Using a contrastive hierarchy to formalize structural similarity as I-proximity in L3 phonology

John Archibald
University of Victoria

In this paper I argue that cross-linguistic similarity in third language acquisition is determined by a structural hierarchy of contrastive phonological features. Such an approach allows us formalize a predictive notion of I-proximity which also provides an explanatory model of L2, and L3 phonological knowledge (represented in an integrated I-grammar). The metrics of phonological similarity (i.e., structural not acoustic) are analogous to morphosyntactic similarity in that both morphosyntactic and phonological approaches can compare the outcomes of parsing the L3 input by the L1 hierarchy and by the L2 hierarchy. From this starting point I propose a conservative, incremental learning theory to guide subsequent reconstruction of the L3 grammar. Under this model, it can be argued that phonology is part of Faculty of Language Narrow (FLN). The (gradient) phonetic material comes from outside the FLN but the linguistic computational system converts it to discrete abstract elements that can be manipulated by the learner.

Keywords: L3 phonology, I-proximity, Linguistic Proximity Model (LPM), contrastive hierarchy, cross-linguistic similarity

1. Introduction

Models of third language acquisition (L3A) such as Rothman (2015; Westergaard (2021) rely on cross-linguistic comparison when attempting to predict or explain whether the L1 or the L2 transfers into the L3. Rothman's Typological Primacy Model (TPM) states that the 'closest language' will transfer to the L3, while Westergaard's Linguistic Proximity Model (LPM) states that the 'closest structure' will determine the L3 form. For example, in a situation where the L1 was Spanish and the L2 was English and the L3 was French, the TPM would predict that the L3

https://doi.org/10.1075/lab.22051.arc | Published online: 4 July 2023
Linguistic Approaches to Bilingualism ISSN 1879-9264 | E-ISSN 1879-9272

learner would transfer the L1 Spanish grammar in its entirety to become the initial state of the L3 French grammar because Spanish and French are typologically closer than English and French are. The LPM, on the other hand, allows property-by-property transfer in that we might see transfer of the L2 English overt subject structure into L3 French (because neither English nor French allow null subjects), but L1 Spanish transfer of reflexive verb properties (because Spanish and French reflexives are more similar than English and French reflexives).

Much of this work is in morphosyntax (e.g., Rothman, et al., 2019), and studies in L3 phonology are still in the minority (Cabrelli, 2012). Each of these models has different ways of determining proximity, though. In this paper I will explore (a) how to formalize the notion of linguistic proximity in L3 phonology using a Contrastive Hierarchy (CH) model (Dresher, 2009), and (b) principles that might underlie the restructuring of such L2/L3 featural hierarchies. I will demonstrate that the CH model is able to account both for (a) common patterns of cross-linguistic influence found between different L1 groups, and (b) patterns of individual variation found within a single group. The basic claim that I am making is that cross-linguistic comparison needs more than a measure of surface acoustic similarity (i.e. not just phonetics) to succeed. We need to formalize a measure of *phonological* similarity (or I-proximity).

2. Invoking similarity

Many traditional accounts of differential L2 transfer (what Weinberger (1997) called *differential* substitution) have relied on the idea that the 'closest' L1 sound will transfer. Why do speakers of some languages substitute [t] for English /θ/ while speakers of others substitute [s] for English /θ/? Many of these studies have been production studies (Lombardi, 2003) but the phenomenon also affects perception (Hanulíková & Weber, 2012; Hancin-Bhatt, 1994). The general form of the argument is that for Russian speakers [t] is closer to /θ/ while for Japanese speakers [s] is closer to /θ/. Clearly, we need a way of measuring closeness.

2.1 Measures of similarity

There are two broad approaches to measuring similarity: local and global. In the phonetics literature, there are many local ways to compare two sounds.

For vowels, one could measure such things as the first formant (F1) and second formant (F2) and then use Euclidean Distance or Mahalanobis Distance (Kartushina et al., 2015) to see how close the two vowels are. For consonants, one might compare milliseconds of Voice Onset Time (VOT) on the release bursts of stops. Such phonetic measures might determine that, say, the English [u] was closer to the Japanese [u] than the French [u] was, or the French [t] is closer to the Spanish [t] than the English [t] is. In the field of L2 speech, models such as the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) and the L2 Perceptual Assimilation Model (PAM-L2; Best & Tyler, 2007) focus explicitly on local cross-linguistic measures of phonetic similarity. The SLM-r refers to acoustic comparisons while the PAM-L2 refers to gestural comparisons.

There are also measures that are used to compare the cross-linguistic similarity of words. Imagine that we wanted to compare words like English *cat*, French *chat*, and Spanish *gato*. Levenshtein distance (Levenshtein, 1966; Beijering, Gooskens & Heeringa, 2008) is one way to independently determine the proximity of two words by counting the number of changes that must be made to change from orthographic form one to form two.

I refer to these approaches as *local* as they compare elements of language segment by segment. Local comparisons may involve a phonological component as well. Brown (2000) compares English, Chinese, and Japanese liquids with reference to the phonological features found in those languages. Her theoretical assumption is that a language with a phonemic /l r/ contrast would have a [±coronal] feature in the representation of /r/ while /l/ would lack this feature. Closely related to Brown's phonological approach are models which use phonological features to predict proximity based on *classes* of sounds (e.g. LaCharité & Prévost, 1999). Take the example of a language which is argued to have a feature [+voice] to represent voiced stop phonemes. This feature would apply to the segments /b/, /d/, and /g/, and these three segments would be predicted to pattern together.

There have also been global proposals made which attempt to compare whole grammars rather than individual segments or classes of sounds. Going back as far as Chomsky (1965) and re-energized in Yang (2017), it has been recognized that the learner's task is to select a grammar which is consistent with the environmental input recognizing that there may be many possible grammars which could be consistent with the data. An evaluation metric is needed to guide the learner's choice of which grammar to select.

3. Phonological machinery

Space precludes me articulating (if I could) a full model of the phonetic/phonology interface, but let me state some of my broad assumptions. I view aspects of continuous gradience to be properties of the production system (Pierrehumbert, et al., 2000). This could be viewed as the phonetic component in a derivational model, or as the weighted constraint ranking in a harmonic serialism model (Tessier & Jesney, 2014). At the phonetic level we can describe a continuum along dimensions like (a) Voice Onset Time (measured in milliseconds) to describe differences between a Spanish [p] and and English [b], or (b) the first formant (measured in Hertz) to describe the difference between the acoustics of an [i] and an [e].

Such gradient properties can also emerge to enhance underlying cognitive categorical contrasts following the model of phonetic enhancement (Keyser & Stevens, 2006; Hall, 2011). For example, English has two phonemes /s/ and /ʃ/ whose contrast is signalled mainly by a place of articulation difference (alveolar versus post-alveolar). However, often in production the /ʃ/ is produced with lip rounding. Such lip rounding is not contrastive on English consonants; it serves solely to enhance the contrast between the two sibilant fricatives. Surface properties are not always direct cues to the underlying structure. [+round] may not be part of English /ʃ/, and may not be a reliable cue to the learner is deciding how to represent /ʃ/.

