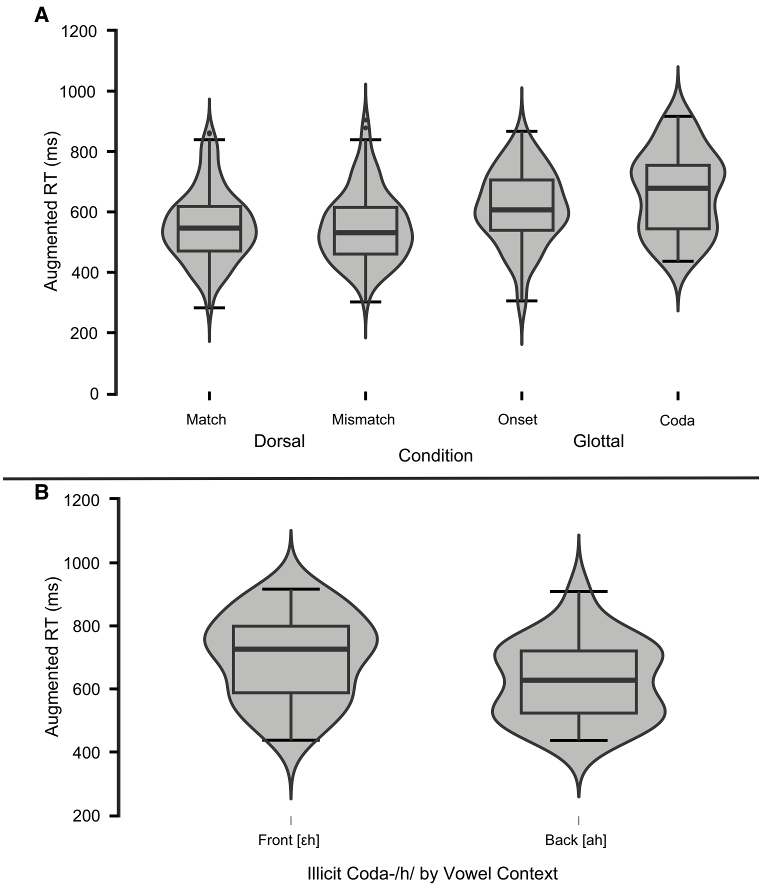


FIGURE 7
Experiment 2 (L1): violin-augmented boxplot showing the median, interquartile range, minimum and maximum, and outliers of augmented reaction time (RT) data **(A)** by fricative (dorsal vs. glottal) and condition (licit vs. illicit) and for **(B)** illicit Coda-/h/ by vowel context ([a] vs. [ɛ]).

Figure 7A depicts the distribution of responses by condition, showing longer RT for both [h] conditions, especially illicit coda-[h], as compared to dorsal fricatives. Supplementary Table G2 shows that all L1 German licit condition data satisfy the assumption of normality; however, this group's glottal fricative conditions tend to have slower RT than the dorsal conditions. In addition, the L1 group's RT means are reliably slower in conditions containing velar [x] than in palatal [ç] conditions, a pattern not shared by the L2+ group.

A linear mixed-effects model was run for the L1 RT data in JASP (JASP Team, 2023; Version 0.18.1). Condition (Licit vs. Illicit) and Fricative (Dorsal vs. Glottal) were declared as fixed effects. Participants and items were declared as random effects grouping factors. The maximal model was incrementally simplified, resulting in a final model that includes Condition, Fricative, and Condition*Fricative as fixed effects with random intercept and Fricative as a random effect under participant (as with the L2+ model). Table 12 reports the estimated marginal means and

TABLE 12 Estimated marginal means (ms, hits only) of experiment 2 (L1), *SE*, and 95% confidence interval (top), with parameter estimate, variability, *SE*, *df*, *t*-value, and *p*-value (bottom).

Condition	Mean RT	*SE*	95% confidence interval	
			Lower bound	Upper bound
Licitness condition				
Licit				
Dorsal [ax ɛç] + t	554	26.6	502	606
Glottal Onset-[h]	619	29.9	560	677
Illicit				
Dorsal *[aç ɛx] + t	556	26.7	503	609
Glottal *Coda-[h]	672	30.2	613	731
Fricative				
Dorsal				
Licit [ax ɛç] + t	554	26.6	502	606
Illicit *[aç ɛx] + t	556	26.7	503	609
Glottal				
Licit Onset-[h]	619	29.9	560	677
Illicit *Coda-[h]	672	30.2	613	731

Fixed effects	Estimate	*SE*	*df*	*t*	*p*
Intercept	600	24.9	8.63	24.07	<0.001
[Condition (1) = Mismatch]	−14	7.16	46.987	−1.92	0.060
[Fricative (1) = Glottal]	−45	9.01	12.31	−5.00	<0.001
[Condition= Mismatch]*[Fricative = Glottal]	13	7.16	46.96	1.79	0.080
Parameter	Variance	σ			
Residual	9,124	96			
Participant	5,138	72			
Fricative	268	16			
Item	933	31			

Reaction time (RT) values rounded to the nearest millisecond. Estimated marginal means are reported. Condition reference level is Match; Fricative reference level is Dorsal. Significance calculation includes Holm–Bonferroni correction for multiple comparisons.

parameter estimates. In contrast to the L2+ data, a clear difference in licit conditions between fricative types is evident in Table 11 and Figure 7A.

With the type III tests of fixed effects, the *F*-tests show a significant main effect for Fricative, $F_{(1,12.31)} = 24.986$, $p < 0.001$. There is a marginal effect of Condition, $F_{(1,46.99)} = 3.701$, $p = 0.060$, and the interaction of these is also marginal, $F_{(1,46.96)} = 3.210$, $p = 0.080$. In summary, the model confirms only that the relatively slower RT for the glottal fricative trials is significant. Unlike the L2+ group, the L1 group's licit and illicit dorsal fricative trials have nearly identical distributions, and the observed difference between licit and illicit glottal fricative trials is not sufficient to drive an unambiguous main effect for Condition or an interaction effect, given that glottal trials represent only one third of the data set. However, an independent samples *t*-test conducted without the dorsal fricative data shows that the difference between licit Onset-/h/ and illicit *Coda-/h/ trials is significant, $t_{(113)} = 2.101$, $p = 0.038$, with a small-to-medium effect size, $d = 0.392$, $SE = 0.190$. Thus, an RT inhibition effect for violating *Coda-/h/ appears to single-handedly drive the marginal trends observed in the model. Like experiment 1, this model provides no clear evidence of sensitivity to violation of DFA in the L1 German group.

It is apparent in Supplementary Table G2 and Figure 7B that the glottal fricative illicit *[ɛh] subcondition is markedly slower than the illicit *[ah] subcondition, although both equally violate *Coda-/h/. To investigate whether the inhibition observed for *Coda-/h/ trials might be specific to one vowel context, these subconditions were compared directly. An independent samples *t*-test shows that this difference is significant, $t_{(53)} = 2.084$, $p = 0.042$, with a medium effect size, $d = 0.564$, $SE = 0.280$. Thus, the illicit *[ɛh] subcondition specifically contributes most strongly to the marginal trends observed in the model. In contrast, the illicit *[ah] subcondition has a distribution more comparable to the licit Onset-/h/ condition and a mean RT relatively closer to the licit [ax] condition. This suggests that there may be a principled difference in the L1 German group's processing of /h/ in syllable codas based on the preceding vowel, such that *[ah] does not reliably trigger (as much) inhibition. I return to this in Section 8.3.2.

8.3 Discussion of experiment 2

8.3.1 Facilitation trend with DFA violations in L2+ learners not found in Swabian Germans

Experiment 2 uses the [t]-detection innovation, which stipulates a Persean listening target that has equivalent phonemic status in English and German, and with stimuli that exhibit a release burst realization appropriate to the coda position in both languages. The dorsal fricative condition undertook to replicate either Weber's (2001a,b, 2002) processing facilitation findings or Lindsey's (2013) processing inhibition findings. Either of these RT effects for either participant group would answer the third research question (RQ3) in the affirmative for each of the L2+ (RQ4a) and L1 German (RQ4b) groups and the methodological question of whether the Persean task design is sensitive enough for investigating progressive assimilation processes (RQ4c). The L2+ group shows only a marginal trend, with a small effect size, suggesting that the early L2+ learners may have a weak tendency toward facilitation (faster RT) in response to violation of DFA (RQ4a; cf. Weber, 2001a,b). The L1 German group shows no facilitative or inhibitory effect (RQ4b; contra Lindsey, 2013, and Weber). Thus, no previously reported RT effect is replicated here (RQ3), yet the adjacent [t]-detection task's ability to detect listener sensitivity to violations of progressive assimilation may be provisionally affirmed (RQ4c), pending more conclusive replication.

The negative result for DFA with L1 German listeners may be due, in part, to the primary location of data collection (Stuttgart). Lipski's (2006) magnetoencephalography study, which found no effect of DFA violation, also collected data in Stuttgart and nearby

Tübingen, where the regional dialect is Swabian. According to Hall (2022, p. 82–85), [x] and [ç] both occur only in a post-sonorant position in Swabian, and they are in an allophonic relationship through DFA (a case of velar fronting). But southern Baden-Württemberg, including Swabia, abuts High Alemannic and transition dialect areas within 80–120 km, where only velar [x] occurs. Both university cities, and Stuttgart as regional capital, afford ample opportunity for exposure to speakers from a broader dialect area and thus for more variable local input in terms of DFA than in Regensburg or Hannover, where Weber sampled. In short, the failure of experiments 1 and 2 to replicate previous DFA findings may arise from the fact that, for many L1 German speakers in Stuttgart, the Front–Back Mismatch [ɛx] sequence, although illicit and unprecedented in Standard German, is not unusual to encounter in natural speech and thus less likely to trigger any reliable RT effects for groups in aggregate—neither facilitation nor inhibition of processing. Scott's (2019a, p. 292, 317–324) additional analysis of individuals' differences of means reveals that this variability is not limited to participants tested in Stuttgart or who reported a Swabian dialect background. This highlights that sensitivity to DFA likely depends greatly on the dialect exposure of individual speakers. The present experiments may not have detected sensitivity to DFA violations simply because that sensitivity is not a robust feature of this L1 German sample group.

Experiment 2 found only a marginal facilitation trend for the L2+ group when presented with violations of DFA. This response pattern, which was less internally variable (more uniform) than for the L1 German group, further suggests that dialect variation among L1 Germanophones plays an important role in the perception of dorsal fricatives. It may also be the case that the choice of listening target, a phone occurring *after* a progressive assimilation, exchanges some degree of task sensitivity to small effects as the methodological price of avoiding explicit attentional focus on conditioning or application environments in phoneme detection: a trade-off for cross-group task equivalence. Additional research is required to address these questions.

8.3.2 Inhibition with Coda-/h/ : broad in L2+ learners, context-dependent for L1 Germans

I included the glottal fricative condition to test the Persean variant of the phoneme detection task's ability to investigate sensitivity to violation of a phonological constraint that is fundamentally different from place assimilation. This is the first phoneme detection study to investigate sensitivity to violation of a constraint on syllable well-formedness: the phonotactic/prosodic constraint *Coda-/h/.[9] The L2+ group's responses to violations of this constraint exhibit unambiguous and pronounced RT inhibition (RQ5a). The L1 German group's response pattern is more selective: Only the *[ɛh] subcondition shows inhibition comparable to the L2+ group (RQ5b). In contrast, RT in the *[ah] subcondition is more comparable to licit conditions such as Onset-/h/ or the [ax] sequence. Together, these results demonstrate that the adjacent [t]-detection task unambiguously detects sensitivity to violations of this phonotactic/prosodic constraint governing syllable well-formedness (RQ5c).