I view aspects of categorical representational phenomena to be properties of the phonological system. I adopt a model where phonemes are stored as part of a lexical representation.

3.1 The contrastive hierarchy

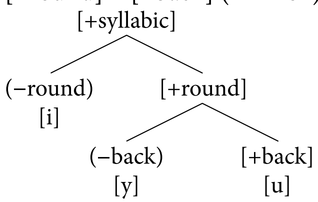
I will demonstrate that Dresher's (2009, 2018) Contrastive Hierarchy (CH) model of phonology is well-suited to formalizing cross-linguistic similarity, and can be used to explain the property-by-property influence witnessed in L3 grammars (Archibald, 2022a, 2022b). The CH has been used to successfully account for L1A (Bohn & Santos, 2018), historical change (Oxford, 2015), and morphosyntax (Cowper & Hall, 2019). Formalizing the notion of structural similarity within this model can provide a unified mechanism of similarity effects across linguistic domains.

Broadly speaking, the CH model is a representational approach where phonological features (e.g [+voice]) are stored in a ranked feature hierarchy which classifies the contrastive sounds of a particular language.[1] Thus, it is a model of contrastive *underspecification* (as non-contrastive information is omitted from the representation). The choice of which features to choose for a particular language is further constrained by the Activity Principle which acknowledges that only fea-

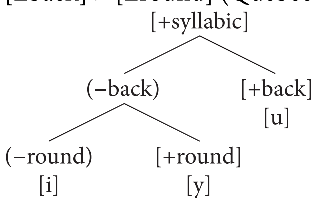
1. Following Dresher (2009), I am adopting a model with binary (i.e., plus or minus) feature values though other others (e.g., Purnell, Raimy & Salmons, 2019) adopt a privative (i.e., present or absent) convention. Nothing crucial hinges on this assumption.

tures which are involved in the computational component of the phonology (i.e., allophony) can be chosen as contrastive features. In this way, the learning of the CH is dependent on the properties of the input. Consider the example (from Cowper & Hall, 2019) given in (1).

1. a. [±round] > [±back] (Finnish)



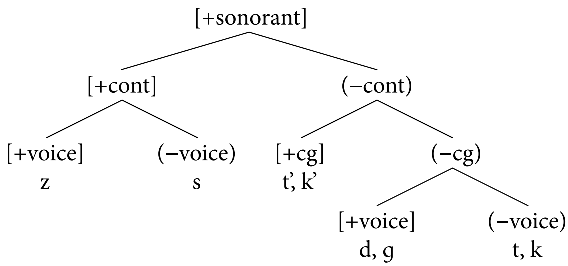
b. [±back] > [±round] (Quebec French)



We see two languages which share the same surface phonetic segments (i/y/u). Note, however, that the phonological representation underlying this 3-vowel system is different in (a) and (b). Note that identical phonetic inventories can be represented – and thus explained –- by different phonological hierarchies. In (a) the [u] is represented by the features [round] > [back] while in (b) the [u] feature has no [round] feature. The implication is that the language in (a) could have a rule spreading [round] from /u/, for example, while the language in (b) could not. Thus, a CH is an explanatory account of differing phonological processes and phonemic inventories.

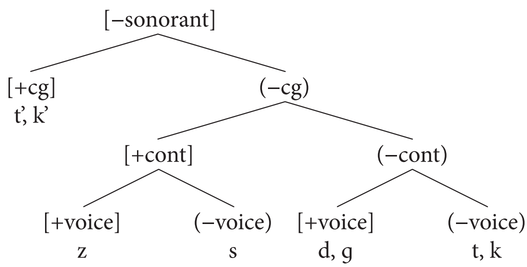
Mackenzie (2011) illustrates this nicely by demonstrating how dissimilar features can block feature spreading. Chaha (a language spoken in Ethiopia) exhibits harmony of oral stops (e.g., [wɨdək']→ [wɨt'ək']). Here the [d] turns into a [t'] because stops in a morphological root have to agree in with respect to the [constricted glottis] feature, so the root surfaces with a [t'] and a [k'], which is well-formed. Fricatives (even though they have phonetic voicing contrasts like [s]/[z]), though, do not participate in the harmony (so a form like /sɨgd/ is well-formed even though it contains both /s/ and /g/). Mackenzie's analysis that the stops ([d, g, t, k]) have a [constricted glottis] feature while fricatives have only a [voice] feature and no [constricted glottis] feature, as shown in (2).

2.



Note that if the ranking of the features was changed, the harmony phenomenon (which bars segments with distinct [constricted glottis] representations) could not be explained, as the fricatives would be (−constricted glottis), as shown in (3).

(3)



Mackenzie (2011: 1401) argues that "phonological similarity is evaluated over contrastive, phonological features," and that consonant harmony patterns support "a phonological notion of similarity based on contrastive feature specifications." Such an analysis shows that both the features *and* the ranking are critical.

3.2 Contrastive hierarchy theory and optimality theory

It should also be noted that while a CH is representational and invokes representational principles such as underspecification, it is translatable into the machinery of Optimality Theory as shown in Mackenzie (2013). Again, space precludes me from demonstrating how such an algorithm works. However, as Hall (p.c.) points out, OT constraints *refer* to representations by encoding notions such as feature, mora, syllable, etc. Thus, while I state my claims in representational terms rather than Optimality constraints, the insights into the nature of learner grammar are not lessened.

Such an assumption (of underlying phonemic representation) can be viewed as being at odds with Optimality Theoretic machinery if we adopt the assumption of Richness of the Base (ROTB) where input representations cannot be constrained language-specifically and contrastive segments emerge as the result of constraint interaction. However, there are OT models which do not assume

ROTB. Nevins & Vaux (2007) argue that learners construct abstract representations which are not identical to surface forms. As Hall (2007) notes, the design challenge is to exclude certain information from underlying phonological representations. For example, we may wish to adopt a model where English underlying representations would include an /r/l/ contrast but Japanese underlying representations would not. This is not compatible with ROTB which states that "the constraint grammar of a language should produce phonotactically well-formed outputs for all conceivable inputs, including those which are not – and could not be – present in that language's lexicon." Rasin & Katzir (2016) propose a version of OT (Minimum Description Length) which is incompatible with ROTB and employs language-specific constraints on underlying representations. Hall (2007) also notes that Lexical Optimization ('which Prince & Smolensky (1993: § 9.3) propose as a procedure for selecting a single underlying form from among several that yield the same output') is similarly problematic. The default assumption is that the underlying form should mirror the surface form. He notes that 'one consequence of this is that it is difficult for underlying representations to be underspecified for any features that are present in their corresponding surface forms: the filling in of unspecified feature values introduces a putatively unnecessary mismatch between input and output.' Tessier & Jesney (2014) also reject the Identity Map in Harmonic Serialism models. And yet the notion of underspecification seems to be well-motivated and empirically essential.