Regarding the *Coda-/h/ constraint, the intergroup results align and differ in important ways. The L2+ group exhibits robust inhibition, but the L1 group's specific sensitivity to *[ɛh] in codas complicates the scenario. Crucially, processing inhibition is mitigated for its counterpart *[ah], not differing markedly from the licit [ax] subcondition. This is a likely candidate for *phonotactic assimilation*, described by Seguí et al. (2001, p. 198), of which they outline three types:

> (1) the listener "ignores" individual phonemes (or stress patterns: Dupoux et al., 1997) that are present in the signal; (2) the listener "perceives" illusory phonemes that have no acoustic correlate in the signal; (3) the listener "transforms" one phoneme in (sic!) another.

The lack of difference in perception between illicit *[ah] and licit [ax] in syllable codas likely falls under the third type—that is, the L1 German group appears to perceptually "transform" the contextually illicit [h] into an instance of a spectrally similar fricative (Supplementary material A) that is licensed to follow back [a]—namely, into [x]—and then reacts accordingly, without RT inhibition. This crucial difference between *[ah] and its corresponding *[ɛh], which is much less easily repaired in an analogous way due to spectral differences between [h] and [ç], detracts from the overall sensitivity to *Coda-/h/ violations for this group. Indeed, how RT experiment methodology interacts with prelexical phonotactic assimilation is an open empirical question relevant to broader questions about prediction during language processing more generally (e.g., Kaan and Grüter, 2021; Key, 2014). Alternatively, as the *[ɛh]-condition is not phonotactically assimilated in this way, this difference may serve as indirect support for the argument that L1 German speakers do, in fact, maintain a phonological distinction between which vowel + dorsal fricative pairs are permitted (RQ4b). This may depend on the interaction of the frontness/backness of each segment, despite the negative result for violations of DFA in the present study.

9 General discussion and conclusion

9.1 Implications and context of results

The central aim of this study was to develop a version of the phoneme detection task with task parity for participant groups with different language backgrounds and to test the method on various phonological phenomena. The first experiment with L1 German speakers unambiguously replicated processing inhibition (slower RT) findings in response to violations of RNA (RQ1; Otake et al., 1996; Weber, 2001a,b). In addition, reanalysis of Scott's (2019a) nasal data with a model that differentiates underspecified

9 In postvocalic coda position, both German and (rarely) English orthography use <h> not to indicate a consonant, but rather as a diacritic of vowel duration (in English, perhaps also vowel quality), for example, German *Stahl* [ʃtaːl], "steel," *Mehl* [meːl], "flour," and English *yeah* [jæː], "yes," *nah* [næː], "no," or *ah* [aː] and *meh* [mɛː] (interjections). These contextual uses of <h> may predispose both groups in this study to process any perceived frication noise in the stimuli as cue evidence *against* the perception of /h/ in the signal. This may contribute to the activation of other fricatives as more viable competitor candidates. As with the dorsal fricatives, stipulating /t/ as the listening target discourages explicit attention to the [h] itself.

mismatch and specified mismatch conditions suggests that there may be a continuum of processing inhibition for assimilation of this type according to the intensity of the RT effect such that phonotactic constraints may trigger more severe inhibition than mere underapplication of place assimilation, while rare but precedented sequences may trigger a lesser effect than both. An attempt to analyze the observed difference between specified mismatch *[ŋp] vs. *[mk] was inconclusive, leaving the phonemic status of [m] vs. the context-dependent allophone [ŋ] an open question for future research with a larger data set (RQ2).

Both experiments attempted to replicate any sort of consistent RT effect in response to violation of DFA with an L1 German group (RQ3), but the samples in this study exhibited neither processing facilitation (Weber, 2001a,b) nor processing inhibition (Lindsey, 2013). Experiment 2 revealed a weak facilitation (faster RT) effect in L2+ learners in reaction to violation of DFA (RQ4a). This result is similar to Weber's (2001a,b) L1 German result, and contra Lindsey's L2+ learner result, but does support the conclusion that phoneme detection with the Persean listening target [t] is sensitive enough to investigate progressive assimilation (RQ4c). An unexpected asymmetry in the results between the illicit Coda-[h] conditions by vowel context suggests that the L1 German group may have some phonotactic preference for front–back agreement between a nuclear vowel [a] or [ɛ] and a following dorsal fricative despite the inconclusive result in the DFA condition (RQ4b).

Finally, to test this method with more than place assimilation, experiment 2 undertook the first phoneme detection investigation of sensitivity to the prosodic/phonotactic constraint against /h/ in syllable codas in both German and English. Both groups exhibited unambiguous processing inhibition in response to illicit Coda-/h/ (RQ5a, RQ5b), confirming that the Persean listening target technique can be fruitfully employed to investigate a non-assimilation type of phonological knowledge (RQ5c). Additionally, the observed asymmetry in the L1 German group's responses—inhibition specifically for the *[ɛh] subcondition but not for the *[ah] subcondition—suggests that this group's prelexical perception may reanalyze the illicit [h] as a licit [x] following the back vowel [a] as a sort of phonotactic assimilation (Seguí et al., 2001).

This study provisionally supports argumentation (e.g., Weber, 2001a,b) that violation of strong phonological expectations (e.g., regressive assimilation, phonotactic/prosodic constraints) yields profound inhibition in phoneme detection, whereas violation of weak phonological expectations (e.g., progressive assimilation) yields a smaller facilitation effect. Future research is needed to establish further systematic predictions about which phoneme detection task and listening target designs elicit which RT effects (facilitation or inhibition) and intensity (small or weak effects) with various populations (L1, L2+), whose representations may diverge in precision and robustness (Bordag et al., 2021; Cook et al., 2016; Darcy et al., 2007, 2013).