So, while I am not adopting an OT framework for my proposals of assessing phonological similarity, I would argue that there are ways in which this could be done. That is to say, if one wanted to recast the proposals I am making in OT terms, it could be done if one rejects the stipulations of ROTB and Lexical Optimization. The result would be to stipulate information in the phonemic inventory (see Hall (2007) for further discussion of how the generative power of such a theory can be constrained by the adoption of such tools as the Successive Division Algorithm of Dresher (2009)).

4. Multilingual acquisition

I will show the utility of this CH model to account for segmental aspects of phonological L2/L3A.[2] A key element is to recognize that learner behaviour is

2. As a model of phonological grammar, CH makes predictions about both multilingual perception and production. Perception may well proceed in advance of production (given the additional motoric demands of production, see Patience & Qian, 2022). Of course, neither perception tasks nor production tasks are direct windows onto grammar.

best viewed at the inventory level rather than by a segment-by-segment comparison. A clear example of an inventory effect is presented in Munro & Derwing's (2008) description of L1 Mandarin acquisition of L2 English vowels as shown by the intelligibility accuracy scores[3] in Table 1.

Table 1. L2 English vowel intelligibility scores (from Munro & Derwing, 2008)

Long (tense) vowels	Short* (lax) vowels
[i]: 97%	[ɪ]: 55%
[e]: 88%	[ɛ]: 57%
[ɑ]: 80%	[æ] 76%
[o]: 85%	[ʌ]: 75%
[u]: 78%	[ʊ]: 70%

* In this paper I will tend to use the terms long/short to refer to this vowel contrast. Others use the terms tense/lax. I will use these pairs of terms synonymously.

This chart reveals that the Mandarin subjects have difficulty with a number of L2 vowels. If we viewed this as having 5 different causes then we would be missing the fact that all of the problematic vowels shared a phonological feature: [−long]. A CH recognizes that phonological knowledge captures *contrast*, and thus when we compare two (or more) languages, we need to look at the contrastive system (not merely individual segments).[4]

4.1 Phonological similarity

The notion of similarity plays a central role in two key models of L3A. Rothman's (2015) TPM predicts that either the L1 or the L2 (whichever is most similar) will transfer in its entirety to become the basis of the L3 grammar (see also Schwartz & Sprouse, 2021). Westergaard's LPM predicts that both L1 and L2 structures can

3. In a delayed repetition task L1 Mandarin adults living in Canada produced 20 English CVC words (which all ended in voiceless consonants, and, thus, contain vowels of short duration). These pronunciations were listened to by 4 judges who clicked a button to identify the vowel in the word spoken. These scores reflect accuracy over all judges (e.g., a speaker says *bit* and the judges click [ɪ] would be labelled intelligible while if they clicked [i] it would be labelled unintelligible. I have reported average scores across tokens after one year of full-time ESL study in Canada in Table 1.

4. Of course, there will be additional (likely phonetic) factors that can lead to intraclass variation (e.g. why the performance on [i/ɪ] is quite distinct while [ɑ/æ]) is closer, but my point is that we need to consider phonological class as well as phonetic properties.

be activated in the L3 grammar depending on which is most similar (L1/L3 or L2/L3). Both models crucially require an independent way of measuring similarity. In exploring L3 phonological data, I will show how a CH approach can give us just such a metric.

4.2 Previous approaches

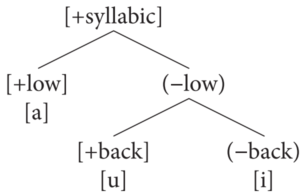
Kwon (2021) provides an empirical study which draws on the Featurally-Underspecified Lexicon (FUL) model of Lahiri & Reetz (2002) to demonstrate how L1 phonological representations account for L2 perceptual similarity judgments in L2 vowel perception. While there have been some studies which have made reference to investigating L3 phonology (Chen & Han, 2019; Chen & Tian, 2021; Llama & Cardoso, 2018), my reading of these studies is that they would be more appropriately labelled L3 phonetics because they deal primarily with gradient phenomena such as VOT rather than the categorical phenomena such as the features underlying the phonetic production.

I acknowledge an anonymous reviewer's comment that there is no consensus view in the field as to the division of labour between phonetics and phonology. Positions range from the detailed phonetic representations of Exemplar Theory (Johnson, 2006) to the Substance-Free representations of Reiss & Volenec (2022). In Section 6.0, I argue that phonetic variation is an important input cue to the learner though not directly encoded in the phonological representations. The interface architecture (see Archibald (in press)) I assume is that of Natvig & Salmons (2021).

4.3 Property-by-property transfer in L3 phonology

Archibald (2022a, 2022b) reanalyzed Benrabah's (1991) data -- which showed that balanced, adult Algerian Arabic/French sequential bilinguals used their French vowels (e.g., French [œ] for English [ʌ]) but Arabic consonants (e.g., pharyngealized stops; [ɦ] for [h]; [ɾ] for [ɹ]; and [t] for [θ]) in a spontaneous L3 English production task– using a Contrastive Hierarchy model. For reasons of space, I will focus on vowels here, though see Archibald (2022b, forthcoming) for a discussion which addresses consonants; the logic of the argument is the same. The Arabic vowel ranking I would propose is shown in (4).

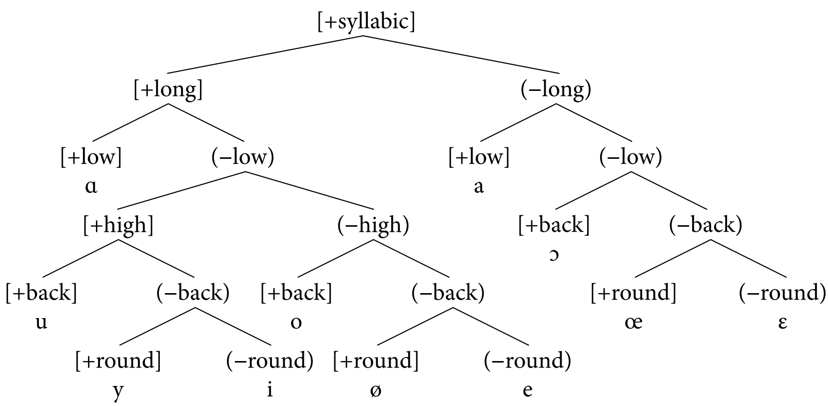
4.



I am adopting a binary feature model where there are markedness relationships built into the representations. The unmarked (−) value is presented in parentheses while the marked [+] value is presented in square brackets. This is consistent with Dresher (2009), Hall (2017), and Natvig & Salmons (2021). Interestingly, there is a growing body of neurolinguistic literature to support such markedness-based approaches to phonological representations (Cummings et al. 2021; Hestvik & Durvasula, 2016; Rhodes et al. 2022).

Turning to another of the languages we are discussing here, the French vowel ranking (based on Hall, 2017) is shown in (5).

5.