The second experiment undertook the first phoneme detection investigation of /h/ ungrammatically located in syllable codas, revealing different patterns of processing inhibition between L1 and L2+ groups. This result has interesting implications for methodological approaches to investigating questions in L2+ phonological theory. The robust processing inhibition effect exhibited in the L2+ learner group clearly validates the utility of the Persean listening target technique for investigating prosodic/phonotactic constraints other than the segmental adjacency effects of place assimilation. The L1 German group's inhibition pattern generally confirms this. But the lack of this effect specifically in the *[ah]+[t] condition in the L1 German group suggests an additional, subtler context-dependent sensitivity. Recalling the three types of phonotactic assimilation outlined by Seguí et al. (2001), finding reports of listeners phonotactically assimilating input in illicit contexts is not difficult. Listeners may fail to detect individual phonemes in the signal (Type 1, e.g., L1 Mandarin failure to discriminate syllables with English coda laterals from open CV syllables; Wang, 2023) or illusorily perceive phonemes that have no acoustic basis in the input (Type 2, e.g., L1 Korean epenthesis of a vowel within word-initial consonant clusters in English; Darcy and Thomas, 2019). In contrast with the L1 German group's significant processing inhibition in response to Coda-[h] following the front vowel [ɛ], the same group's lack of RT effect in response to Coda-[h] following the back vowel [a] suggests that they did not perceive the [h] as /h/ at all, but instead as [x], an allophone of the dorsal fricative that would be licit following /a/ (Type 3, transformation of one phoneme to another). Interestingly, the fact that this prelexical perceptual repair strategy does not appear to be viable in the front vowel [ɛ] context suggests that this group has a preference for agreement of place between a vowel and a following dorsal fricative, despite the negative result for the DFA conditions. Nonetheless, phonotactic assimilation is constrained to some degree by an acoustic similarity between the signal (here, [h]) and the potential percept (here, [x] in the back context and [ç] in the front environment). This interaction of two types of phonological knowledge suggests that further research may be able to intentionally leverage such subtleties to investigate one phenomenon of interest (e.g., place assimilation) through phoneme detection tasks aimed at obliquely related phenomena (e.g., prosodic/phonotactic constraints on placement of a phone with similar acoustic cues).

9.2 Design phoneme detection experiments for statistical power *a priori*

Particular care is warranted in task design to maintain statistical viability when investigating smaller effects such as processing facilitation in response to violation of weak phonological expectations (e.g., progressive assimilation). The present reanalyses draw finer distinctions within factors than Scott (2019a) initially designed for (i.e., two subtypes of nasal mismatch and sorting coda glottal fricative trials by vowel). This subdivides relatively small data sets to explore additional questions *ad hoc*. In practice, the bar of maximal models set by Barr et al. (2013) is too high for many data sets to meet (Matuschek et al., 2017). The present study simplifies model dimensions so that the results derive sufficient support from the available data. These results may lack enough statistical power to conclusively address specific empirical questions about each of the speech perception phenomena investigated, particularly for small effects. For example, the overall results signal that violation of RNA and *Coda-/h/ induces RT inhibition generally, but smaller RT effects, such as facilitation, may be marginal, risking Type II error (false negative). To model several fixed and random effects simultaneously, this study employs linear mixed-effects models. When these data were collected in 2015, there was no standard for *a priori* power analysis of such models with multilevel factors (Kumle et al., 2021, p. 2,528–2,529). Statistical practice also indicates against post-experimental power calculations (so-called *observed power*; e.g., Hoenig and Heisey, 2012; Lydersen, 2019). Thus, no power analysis is provided here. Future studies using phoneme detection to investigate phonological and phonotactic sensitivity should include simulation-based *a priori* power analysis at the design stage to determine the appropriate number of trials and sample sizes (see Kumle et al., 2021). Additionally, more statistical research is needed to establish appropriate field standards for how to statistically model RT task data sets, which often violate assumptions of normality.

9.3 Design phoneme detection L2+ experiments for task parity

Most phoneme detection studies have focused on phonological interactions between adjacent segments (e.g., place assimilation) and the effects of phonetic transition cues with L1 speakers (e.g., Foss, 1969; Frauenfelder and Seguí, 1989). Although some include non-learner comparison groups from other language backgrounds (cross-language listeners; e.g., Otake et al., 1996; Weber, 2001a,b), phoneme detection investigations of L2+ learner groups are less common (e.g., advanced L2+ learners, Lindsey, 2013). Our collective awareness and comfort with RT psycholinguistic methods generally in L2+ acquisition research is still developing with necessary caution (Hui and Jia, 2024).

When building RT experiments for L2+, thorough methodological consideration is necessary to reduce noise from unintended sources of RT differences between groups, such as from hardware and software latency, stimuli that could be phonetically ambiguous for one language group, and the like. Phoneme detection tasks may require presenting the listening target within a very narrow temporal window (e.g., adjacent phones) for small RT effects to remain detectable (i.e., to avoid Type II error).

One underexamined challenge is achieving task parity between language background and target language proficiency groups in terms of phonological and orthographic representation as discussed in this study. Phoneme detection studies historically vary by the relationship between the listening target and the phenomenon of interest. Listening targets may be subject to a phonological principle (the object of interest *x*; e.g., nasals in RNA, fricatives in DFA and *Coda-/h/ contexts), they may trigger application of a phonological principle (e.g., obstruents in RNA, vowels in DFA), or they may be merely adjacent without being implicated in the phonological principle of interest (a Persean listening target *y*, e.g., [t] following fricatives in DFA and *Coda-/h/ contexts). There are countless fundamental representational differences between the phonology, orthography, and the GPCs of any given listening target for different L1 and L2+ populations. As a result, using phoneme detection—or any task design that requires explicit labels for phones or phonemes—in L2+ acquisition research requires careful consideration of the equivalence of the relationship of the listening target and its label to potential mental representations for each language group under investigation. When learner and L1 language groups have equivalent relationships with the experimental task, the experimenter can be more confident that RT results show evidence that the L2+ mental representation is different from the L1 mental representation for the phenomenon of interest, not the trivial result that different cognitive tasks have different performance speeds. As we improve our understanding of the representational challenges that come with listening target labels in phoneme detection, we gain a powerful methodological tool for investigating phonological knowledge in a wide variety of cross-language and L2+ learning scenarios.