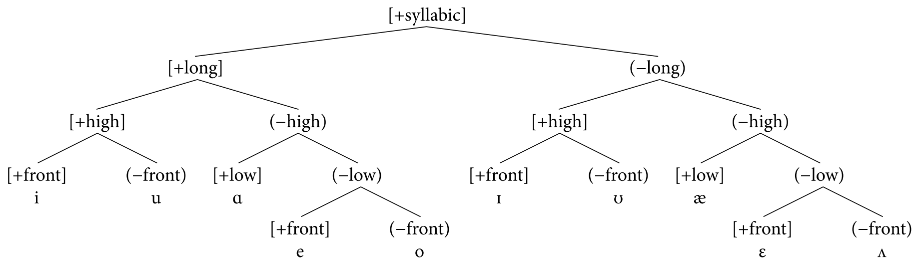
Such a representational system then allows us to compare L1, L2, and L3 vowels to explain why the participants used French vowels for L3 English but Arabic consonants. Let us compare the relevant vocalic feature rankings for the three languages, as shown in Table 2.

Table 2. Feature rankings of Arabic, French, and English

Arabic (Archibald, 2022a)	[low] > [back]
French (Hall, 2017)	[nasal] > [long] > [low] > [high] > [back] > [round]
English (Gardner & Roeder, 2022)	[long] > [high] > [low] > [front]

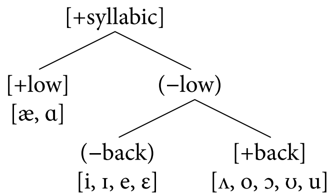
Before comparing the Arabic and French parses of English, let us consider the CH of the English vocalic inventory, shown in (6).

6.



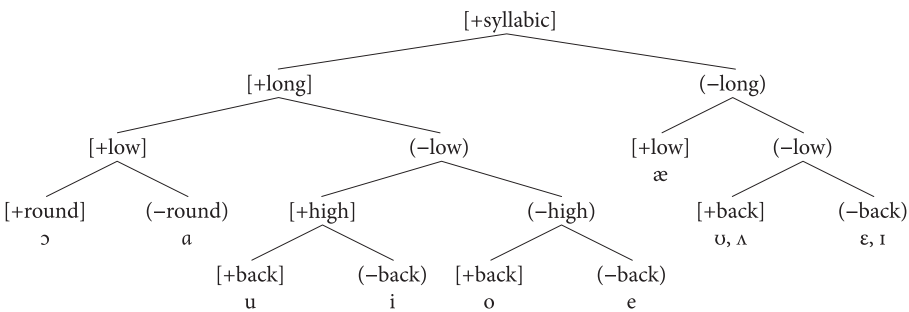
As Archibald (2022b) showed, using the Arabic vocalic features to parse the L2 English input does not allow the learner to uniquely parse all the L3 phonemes (eight English vowels would be ambiguous), as shown in (7).

7.



However, using the French features can uniquely parse all but two vowels in the L3 English input, as shown in (8).

8.



Such a parsing comparison would support the choice of the learner to adopt French vocalic features as being more similar to (i.e., better able to parse) the English input.

4.4 Restructuring a contrastive hierarchy

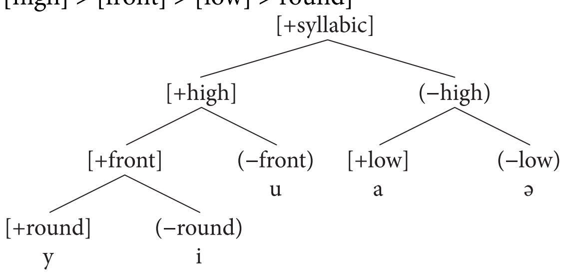
I view the L3 grammar as being constructed and represented from the existing components of the integrated I-grammar (see López, 2020). My position is that within the multilingual integrated I-grammar there are representational means of assembling individual languages. Someone who knows Arabic, French, and

English is clearly able to suppress, say, English and French lexical items when speaking Arabic, so the languages must be identifiable. Automatic spreading activation is not so easy to suppress in comprehension (Dijkstra, Grainger, & Van Heuven, 1999) but the basic production facts demonstrate that multilinguals can reference such constructs as *Arabic vowels* or *French consonants* and suppress them. Perhaps the mechanism to achieve such a result is language tagging (Green, 1998) in which particular structures in the integrated repository are tagged for which language they belong to. Thus, the generation of L3 utterances/structures can proceed on the basis of consulting the components of the repository. Once a decision has been made then structures will start to be tagged as L3 in the repository (see Archibald, 2022a) too. In acquisition, though, grammars (built as a result of parsing) need to be revised or restructured.

5. Principles of restructuring

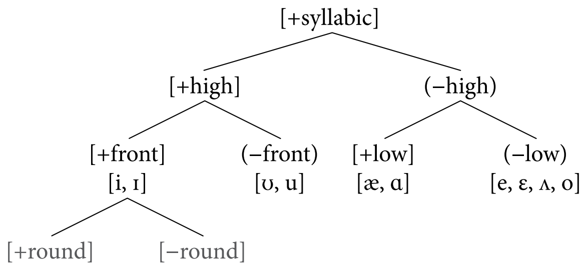
This section will draw on a case study presented by Wu (2021) analyzing the CH of the Mandarin phonemic vowel inventory. She proposed the structure given in (9) to account for both the contrastive phonemes and the common phonological processes of Mandarin.

9. Mandarin vowel Contrastive Hierarchy.
   [high] > [front] > [low] > round]



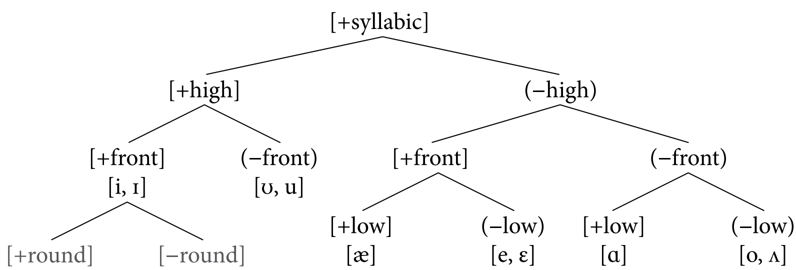
Under this analysis, Mandarin has five contrastive vowel phonemes. There are also productive phonological processes in Mandarin (Duanmu, 2007) where (i) /ə/ fronts to [e] in the environment of the [+front] vowels [i] and [y], and (ii) /ə/ backs to [o] in the environment of the (−front) vowel [u]. English has more than five phonemic vowels, and, thus, a more complex Contrastive Hierarchy, as we saw in (6). From a learnability perspective, we have to ask: what actions do the Mandarin learners have to take in order to be able to parse the English vocalic input? Let us first note in (10), how the Mandarin CH could parse the English[5] input.

10.



First of all, note that the lax/short vowels (ɪ, ʊ, ɛ, æ, ʌ) cannot be distinguished from the tense/long vowels. Note also that the feature [round] is not needed to uniquely parse the English high, front vowels. What restructuring actions could the Mandarin learners take? One step would be to take the [front] feature used for [+high] vowels and use it for the [−high] vowels. As we saw in (9), the ranking of features in Mandarin has [front] above [low]. Assuming that CH rankings transfer, this would produce the structure shown in (11).

11.

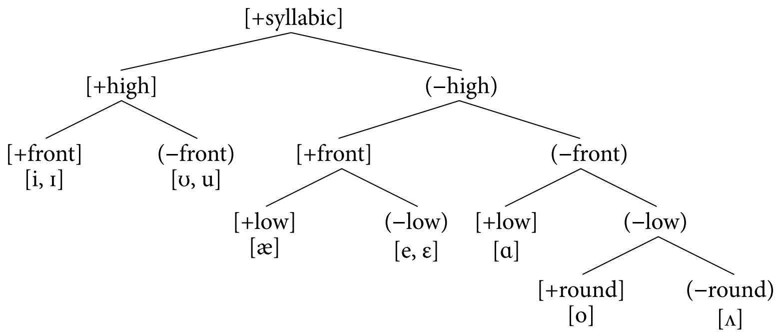


Such a change would now allow the learner to contrast /æ/ from /ɑ/ and /e, ɛ/ from /o, ʌ/.

Another step would be to redeploy the [round] feature (which is used for [+high] vowels in Mandarin) to introduce new L2 contrasts for (−high) vowels as shown in (12).

5. In this section both for reasons of space and because Wu's studies involve Canadian English, I am not including /ɔ/. The restructuring principles would remain the same if the target dialect had an /ɔ/~/ ɑ/ contrast.

12.



Invoking two such local changes would restructure the Mandarin CH to uniquely parse all but the vowels [ɪ/i, ɛ/e, ʊ/u].

Clearly, the grammar would have to be restructured in a more major way to become targetlike. Such an account is actually in line with the empirical investigations (Jin & Liu, 2014; Yuan & Archibald, 2022) which show that Mandarin learners of English have difficulty perceiving and producing lax/short vowel contrasts in English. According to Oxford (2015); Purnell et al. (2019) and Gardner & Roeder (2022) the top-ranked English feature is [long]. Mandarin learners of English would have to acquire a new, highly-ranked feature.

5.1 Acquiring a new highly-ranked feature

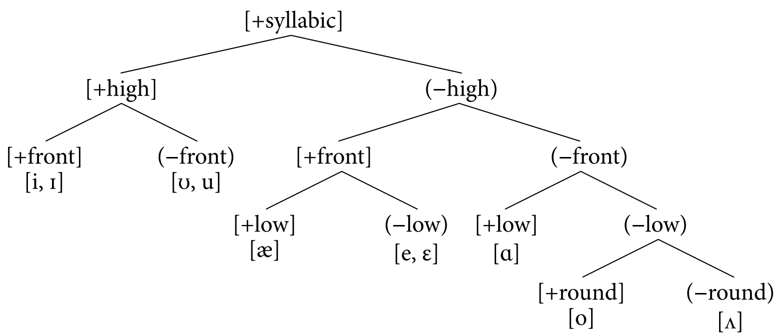
The question is: what steps would the learners posit in arriving at a ranking with the [long] feature at the top? Consistent with the general approach of learnability (Wexler & Culicover, 1983), I assume a conservative learner; one who makes small incremental changes based on positive evidence in the input, and parsing failures. Practically speaking, this means that restructuring will start at the bottom of a given tree to avoid overgeneralization and the need of potential backtracking to undo decisions. This is what Oxford (2015) referred to as the *sisterhood* condition where the most conservative change to a structure is to change the sister[6] of an existing node. If more changes are needed, then the learner would restructure by moving a feature one level higher in the tree until the structure could successfully parse the input data. Intermediate structures must be licensed by Universal Grammar. This predicts then, that if a given feature is low-ranked in the L1 but high-ranked in the L2 that it will be incrementally promoted until it reaches the top level.

6. Oxford (2015: 316) defines sister as "any two nodes that are immediately dominated by the same node."

5.2 Back to Mandarin

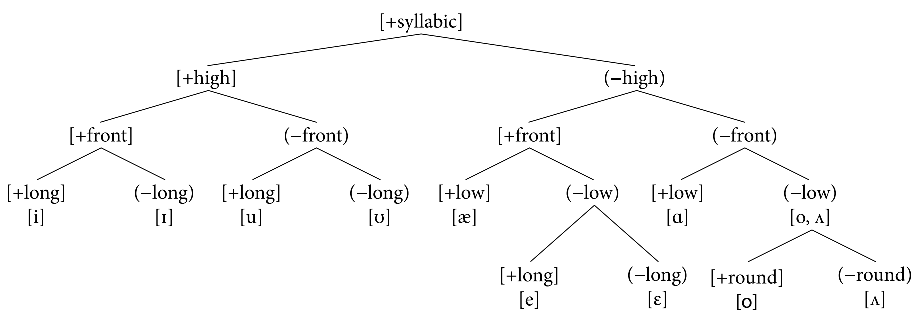
Let us assume, for the sake of argument that the Mandarin learners of English complete the redeployment steps outlined above, and have landed on the representation shown in (13).

13.



At this point they have yet to acquire the [long] feature. The first step would be to add the new feature [±long] dominating any pair of vowels which have yet to be uniquely parsed as shown in (14).

14.



Note that by beginning the restructuring at the bottom of the tree, the change is minimal. If it turned out that the new grammar was not better able to parse the input the retreat would be minimal. Note that the restructuring in (12) was motivated by a transfer of L1 CH rankings while the restructuring in (13) results from the triggering of a new feature not found in the L1. Contrast this with a learner who, in an attempt to disambiguate [e, ɛ], [u, ʊ], and [i, ɪ] made a change by adding [±long] at the top of the tree. This would require a major restructuring of the contrastive hierarchy. A change at the top would require *every* vowel to be specified [±long] while the restructuring in (12) would only require undifferentiated vowels to be specific in this way. Therefore, I assume that learning starts at the bottom and proceeds incrementally. For learnability reasons, we prefer the latter but the question requires empirical investigation.

5.3 Individual variation

One of the advantages of the CH approach to phonemic inventories is that it accommodates individual variation in developmental path easily. Munro (2018) looks at Mandarin learners' production of English vowels [i]/[ɪ] and [u]/[ʊ], while Munro (2021) looks at Cantonese production of the same vowels. In an intelligibility paradigm both studies note considerable individual variation in the intelligibility of the vowels in question. Some learners may do better on the front vowels while others do better on the back vowels. Our previous discussion can provide an explanation. Given that there were three contrasts to be acquired, three ambiguous pairs to be distinguished, learners may elaborate the nodes in different orders. There is nothing contingent on elaborating the [i]/[ɪ] node with elaborating the [ʊ, u] node or vice versa. And elaborating the [e, ɛ] node is also a separate decision.

This is, after all, what we see in L1A. Research such as Fikkert (1994); Rice & Avery (1995), up to Bohn & Santos (2018) has shown that children often take different routes to arrive at the same final representational destination. However, what such studies in individual variation also reveal is that the learners are conservative, and proceed incrementally so as to avoid having to retreat from an unjustified overgeneralization. It is unsurprising that we see the same in L2A and L3A. While there is a rich literature in individual variation in SLA (Dornyei & Ryan, 2015), little of this research has connected explicitly with a representational model of generative grammar which is the main reason for its sparse coverage in this paper. I would also say that the Contrastive Hierarchy approach is valuable in being able to account for such individual variation. Munro (2018) questions the utility of linguistic theory in accounting for pronunciation errors largely because not all speakers of the same L1 evidence the same profiles. What the CH approach recognizes is that different learners may take different paths and yet the choice of paths is still highly constrained.