9.4 Design phoneme detection experiments for L1 varieties and L2+ learner trajectories

This study suggests that the phoneme detection paradigm can be leveraged to investigate underlying mental representations, such as theories of feature underspecification. We can ask a variety of interesting theoretical questions: (How) do L2+ learners' representations of phonological (un)grammaticality change from L1-based to L2-based in the course of IL development? What do the earliest steps of IL development look like? What does ultimate attainment look like in L2+ phonological perception, and does it ever become target-like/optimal? By stipulating an adjacent, phonologically uninvolved, and acoustically consistent Persean listening target, phoneme detection can be a cognitively equivalent task for different groups when investigating important phonological questions with a variety of L1 dialect and regiolect groups and L2+ learners at different stages of IL development. Phoneme detection provides an instrument that enables investigation of implicit, prelexical processing, even with participants who lack any TL experience. Future research should also consider a wider variety of scenarios. This instrument is readily adaptable for uninstructed L2+ learners or non-reading immigrant groups in a TL environment or for comparing groups whose L1 literacies use different non-alphabetic scripts.

9.5 Design phoneme detection experiments with Persean listening targets

This study addresses the labeling problem of phonology, orthography, and GPCs for perception research in both L1 and L2+ scenarios with a methodological solution. The experiments presented here establish that directing listener focus away from application and conditioning environments to a Persean listening target (*y*) can meet the challenge of stipulating representationally equivalent listening targets across groups while also more clearly tapping into implicit knowledge of the objects of interest (*x*) that are not attended to explicitly. This is achieved while retaining sufficient task sensitivity to investigate implicit or automatized explicit linguistic knowledge in speech processing for a variety of assimilation phenomena, cue weighting and fusion strategies, and prosodic/phonotactic constraints.[10] Different types of phenomena

10 For similar work in other areas of L2+ grammar, see Rebuschat (2013) and Suzuki (2017).

may need to be investigated separately, as different RT effects (processing facilitation or inhibition) may not lend themselves to combined models.

Directing attention to a specific phone, such as in identification and perceptual assimilation tasks (Best, 1995; Best and Tyler, 2007), probes perception at a highly explicit level rather than implicit phonological knowledge (cf. *optimal*, Bordag et al., 2021; *automatic*, Strange, 2011). When we are more interested in implicit knowledge including abstract phonological representations such as phonemes or phonotactics, stipulating the object of investigation x as the listening target can place unintended emphasis on subphonemic phonetic detail in x. Moreover, for investigation of L2+ *acquisition* of phonological knowledge, stipulating the object of investigation x as the listening target can confound both intergroup congruence (of phonological and orthographic domains and GPC mappings) and assessment of pre-learner and IL developmental stages in L2+ scenarios. Avoiding the target of investigation as the target of listener attention and redirecting focus to a reliable and congruent Persean listening target y affords access to crucial questions in L2+ phonology, just as the shield's reflection enabled Perseus to strike true.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation. The datasets analyzed for this study, including language background questionnaires, stimuli, experiment code, and anonymized data set, can be found in the Open Science Framework at https://osf.io/prfyw/.

Ethics statement

The studies involving humans were approved by Indiana University, Bloomington and Universität Stuttgart. The studies were conducted in accordance with the local legislation and institutional requirements. The ethics committee/institutional review board waived the requirement of written informed consent for participation from the participants or the participants' legal guardians/next of kin because collection of signed consent forms would have created additional identifying records of participants for otherwise de-identified data, increasing risk of loss of confidentiality.

Author contributions

JS: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research was supported by a short-term research grant from the Universität Stuttgart (Sonderforschungsbereich 732), a 2015 Language Learning Dissertation Grant, and a 2016 Graduate Travel Award from the College of Arts and Sciences at Indiana University. Partial funding for open access provided by the University of Maryland School of Languages, Literatures, and Cultures and by UMD Libraries' Open Access Publishing Fund.

Acknowledgments

I would like to thank Isabelle Darcy, Ken de Jong, Phil Lesourd, and Susanne Even for their guidance and feedback on the dissertation that led to this article. Thanks also to Christiane Kaden for patient hours in a recording booth, Bronson Hui for guidance with statistical analysis in JASP, Janet Scott for reading drafts and graphics assistance, the reviewers for their insightful comments, and the editors of this special collection. This article is based in part on parts of the author's dissertation and a conference proceedings article by the author, cited herein as Scott (2019a,b).

Conflict of interest

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

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/flang.2024.1254956/full#supplementary-material

References

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

Bassetti, B., Escudero, P., and Hayes-Harb, R. (eds.) (2015a). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist.* 36.

Bassetti, B., Escudero, P., and Hayes-Harb, R. (2015b). Second language phonology at the interface between acoustic and orthographic input. *Appl. Psycholinguist*. 36:393. doi: 10.1017/S0142716414000393

Bertelson, P. (ed.). (1986a). The onset of literacy. *Cognition* 24.

Bertelson, P. (1986b). The onset of literacy: liminal remarks. *Cognition* 24, 1–30. doi: 10.1016/0010-0277(86)90002-8

Best, C. T. (1995). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*, ed. W. Strange (Timonium, MD: York Press), 171–204.

Best, C. T., and Tyler, M. (2007). "Nonnative and second-language speech perception. Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds. O. S. Bohn and M. J. Munro (Philadelphia: Benjamins), 13–34.

Boersma, P., and Weenink, D. (2014). *Praat: Doing Phonetics by Computer* (Version 5.3.66) [Computer Program]. Available at: http://www.praat.org/ (accessed September 27, 2024).

Bordag, D., Gor, K., and Opitz, A. (2021). Ontogenesis model of the L2 lexical representation. *Bilingualism* 25, 185–201. doi: 10.1017/S1366728921000250

Castro-Caldas, A. (2004). Targeting regions of interest for the study of the illiterate brain. *Int. J. Psychol*. 39, 5–17. doi: 10.1080/00207590344000240

Christie, J., and Klein, R. (1996). Assessing the evidence for novel popout. *J. Exp. Psychol. Gen*. 125, 201–207. doi: 10.1037/0096-3445.125.2.201

Connine, C. M. (1994). "Vertical and horizontal similarity in spoken word recognition," in *Perspectives on Sentence Processing*, eds. C. Clifton, Jr., L. Frazier, and K. Rayner (Hillsdale, NJ: Lawrence Erlbaum Associates), 107–120.