Consider the following scenario. The Mandarin vowel inventory is given in (9). The Mandarin CH cannot uniquely parse the English vowel phonemes as shown in (10). Further changes still need to be made to uniquely parse all the English vowels, and this is where individual variation comes into play. Let us assume that there are two possible actions the learner can take: (a) redeploy [±front] from the [+high] branch and apply it to the (−high) branch, and (b) trigger a new feature [±long] for all nodes with unparsed segments. Some learners may choose to implement (a) before (b), others the reverse. Even within (b) there is the possibility of individual variation as the new feature could be triggered independently under each of the four terminal nodes in (10). Of course, we would like to propose a learning theory which explains the developmental paths taken and why, but that goes beyond the scope of this paper.

There is still much to be explored in applying the CH to multilingual phonology. Questions such as degrees of difficulty in adding new features as opposed to culling old features need to be tackled. The role of semantics in triggering the features which allow new minimal pairs has been under-addressed. Future research will have much to report on. Archibald (2022a, 2022b) makes an argument consistent with Westergaard's (2021) notion of property-by-property cross-linguistic influence. Benrabah (1991) showed that L3 vowels and consonants had distinct linguistic sources, while Ghazali and Bouchhioua (2003) and Bouchhioua (2016, 2017) showed that stress and rhythm could have different sources as well; *both* the L1 and the L2 can influence the L3. This raises the empirical question of how much language mixing (i.e., L1 & L2 influence on the L3) within and across linguistic domains can be observed. Drawing on the codeswitching literature (e.g., Stefanich et al., 2019) our starting position would be to note that mixing can be very prolific. The literature documents that individuals can change languages within conversations, within sentences, and within words. As we have seen in the L3 data, 'switches'[7] for consonants and vowels, for stress and rhythm have been reported, but what about finer-grained distinctions? Could someone produce L1 phonology in nouns but L2 phonology in verbs? A similar situation does arise in Michif (Pappen, 2003). Could we find L1 phonology in *some* consonants (e.g., sonorants) but L2 phonology in other consonants (e.g., obstruents)? These are empirical questions that still need to be explored as we seek understanding of the nature of the multilingual I-grammar.

6. Learning features

Up until now I have mainly been focusing on how hierarchies are restructured by the learner. Let us turn to explore the question of how features can be learned. I have been assuming binary features. Others (e.g., Purnell, Raimy & Salmons, 2019) assume privative features but what follows could be implemented in either privative or binary feature models. In this section, I want to explore briefly how new feature values could be learned. Cowper & Hall (2014) present some interesting phonetic data (from Abou Haider, 1994), shown in Figure 1, on the production of the three Arabic vowels which is highly relevant to the acquisition questions we

7. I put this in 'scare' quotes to acknowledge that following the integrated-I-grammar approach of López *switch* implies a separationist architecture but it, nonetheless, captures the surface facts.

have been exploring. Each vowel symbol represents the mean value for each of the eight male speakers Abou Haider recorded in a word reading task consisting of 232 words.

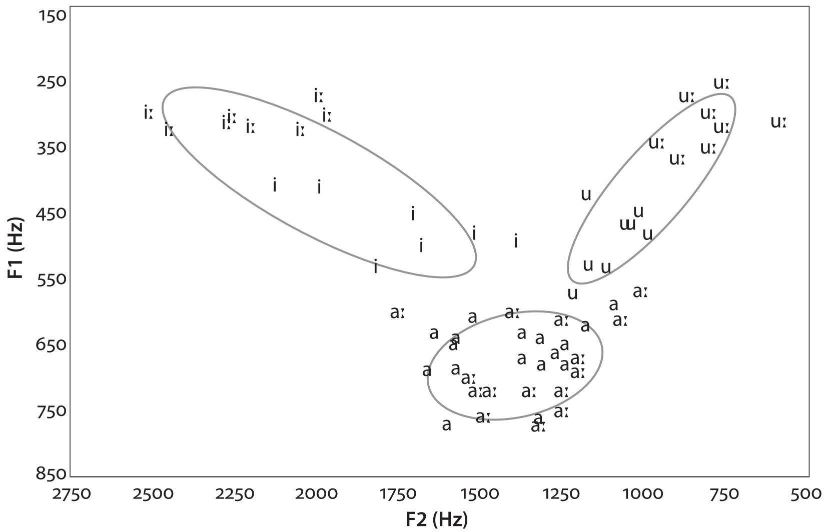


Figure 1. Variation in Standard Arabic vowels (Hall, 2011; data from Abou Haidar, 1994)

In order to understand why such phonetic variation is of interest, we need to note the claims of Purnell (2022) and Natvig & Salmons (2021) who argue that variation is a cue to the *unmarked* feature.[8] That is to say, we see much more variation in the phonetic implementation of an unmarked feature compared to the implementation of a marked feature. If we look at the height (F1) variation on the [+low] vowel, it appears to be about 100 Hz while the variation on the [−low] vowels seems to be about 300 Hz. Therefore, there is greater variation on the high vowels than on the low vowel. This variation in the input would provide a cue to the learner that [+low] is the marked feature and (−low) is the unmarked feature in Arabic. If we look at the backness variation, the [+back] vowel ranges along F2 by about 500 Hz while the [−back] vowel varies by about 1000 Hz. This variation would provide a cue to the learner that [+back] is the marked feature and (−back) is the unmarked feature in Arabic.

Extrapolating beyond the discussion in either Purnell (2022) or Natvig & Salmons (2021), I would suggest that variation might also provide the learner with a cue to the rankings of the features. The F1 variation (height) is about 300 Hz

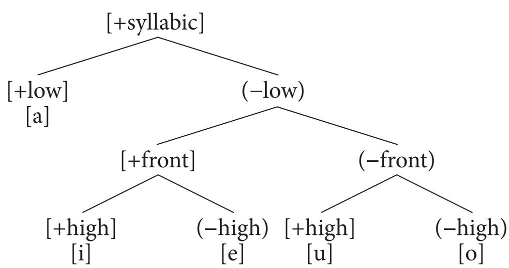
8. Broadly speaking, the theory of markedness attempts to account for what is frequent or natural (unmarked) in a language versus what is infrequent or unusual (marked).

while the F2 variation (backness) is about 700 Hz. This might suggest to the learner that [low] (with less variation) is ranked above [back] (with more variation). This is a proposal which would require further empirical investigation, and which would need to incorporate perceptual models of pitch more sophisticated than simple comparison of Hertz (e.g. semitone scales).

The basic insight from these approaches is that there is a connection between phonetic variation and phonological structure. Labov, Ash & Boberg (2006) had noted that back vowels (with the unmarked value of (−front) tend to vary along the front/back dimension, but front vowels (with the marked value of [+front]) tend not to move back in the process of language change. Similarly, non-high vowels (the unmarked value of (−high)) will vary along the high/low dimension, but high vowels (the marked value of [+high]) won't lower in the process of language change. For Labov these were primitives but not for Purnell; they are motivated by the markedness of phonological features.

Purnell shows us that the implementation of a contrastive hierarchy results in asymmetric phonetic variation with the the unmarked phonological features triggering more phonetic variation. Consider a vowel hierarchy (adapted from Purnell, 2022) as shown in (15).

15.



The empirical claim is that there would be more variation in the production of the [e] and [o] vowels compared to the [i] and [u] vowels. Turning that around, such variation is what characterizes the input to the learner. I think it is uncontroversial to assume the learner detects such variation; if the production system is built to generate these cues, then the perception system is built to reverse engineer the cues. What this means is that as the learner analyzes the phonetic variation in the input, a decision can be made as to the representation of the feature values.

Such a procedure allows the feature values to emerge from an analysis of the input data. What is innate is not a universal inventory of possible features but rather a learning algorithm to seek out and represent *contrast*. In this sense, we can harmonize the representations and processes need to account for L1A, L2A, and L3A (and indeed historical linguistics and language variation).

Under this model, we can argue (following Dresher, 2018) that phonology is part of Faculty of Language Narrow (Fitch, Hauser & Chomsky, 2005). The (gradient) phonetic material comes from outside the FLN but the linguistic computational system converts it to discrete abstract elements that can be manipulated.

7. Conclusion

Cross-linguistic similarity is determined by a structural hierarchy of contrastive phonological features. In this paper, I have focused on one sub-domain of phonological knowledge: vocalic features. While technically, such a model could be used to compare whole languages in an attempt to predict which language would transfer in a TPM approach to L3A, it seems to me that the complexity revealed in our analysis of vowel features suggests that if we assembled feature hierarchies for vowels, consonants, and stress or tone, it might be, to say the least, challenging to arrive at a single measure of similarity which would inform the Big Decision (to use Schwartz & Sprouse's, 2021, phrase) of what should transfer to the L3. I feel that Contrastive Hierarchy Theory is well-suited to predicting and explaining the property-by-property cross-linguistic influence we have empirically observed in L3A (Archibald, 2022a, 2022b). Such an approach allows us formalize a predictive and explanatory notion of I-proximity. Dresher's Contrastive Hierarchy provides us with an explanatory model of L1, L2, and L3 phonological knowledge which is represented in an integrated I-grammar. The metrics of phonological similarity (i.e., structural not acoustic) appear to be analogous to morphosyntactic similarity (Jensen et al., 2021). Both morphosyntactic and phonological approaches can compare the outcomes of parsing the L3 input by the L1 hierarchy and by the L2 hierarchy. However, this is the *starting* point (see Archibald, 2021); we still need a learning theory to guide subsequent reconstruction of the L3 grammar. I have attempted to sketch out some preliminary thoughts on how such a process might proceed. I think there is more to be explored in understanding the role of phonologically *active* features in acquisition. Raimy (pc) suggests that activity 'culls the hypothesis space' in that learners would only consider selecting features to be triggered/re-ranked in the L2 or L3 if there was evidence in the input that they were active in the phonology.

Formalizing a measure of cross-linguistic similarity is important in many linguistic domains, and helps to address many linguistic questions. Work within a Contrastive Hierarchy model shows, I believe, that comparison – in this case phonological comparision – is part of the Learning Theory which learners invoke, not only a task for the linguist.

Funding

Research funded by SSHRCC (435-2022-1097) to John Archibald.

References

Abou Haidar, L. (1994). Norme linguistique et variabilité dialectale: Analyse formantique du système vocalique de la langue arabe. *Revue de Phonétique Appliquée*, *110*, 1–15.

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

Archibald, J. (2022a). Phonological parsing via an integrated I-language: the emergence of property-by-property transfer effects in L3 phonology. *Linguistic Approaches to Bilingualism.*

Archibald, J. (2022b). Segmental and prosodic evidence for property-by-property phonological transfer in L3 English in northern Africa. *Languages 7*(1),28;

Beijering, K., C. Gooskens, & W. Heeringa. (2008). Predicting intelligibility and perceived linguistic distances by means of the Levenshtein algorithm. *Linguistics in the Netherlands*, *1*, 13–24.

Bohn, G. & R. Santos. (2018). The acquisition of pre-tonic vowels in Brazilian Portuguese. *Alfa 62*(1), 191–221.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald, (Ed.), *Second Language Acquisition and Linguistic Theory*. Blackwell. Pp.4–63.

Cabrelli, J. (2012). L3 phonology: An understudied domain. In J. Cabrelli, S. Flynn & J. Rothman, (Eds.). *Third language acquisition in adulthood* (pp. 33–60). John Benjamins.

Chen, H.C. & Q.W. Han. (2019). L3 phonology: contributions of L1 and L2 to L3 pronunciation learning by Hong Kong speakers. *International Journal of Multilingualism 16*(4), 492–512.

Chen, H.C. & J.X. Tian. (2021). The roles of Cantonese speakers' L1 and L2 phonological features in L3 pronunciation acquisition. *International Journal of Multilingualism.*

Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press, Cambridge, MA.

Cowper, E., & D.C. Hall. (2019). Scope variation in contrastive hierarchies of morphosyntactic features. In D. Lightfoot & J. Havenhill (Eds.) *Variable properties in language: Their nature and acquisition* (pp. 27–41). Georgetown University Press.

Cowper, E. & D.C. Hall. (2014). *Reductio ad discriminem*: Where features come from. *Nordlyd 41*(2), 145–164.

Cummings, A., Y. Wu & D. Ogiela. (2021). Phonological underspecification: an explanation for how a rake can become awake. *Frontiers in Human Neuroscience 15*:585817.

Dornyei, Z. & S. Ryan. (2015). *The Psychology of the Langauge Learners Revisited*. Taylor and Francis.

Dresher, B.E. (2009). *The contrastive hierarchy in phonology*. Cambridge University Press.

Dresher, B. E. (2018). Contrastive hierarchy theory and the nature of features. In Wm. G. Bennett, Lindsay Hracs, and Dennis Ryan Storoshenko, (Eds.), *Proceedings of the 35th West Coast Conference on Formal Linguistics*, (pp.18–29). Cascadilla Proceedings Project.

Duanmu, S. (2007). *The phonology of standard Chinese*. Oxford University Press.

Fikkert, P. (1994). *On the acquisition of prosodic stucture*. Leiden.

Fitch, T., M. Hauser, & N. Chomsky. (2005). The evolution of the language faculty: clarifications and implications. *Cognition 97*, 179–210.

Gardner, H. & R. Roeder. (2022). Phonological mergers have systemic phonetics consequences: PALM, trees, and the low back merger shift. *Language Variation and Change*.

Green, D. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition*, *1*(2), 67–81.

Hall, D. C. (2007). Laryngeal underspecification and Richness of the Base. In S. Blaho, P. Bye & M. Krämer, (Eds.) *Freedom of Analysis?* (p. 11–34). De Gruyter Mouton.