Connine, C. M., and Titone, D. (1996). Phoneme monitoring. *Lang. Cogn. Process*. 11, 635–646. doi: 10.1080/016909696387042

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

Cutler, A., Butterfield, S., and Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *J. Mem. Lang*. 26, 406–418. doi: 10.1016/0749-596X(87)90099-4

Cutler, A., and Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *J. Mem. Lang*. 33, 824–844. doi: 10.1006/jmla.1994.1039

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

Darcy, I., Peperkamp, S., and Dupoux, E. (2007). "Bilinguals play by the rules: perceptual compensation for assimilation in late L2-learners," in *Laboratory Phonology 9*, eds. J. Cole and J. I. Hualde (Berlin: De Gruyter Mouton), 411–442.

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

Davis, S., and Cho, M.-H. (2003). The distribution of aspirated stops and /h/ in American English and Korean: an alignment approach with typological implications. *Linguistics* 41, 607–652. doi: 10.1515/ling.2003.020

Dehaene-Lambertz, G., Dupoux, E., and Gout, A. (2000). Electrophysiological correlates of phonological processing: a cross-linguistic study. *J. Cogn. Neurosci*. 12, 635–647. doi: 10.1162/089892900562390

Dollenmayer, D. B., Crocker, E. W., and Hansen, T. S. (2014). *Neue Horizonte: Student Activities Manual, 8th Edn*. Boston, MA: Heinle, Cengage Learning.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: a perceptual illusion? *J. Exp. Psychol. Hum. Percept. Perform*. 25, 1568–1578. doi: 10.1037//0096-1523.25.6.1568

Dupoux, E., Pallier, C., Sebastian, N., and Mehler, J. (1997). A destressing deafness in French? *J. Mem. Lang*. 36, 406–421. doi: 10.1006/jmla.1996.2500

Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., et al. (2014). Log-transformation and its implications for data analysis. *Shangai Archiv. Psychiat*. 26, 105–109. doi: 10.3969/j.issn.1002-0829.2014.02.009

Foss, D. J. (1969). Decision processes during sentence comprehension: effects of lexical item difficulty and position upon decision times. *J. Verb. Learn. Verb. Behav*. 8, 457–462. doi: 10.1016/S0022-5371(69)80089-7

Frauenfelder, U. H., and Seguí, J. (1989). Phoneme monitoring and lexical processing: evidence for associative context effects. *Mem. Cogn*. 17, 134–140. doi: 10.3758/BF03197063

Glover, J. (2014). *Liquid Vocalizations and Underspecification in German Dialects* (Dissertation). Indiana University, Bloomington, IN, United States.

Hall, T. A. (1989). Lexical phonology and the distribution of German [ç] and [x]. *Phonology* 6, 1–17. doi: 10.1017/S0952675700000920

Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German*. Tübingen: Max Niemeyer.

Hall, T. A. (1995). "Remarks on coronal underspecification," in *Leiden in Last. HIL Phonology Papers I*, eds. H. van der Hulst and J. van de Weijer (The Hague: Holland Academic), 187–203.

Hall, T. A. (1997). *The Phonology of Coronals*. Amsterdam: John Benjamins.

Hall, T. A. (2010). Nasal place assimilation in Emsland German and its theoretical implications. *Zeitschrift für Dialektologie und Linguistik* 77, 129–144. doi: 10.25162/zdl-2010-0005

Hall, T. A. (2014). Alveolopalatalization in Central German as markedness reduction. *Trans. Philol. Soc*. 112, 143–166. doi: 10.1111/1467-968X.12002

Hall, T. A. (2022). *Velar Fronting in German Dialects: A Study in Synchronic and Diachronic Phonology (Open Germanic Linguistics 3)*. Berlin: Language Science Press. doi: 10.5281/zenodo.7185567

Heeger, D. (1997). *Signal Detection Theory*. New York, NY: New York University. Available at: http://www.cns.nyu.edu/%7edavid/handouts/sdt-advanced.pdf (accessed September 27, 2025).

Hoenig, J. M., and Heisey, D. M. (2012). The abuse of power: the pervasive fallacy of power calculations for data analysis. *Am. Statist*. 55, 19–24. doi: 10.1198/000313001300339897

Hui, B., and Jia, R. (2024). Reflecting on the use of response times to index linguistic knowledge in SLA. *Ann. Rev. Appl. Linguist*. 2024, 1–11. doi: 10.1017/S0267190524000047

Hura, S. L., Lindblom, B., and Diehl, R. L. (1992). On the role of perception in shaping phonological assimilation rules. *Lang. Speech* 35, 59–72. doi: 10.1177/002383099203500206

Itô, J. (1989). A prosodic theory of epenthesis. *Nat. Lang. Linguist. Theor*. 7:2172259. doi: 10.1007/BF00138077

Iverson, G. K., and Salmons, J. (1992). The place of structure preservation in German diminutive formation. *Phonology* 9, 137–143. Available at: https://www.jstor.org/stable/4420048

JASP Team (2023). *JASP* (Version 0.18.1) [Computer Software]. Available at: https://jasp-stats.org/download/ (accessed September 27, 2024).

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German*. Amsterdam: John Benjamins.

Johnston, W. A., and Schwarting, I. S. (1996). Reassessing the evidence for novel popout. *J. Exp. Psychol: Gen* . 125, 208–212. doi: 10.1037/0096-3445.125.2.208

Johnston, W. A., and Schwarting, I. S. (1997). Novel popout: an enigma for conventional theories of attention. *J. Exp. Psychol. Hum. Percept. Perform*. 23, 622–631. doi: 10.1037//0096-1523.23.3.622

Kaan, E., and Grüter, T. (2021). *Prediction in Second Language Processing and Learning*. Amsterdam: John Benjamins. doi: 10.1075/bpa.12

Kalmar, T. M. (2015). *Illegal Alphabets and Adult Biliteracy: Latino Migrants Crossing the Linguistic Border, 2nd Edn*. New York, NY: Routledge.