Hall, D. C. (2011). Phonological contrast and its phonetic enhancement: Dispersedness without dispersion. *Phonology*, *28*(1), 1–54.

Hancin-Bhatt, B. (1994). Segmental transfer: a consequence of a dynamic system. *Second Language Research 10*(3), 241–269.

Hanulíková, A. & A. Weber. (2012). Sink positive: linguistic experience with th substituitions influences nonnative word recognition. *Attention Perception and Psychophysics 74*, 613–629.

Hestvik, A. & K. Durvasula. (2016). Neurobiological evidence for voicing underspecification in English. *Brain & Language 152*, 28–43.

Jensen, I., N. Mitrofanova, M. Anderssen, Y. Rodina, R. Slabakova, & M. Westergaard. (2022). Crosslinguistic influence in L3 acquisition across linguistic modules. *International Journal of Multilingualism*.

Jin, S. H., & C. Liu. (2014). Intelligibility of American English vowels and consonants spoken by international students in the United States. *Journal of Speech, Language, and Hearing Research*, *57*(2), 583–596.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics 34*(4), 485–499.

Kartushina, N., A. Hervais-Adelman, U. Frauenfelder & N. Golestani. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *Journal of the Acoustical Society of America 138*(2), 817.

Keyser, S. J. & K. Stevens. (2006). Enhancement and overlap in the speech chain. *Language 82*(1), 33–63.

Kwon, J. (2021). Defining perceptual similarity with phonological levels of representations: Feature (mis)match in Korean and English. Ph.D. dissertation. University of Wisconsin-Madison.

Labov, W., S. Ash, & C. Boberg. (2008). *The atlas of North American English: phonetics,phonology and sound change*. De Gruyter Mouton.

LaCharité, D., and P. Prévost. (1999). The role of L1 and of teaching in the acquisition of English sounds by Francophones. In A. Greenhill, H. Littlefield, and C. Tano (Eds.). *Proceedings of BUCLD 23*. (pp. 373–385). Cascadilla Press.

Lahiri, A. & H. Reetz. (2002). Underspecified recognition. In C. Gussenhoven and N. Warner, (Eds.), *Laboratory Phonology 7*. (pp. 637–675). De Gruyter Mouton.

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

doi Llama, R. & W. Cardoso. (2018). Revisiting (non-)native influence in VOT production: insights from advanced L3 Spanish. *Languages*, *3*(3).

doi Lombardi, L. (2003). Second language data and constraints on Manner: explaining substitutions for the English interdentals. *Second Language Research*, *18*(3), 225–250.

doi Mackenzie, S. (2011). Contrast and the evaluation of similarity: evidence from consonant harmony. *Lingua*, *121*, 1401–1423.

doi Munro, M. (2021). On the difficulty of defining 'difficult' in second-language vowel acquisition. *Frontiers in Communication: Language Sciences*, *6*, 639398.

Munro, M. (2018). How well can we predict second language learners; pronunciation difficulties? *CATESOL Journal 30.1*, 267–281.

doi Munro, M. & T. Derwing. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning 58*(3), 479–502.

doi Natvig, D. & J. Salmons. (2021). Connecting structure and variation in sound change. *Cadernos de Lingüistica 2*(1), 1–20.

doi Nevins, A. & B. Vaux. (2007). Underlying representations that do not minimize grammatical violations. In S. Blaho, P. Bye & M. Krämer, (Eds.) *Freedom of Analysis?* (pp. 35–62). De Gruyter Mouton.

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

Papen, R. (2003). Michif: One phonology or two? *University of British Columbia Working Papers in Linguistics (UBCWPL)*, *12*(1), 47–58.

doi Patience, M. & W. Qian. (2022). The role of task complexity and dominant articulatory routines in the acquisition of L3 Spanish. *Languages 7*(2), 90.

Pierrehumbert, J., M. Beckman & R. Ladd. (2000). Conceptual foundations of phonology as a laboratory science. In Burton-Roberts et al., (Eds.). *Phonological Knowledge: Conceptual and Empirical Issues*. (pp 273–304). Oxford University Press.

Purnell, T. (2022). Dialect vowel variation reflects phonology: Evidence from American Dialect Society Publications. Talk at the University of Victoria.

doi Purnell, T., E. Raimy & J. Salmons. (2019). Old English vowels: diachrony, privativity, and phonological representations. *Language 95*(4), e447–e473.

doi Rasin, E. & R. Katzir. (2016). On evaluation metrics in Optimality Theory. *Linguistic Inquiry 47*(2), 235–282.

doi Reiss, C. & V. Volenec. (2022). Conquer primal fear: Phonological features are innate and substance free. *Canadian Journal of Linguistics*.

doi Rhodes, R., E. Avcu, C. Han & A. Hestvik. (2022). Auditory predictions are phonological when phonetic information is variable. *Language, Cognition and Neuroscience*.

Rice, K. & P. Avery. (1995). Variability in a deterministic model of language acquisition: A theory of segmental acquisition. In J. Archibald (Ed.). *Phonological acquisition and phonological theory* (pp. 23–42). Lawrence Erlbaum Associates.

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: Language and Cognition*, *18*(2), 179–190.

Rothman, J., J. González Alonso, & E. Puig-Mayenco. (2019). *Third language acquisition and linguistic transfer*. Cambridge University Press.

Stefanich, S., J. Cabrelli, D. Hilderman, & J. Archibald. (2019). The morphophonology of intraword codeswitching: representation and processing. *Frontiers in Communication: Language Sciences*.

Tessier, A.-M. & K. Jesney. (2014). Learning in Harmonic Serialism and the necessity of a richer base. *Phonology 31*(1), 155–178.

Weinberger, Steven H. 1997. Minimal segments in second language phonology. In A. James and J. Leather (Eds.) *Second-Language Speech: Structure and Process* (pp. 263–311). Mouton de Gruyter.

Westergaard, M. (2021). Microvariation in multilingual situations: the importance of property-by-property acquisition. *Second Language Research*, *37*(3), 397–407.

Wu, J. (2021). A contrastive hierarchy analysis of the Mandarin vowel system. In A. Hernández & C. Plyley, (Eds.) *Proceedings of the CLA Annual conference*. https://cla-acl.ca/pdfs/actes-2021/Wu-Contrastive_hierarchy_analysis_of_the_Mandarin_vowel_system.pdf

Yang, C. (2017). Rage against the machine: evaluation metrics in the 21st century. *Language Acquisition*

Yuan, Q. & J. Archibald. (2022). Modified input training and cue reweighting in second language vowel perception. *Frontiers in Educational Research*.

Wexler, K. & P. Culicover. (1983). *Formal principles of language acquisition*. MIT Press.

Address for correspondence

John Archibald
Department of Linguistics, Clearihue D365
University of Victoria
Victoria, British Columbia
Canada

johnarch@uvic.ca
https://orcid.org/0000-0001-8500-0162

Publication history

Date received: 4 July 2022
Date accepted: 18 May 2023
Published online: 4 July 2023