Keating, G. D., and Jegerski, J. (2015). Experimental designs in sentence processing research. *Stud. Sec. Lang. Acquisit*. 37, 1–32. doi: 10.1017/S0272263114000187

Key, M. (2014). Positive expectation in the processing of allophones. *J. Acoust. Soc. Am*. 135, EL350–EL356. doi: 10.1121/1.4879669

Koda, K. (1998). The role of phonemic awareness in second language reading. *Sec. Lang. Res*. 14, 194–215. doi: 10.1191/026765898676398460

Kumle, L., Võ, M. L.-H., and Draschkow, D. (2021). Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R. *Behav. Res. Methods* 53, 2528–2543. doi: 10.3758/s13428-021-01546-0

Lahiri, A., and Reetz, H. (2010). Distinctive features: phonological underspecification in representation and processing. *J. Phonet*. 38, 44–59. doi: 10.1016/j.wocn.2010.01.002

Lindsey, J. (2013). *L2 Acquisition of the Context-Dependent [ç]-[x] Alternation in American English Learners of German* (Unpublished undergraduate thesis). Indiana University, Bloomington, IN, United States.

Lipski, S. C. (2006). *Neural Correlates of Fricative Contrasts Across Language Boundaries* (Dissertation). University of Stuttgart, Stuttgart, Germany. Availabe at: http://elib.uni-stuttgart.de/

Lydersen, S. (2019). Statistisk styrke - før, men ikke etter! [Statistical power: Before, but not after!] *Tidsskrift for Den norske legeforening* 139:847. doi: 10.4045/tidsskr.18.0847

Mann, V. A. (1986). Phonological awareness: the role of reading experience. *Cognition* 24, 65–92. doi: 10.1016/0010-0277(86)90005-3

Marslen-Wilson, W., and Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychol. Rev*. 101, 653–675. doi: 10.1037/0033-295X.101.4.653

Martin, J. G., and Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects. *J. Acoust. Soc. Am*. 69, 559–567. doi: 10.1121/1.385484

Martin, J. G., and Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *J. Exp. Psychol. Hum. Percept. Perform.* 8, 473–488. doi: 10.1037//0096-1523.8.3.473

Mathôt, S., Schreij, D., and Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences (Version 2.8.0). *Behav. Res. Methods* 44, 314–324. doi: 10.3758/s13428-011-0168-7

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., and Bates, D. (2017). Balancing Type I error and power in linear mixed models. *J. Mem. Lang.* 94, 305–315. doi: 10.1016/j.jml.2017.01.001

McMurray, B., and Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychol. Rev.* 118, 219–246. doi: 10.1037/a0022325

Mills, C. B. (1980). Effects of context on reaction time to phonemes. *J. Verb. Learn. Verb. Behav.* 19, 75–83. doi: 10.1016/S0022-5371(80)90536-8

Morais, J., Bertelson, P., Cary, L., and Alegria, J. (1986). Literacy training and speech segmentation. *Cognition* 24, 45–64. doi: 10.1016/0010-0277(86)90004-1

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *J. Acoust. Soc. Am.* 99, 1718–1725. doi: 10.1121/1.414696

O'Hara, R. B., and Kotze, J. (2010). Do not log-transform count data. *Methods Ecol. Evol.* 1, 118–122. doi: 10.1111/j.2041-210X.2010.00021.x

Otake, T., Yoneyama, K., Cutler, A., and Van der Lugt, A. (1996). The representation of Japanese moraic nasals. *J. Acoust. Soc. Am.* 100, 3831–3842. doi: 10.1121/1.417239

Plag, I., Braun, M., Lappe, S., and Schramm, M. (2009). *Introduction to English Linguistics, 2nd Edn*. Berlin: Mouton de Gruyter.

Posner, M., Nissen, M., and Ogden, W. (1978). "Attended and unattended processing modes: the role of set for spatial location," in *Modes of Perceiving and Processing Information*, eds. H. L. Pick, Jr., and E. Saltzman (Hillsdale, NJ: Lawrence Erlbaum Associates), 137–157.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Lang. Learn.* 63, 595–626. doi: 10.1111/lang.12010

Robinson, O. W. (2001). *Whose German? The Ach/Ich Alternation and Related Phenomena in "Standard" and "Colloquial" [ProQuest Ebook Central Version]*. doi: 10.1075/cilt.208

Scott, J. H. G. (2019a). *Phonemic and Phonotactic Inference in Early Interlanguage: Americans Learning German Fricatives in L2 Acquisition* (Dissertation). Indiana University, Bloomington, IN, United States. Available at: https://www.proquest.com/dissertations-theses/phonemic-phonotactic-inference-early/docview/2235974900/se-2 (accessed October 4, 2024).

Scott, J. H. G. (2019b). "Who follows the rules? Differential robustness of phonological principles," in *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, ISSN 2380-9566*, eds. J. Levis, C. Nagle, and E. Todey (Ames, IA: Iowa State University), 213–225. Available at: https://apling.engl.iastate.edu/conferences/pronunciation-in-second-language-learning-and-teaching-conference/psllt-archive/ (access October 4, 2024).

Scott, J. H. G., and Darcy, I. (2023). Prosodic location modulates listeners' perception of novel German sounds. *Lab. Phonol.* 14, 1–42. doi: 10.16995/labphon.6428

Scott, J. H. G., Lim, R. Z. J., and Russell, C. B. (2022). "Sound-spelling correspondences in FL instruction: same script, different rules," in *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference, held June 2021 virtually at Brock University, St. Catharines, ON*, eds. J. Levis and A. Guskaroska. St. Catharines, ON. doi: 10.31274/psllt.13361

Seguí, J., Frauenfelder, U., and Hallé, P. (2001). "Phonotactic constraints shape speech perception: implications for sublexical and lexical processing," in *Language, Brain, and Cognitive Development: Essays in honor of Jacques Mehler*, ed. E. Dupoux (Cambridge, MA: MIT Press), 195–208.

Seguí, J., and Frauenfelder, U. H. (1986). "The effect of lexical constraints on speech perception," in *Human Memory and Cognitive Abilities: Symposium in Memoriam Hermann Ebbinghaus*, eds. F. Klix and H. Hagendorf (Amsterdam, North-Holland), 795–808.

Selkirk, E. (1984). "On the major class features and syllable theory," in *Language Sound Structure: Studies in Phonology Presented to Morris Halle by His Teacher and Students*, eds. M Aronoff, F. Kelley, and B. K. Stephens (Cambridge, MA: MIT Press), 107–136.

Speeter Beddor, P., and Evans-Romaine, D. (1995). Acoustic-perceptual factors in phonological assimilations: a study of syllable-final nasals. *Rivista di Linguistica* 7, 145–174. Available at: http://linguistica.sns.it/RdL/1995.html (accessed October 4, 2024).

Steinberg, J. (2014). *Phonotaktisches Wissen: Zur prä-attentativen Verarbeitung phonotaktischer Illegalität. [Phonotactic knowledge: On the pre-attentive processing of phonotactic illegality]*. Berlin: De Gruyter.

Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010a). Preattentive phonotactic processing as indexed by the Mismatch Negativity. *J. Cogn. Neurosci.* 22, 2174–2185. doi: 10.1162/jocn.2009.21408

Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2010b). Activation and application of an obligatory phonotactic constraint in German during automatic speech processing is revealed by human event-related potentials. *Int. J. Psychophysiol.* 77, 13–20. doi: 10.1016/j.ijpsycho.2010.03.011

Steinberg, J., Truckenbrodt, H., and Jacobsen, T. (2011). Phonotactic constraint violations in German grammar are detected in auditory speech processing: a human event-related potentials study. *Psychophysiology* 48, 1208–1216. doi: 10.1111/j.1469-8986.2011.01200.x

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phonet.* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001

Streeter, L. A., and Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *J. Acoust. Soc. Am.* 65, 1533–1541. doi: 10.1121/1.382917

Suzuki, Y. (2017). Validity of new measures of implicit knowledge: distinguishing implicit knowledge from automatized explicit knowledge. *Appl. Psycholinguist.* 38, 1229–1261. doi: 10.1017/S014271641700011X

Swinney, D. A., and Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: the variable role of uncertainty about vowel contexts. *Percept. Psychophys.* 27, 104–110. doi: 10.3758/BF03204296

Thomson, R. I. (2011). Computer assisted pronunciation training: targeting second language vowel perception improves pronunciation. *CALICO J.* 28, 744–765. doi: 10.11139/cj.28.3.744-765

Thomson, R. I. (2018). *English Accent Coach [Computer Program]. Version 4.0*. Available at: https://www.englishaccentcoach.com/ (accessed September 27, 2024).

Valaczkai, L. (1998). *Atlas deutscher Sprechlaute: Instrumentalphonetische Untersuchung der Realisierung deutscher Phoneme als Sprechlaute [Atlas of German speech sounds: Instrumental phonetic examination of the realization of German phonemes as speech sounds]*. Vienna: Edition Praesens.

Vance, T. J. (1987). *An Introduction to Japanese Phonology*. Albany, NY: SUNY.

Wang, Y. (2023). Processing of English coda laterals in L2 listeners: an eye-tracking study. *Lang. Speech* 2023, 1–15. doi: 10.1177/00238309231203899

Weber, A. (2001a). Help or hindrance: how violation of different assimilation rules affects spoken language processing. *Lang. Speech* 44, 95–118. doi: 10.1177/00238309010440010401

Weber, A. (2001b). *Language-Specific Listening: The Case of Phonetic Sequences*. Wageningen: Ponsen & Looijen.

Weber, A. (2002). Assimilation violation and spoken-language processing: a supplementary report. *Lang. Speech* 45, 37–46. doi: 10.1177/00238309020450010201

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Percept. Psychophys.* 35, 49–64. doi: 10.3758/BF03205924

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Percept. Psychophys.* 50, 351–360. doi: 10.3758/BF03212227

Wiese, R. (1996). *The Phonology of German*. Oxford: Clarendon Press.

Wiese, R. (2011). *Phonetik und Phonologie [Phonetics and Phonology]*. Paderborn: Wilhelm Fink.

Winters, S. J. (2003). *Empirical Investigations Into the Perceptual and Articulatory Origins of Cross-Linguistic Asymmetries in Place Assimilation* (Dissertation). The Ohio State University, Columbus, OH, United States. Available at: https://linguistics.osu.edu/sites/linguistics.osu.edu/files/dissertations/winters2003.pdf (accessed October 4, 2024).

Frontiers in Language Sciences

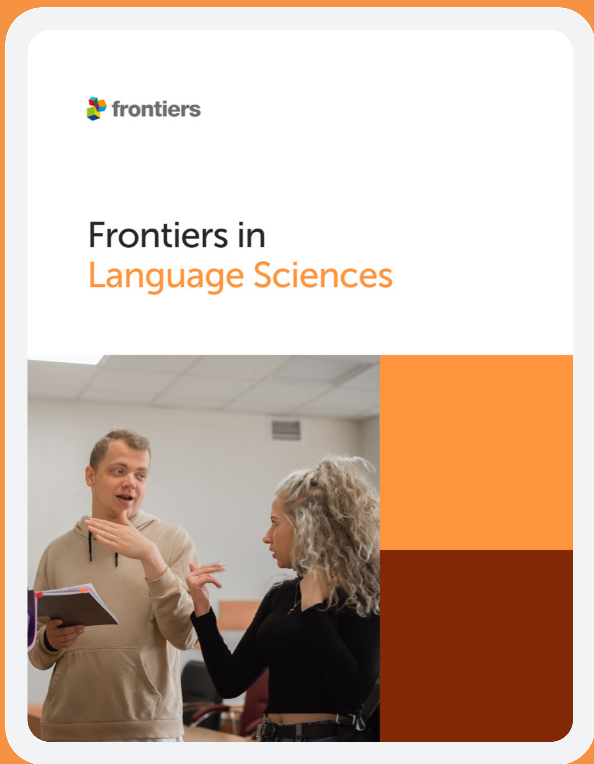
Explores the formal, cognitive and brain mechanisms of human language

A journal dedicated to developing knowledge and understanding of language processing and the ways in which it effects the brain and human development

---

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact