# Mismatching systems: L2 and loanword phonology

## SUMMARY

To a first approximation, L2 phonology can be modeled as using native-like URs, but inappropriately applying the L1 phonological grammar. A second, even better approximation is obtained by *interpolating* between the L1 and L2 grammars using a variational weighted constraints model (e.g. Maximum Entropy Harmonic Grammar).

To a first approximation, loanword phonology can be handled as borrowing the UR from the source language and simply applying the phonology of the borrowing language (analogous to L2 phonology). However, when loanwords contain sequences that are ill-formed in the borrowing language, they are preferentially repaired by epenthesis (especially initially and medially), even when the borrowing language does not otherwise exhibit epenthesis, a fact which appears to call for loanword-specific faithfulness. An alternative theory is that loanwords are borrowed at the level of SRs, with epenthesis occuring because of the perceptual grammar. I will argue that both of these perspectives are true.

## PART I: L2 PHONOLOGY

### I.0 Examples of non-native-like speech in Spanish speakers acquiring English as L2

(1) a. *book* ~> [buk]/*[bʊk]
  b. *bit* ~> [bit]/*[bɪt]
  c. *bin* ~> [bin]/*[bɪn]
(2) a. *speed* ~> [espid]
  b. *stink* ~> [estiŋk]
  c. *skin* ~> [eskin]

Errors like in (1) result from a failure to fully acquire the tense/lax contrast. Subjectively, I experience the result as neutralization to tensed vowels. This kind of error results from the fact that English has a *segmental* contrast which is not present in Spanish. Errors like in (2) result from a repair process which is applied to resolve a *phonotactically illegal sequence*, *#s[+cons]*. The former case has been studied in the literature as “non-native phonetic categorization/learning”, while the latter case is what people think of when they say “L2 phonology”. However, under the modern, OT-ish view that segmental contrasts and phonotactics *both* emerge from the grammar, errors like (1) and (2) both potentially fall under the domain of L2 phonology.

### I.1 Phonetic categories

### I.1.1 *The discrete-continuous mapping*

Phonological representations are discrete symbols, embedded in paradigmatically contrasting, hierarchical structures. The body must convert these representations to continuous sound waves, and then the speech perception system must invert the mapping to recover all or most of the intended representations. Since the early days of spectrograms, we have seen two opposing truths that characterize the discrete-continuous mapping:

- given a phonological category X in prosodic structure Y (X/Y), there are identifiable temporal and/or spectral cues which on average make X different from other paradigmatic alternatives (X/Y ≠ X'/Y)
- the within-speaker, between-speaker, within-category, and between-category is in many cases so large that distinguishing two categories from acoustic values alone is not possible (*s*/Y --?--> X, X')

Despite this, monolingual speakers are almost aways able to recover exactly the intended meaning when it is produced by typical speaker from the same discourse community.

*This figure is from Hillenbrand et al. (1995), a replication of the original Peterson & Barney study on American English vowels. What you are supposed to see is that the within-category variability is on par with the between-category differences. Assuming the ellipses represent 2 confidence intervals (99% of all data), then perhaps half of the /æ/~/ɛ/ tokens are mutually confusable. Most of the tense/lax contrasts are partially ambiguous. Unsurprisingly, these turn out to be some of the most difficult segmental contrasts for non-native speakers of English to acquire.*

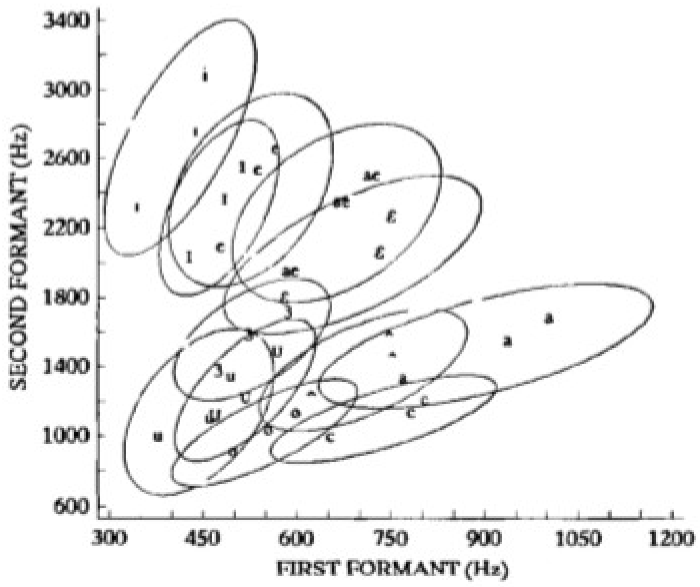


FIG. 3. Average values of $F1$ and $F2$ for men, women, and child talkers for 12 vowels with ellipses fit to the data ("ae"=/æ/, "a"=/ɑ/, "c"=/ɔ/, "ʌ"=/ʌ/, "3"=/ɝ/).

### *I.1.2 Categorical perception in monolingual listeners*

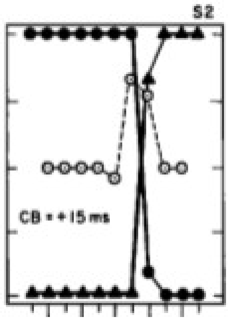
A *category* is a mental construct which groups together some stimuli as being cognitively the 'same', and all other stimuli as 'different'. The Hillenbrand et al. (1995) data is just one of thousands of examples suggesting that categories cannot be defined on a purely objective, language-independent basis alone. How do we diagnose categories?

- *discrimination* (often AX)
  - poor within-category discrimination
  - good between-category discrimination
- *identification*
  - test items are compared to one or more reference items
  - ABX task

There is an extensive literature on categorical perception. Important concepts include:

- psychoacoustic scale
- stimulus continuum
- JND -- just noticable difference
- category boundary

In typical experiments, one sets up a stimulus continuum, e.g. investigating vowel height by holding F2 constant, while F1 varies (*x* Barks, *x+2* Barks, *x+3* Barks, etc..). Discrimination is often plotted as the percentage of 'different' responses for neighboring items; identification is often plotted as the percentage of responses that are identified as the same as a known exemplar. Maximal discrimination is always found at the maximum slope of the identification function. Therefore this region is identified as the 'category boundary'. Other research shows that the category 'boundary' is somewhat mutable according to various contextual influences. For example, the /k/~/g/ boundary is shifted away from /g/ in *gift~kift,* but toward /g/ in *giss~kiss,* because of the lexical status of the items in which /g/~/k/ are embedded (Ganong, 1981).



*Pisoni (1977): the participant heard a low tone (500 Hz) and a high tone (1500 Hz), whose onsets were manipulated (-50 ms, -30 ms, 0 ms, +30 ms, +50 ms; -50 means low tone onset preceded high tone onset by 50 ms). __Identification__: Participants learned to associate -50 ms item with __Left__ button and +50 ms item with __Right__ button. __Discrimination__: For all pairs on the continuum, participants heard the more __Left__ward item, and the more __Right__ward item, and then a repeat of one of those two. They indicated whether the third (X) item was more like the first or second. Filled bubbles (black) indicates identification; open bubbles (dashed) discrimination*

### *I.1.3 Non-native phonetic categories in <u>completely inexperienced</u> speakers/listeners*

Phonetic categories differ from language to language. For example, English has two liquids /l/ and /ɹ/, which may contrast in various positions, while Japanese has a single liquid /ɻ/. We may formalize the notion of *difficult phonetic contrast* as follows:

(3) a. Two sounds [x] and [y] form a *phonetically difficult contrast* if

- [x] and [y] are acoustically similar (e.g. similar values for continuancy, place, etc..)
- but [x] and [y] lexically contrast in one or more languages

b. Two sounds [x] and [y] are said to be a *non-native contrast* for a speaker *s* if

- [x] and [y] lexically contrast in one or more languages, but
- [x] and [y] do not form a lexical contrast in the native language of *s*

The generalizations that emerge from the vast literature on non-native speech perception are:

- if [x] and [y] are lexically contrastive in *s*'s native language, an adult, monolingual speaker *s* is assured to exhibit good discrimination/identification (expected: native-like contrast)
- if [x] and [y] make a phonetically difficult, non-native contrast, an adult, monolingual speaker *s* is very likely to exhibit poor discrimination

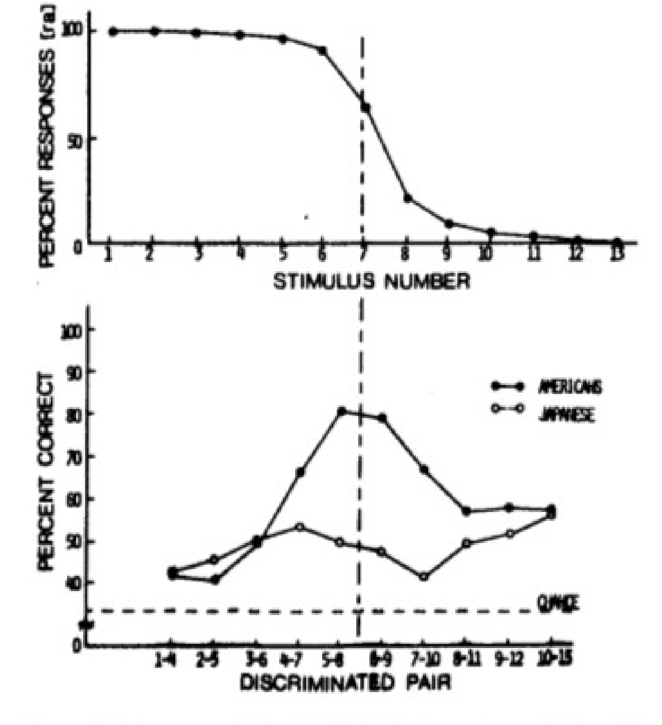


**Figure 2. Upper graph: Pooled identification of speech stimuli by Americans. Lower graph: Pooled discrimination by Americans (closed circles) and Japanese (open circles).**

*English word-initial [l] vs. [ɹ] is* ***the*** *canonical case of a phonetically difficult, non-native contrast (for Japanese listeners).*

*This data, from Miyawaki et al. (1975), shows the discrimination and identification curves for relatively monolingual Japanese listeners, as well as relatively monolingual American English listeners. Identification was not done for Japanese listeners since /l/ and /r/ are not separate sound categories in Japanese. The stimuli in this experiment consisted of artificially synthesized [Ca] tokens, differing only in the trajectory of F3, with item 1 most resembling American English [l] and item [13] most resembling American English [ɹ]*

*The stimulus scales are aligned to show that maximal discrimination occurs at the identification 'boundary'.*

### *I.1.4 The Perceptual Assimilation Model (Best and colleagues)*

Best, McRoberts, and Sithole (1988) proposed the Perceptual Assimilation Model (PAM). The core ideas are as follows:

- listeners are attuned to the language-specific cues that signal phonological categories
- speech sounds that are phonetically similar to one or more existing categories are assimilated to the perceptually closest category
- speech sounds that are not phonetically similar to any existing categories are treated as non-speech; these will be discriminated as well or as poorly as the mammalian auditory system allows

| (4) Contrast | Listener | PAM's predictions | confirmed? |
|---|---|---|---|
| English [l]~[ɹ] | Americans | [l] -> /l/; [ɹ] -> /ɹ/<br>assimilated to *distinct* native cat's<br>good discrimination | √ |
| English [l]~[ɹ] | Japanese | [l] -> /ɽ/; [ɹ] -> /ɽ/<br>assimilated to *same* native cat<br>poor discrimination | √ |
| Zulu clicks* | American | clicks treated as non-speech<br>not assimilated to native cat's<br>good discrimination | √ |
| Swedish [i]~[y] | American | Swedish [i] -> English /i/; Swedish [y] closest to English /i/, but a very poor exemplar; some discrimination | √ |

*click contrasts are as shown in the following table:

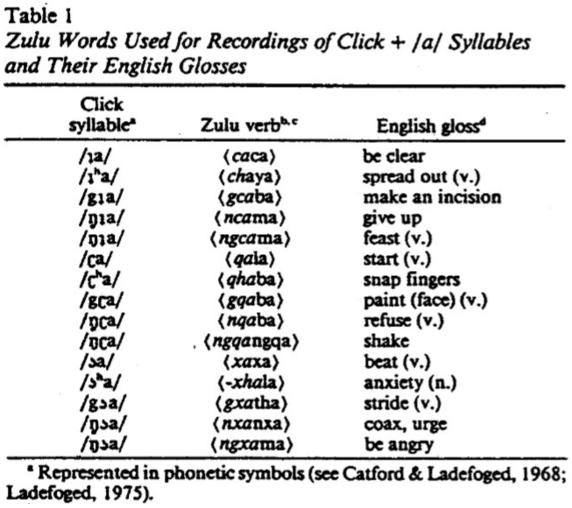


**Table 1**

*Zulu Words Used for Recordings of Click + /a/ Syllables and Their English Glosses*

| Click syllable[a] | Zulu verb[b,c] | English gloss[d] |
|---|---|---|
| /ʇa/ | ⟨*caca*⟩ | be clear |
| /ʇʰa/ | ⟨*chaya*⟩ | spread out (v.) |
| /gʇa/ | ⟨*gcaba*⟩ | make an incision |
| /ŋʇa/ | ⟨*ncama*⟩ | give up |
| /ŋʇa/ | ⟨*ngcama*⟩ | feast (v.) |
| /ʗa/ | ⟨*qala*⟩ | start (v.) |
| /ʗʰa/ | ⟨*qhaba*⟩ | snap fingers |
| /gʗa/ | ⟨*gqaba*⟩ | paint (face) (v.) |
| /ŋʗa/ | ⟨*nqaba*⟩ | refuse (v.) |
| /ŋʗa/ | ⟨*ngqangqa*⟩ | shake |
| /ʖa/ | ⟨*xaxa*⟩ | beat (v.) |
| /ʖʰa/ | ⟨*-xhala*⟩ | anxiety (n.) |
| /gʖa/ | ⟨*gxatha*⟩ | stride (v.) |
| /ŋʖa/ | ⟨*nxanxa*⟩ | coax, urge |
| /ŋʖa/ | ⟨*ngxama*⟩ | be angry |

[a] Represented in phonetic symbols (see Catford & Ladefoged, 1968; Ladefoged, 1975).

A final possibility is that a single native category is the closest perceptual match for both non-native sounds, but it is a good match for only one of them. This is the case for Swedish [i]~[y] (English /i/), Thomspon Salish [k]~[q] (English /k/), and many other cases. Note that borderline cases may arise, such as the Spanish voicing contrast (prevoiced [b] vs. medium-lag [p]) as perceived by English listeners (word-initial Spanish [p] sometimes perceived as English long-lag /p/, sometimes as short-lag /b/).

### *I.1.5 Non-native phonetic categories in <u>experienced speakers/listeners</u>*

All of the above is background for L2 phonology, but differs from L2 phonology in that the listeners do not have any experience with the non-native contrast. One might imagine that people learn with experience. That turns out to only sometimes be the case. There are two sources of evidence for this: studies with speakers who are immersed in an L2 for a long time, and training studies. The general conclusion is the same from both: intra- and between-speaker variation (see refs. below)

- some L2 speakers are able to perfectly acquire non-native segmental contrasts and phonetics, as evident from acoustic measures and/or nativelikeness ratings of isolated words/syllables
- most L2 speakers acquire the contrast imperfectly, in that productions of the contrast manifest acoustic differences on most or all of the same dimensions as natives, but not to the same extent, or variably so
- particularly difficult phonetic contrasts tend to be completely merged in the production or perception of all L2 speakers, and even highly concentrated training with lots of variability does not yield improvements that generalize beyond the training stimuli
    - English [l]~[ɹ] for Japanese listeners is one such case (Bradlow et al., 1999)

Here are some illustrative graphs from the work of Flege:

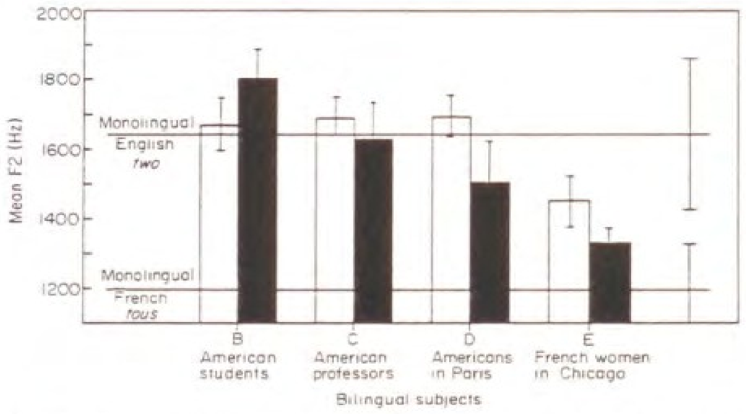
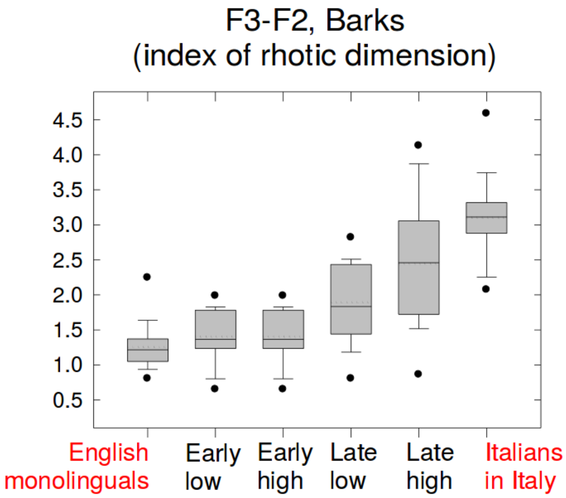


**Figure 2.** The mean F$_2$ frequency, in Hz, of the /u/ tokens in *tous* (■) and *two* (□) produced by L2 learners in four groups (represented by bars) and by monolingual native speakers of English and French (represented by horizontal lines). Most means are based on 70 observations; the brackets enclose ± 1 standard deviation.

*Flege (1987) conducted a production study of the French /y/-/u/ vowel contrast by native English speakers. French monolinguals distinguish these categories with a robust F2 contrast, while Americans living in Paris exhibit an overlapping distribution that does not fully match the French target, so that the intended contrast will only sometimes be recoverable from the phonetics.*



*Flege (2005; presenting data originally mentioned in Flege, Schirru, & MacKay, 2003) studies acquisition of the rounded central vowel schwar (as in bird) by Italian-Americans. F3-F2 in barks is used as an index of rhoticity. "Early"/"Late" refers to whether the speaker began acquiring English during early adolescence or after adolescence; "low"/"high" refer to the degree of Italian usage. L2 speakers exhibit a rhoticity value that is intermediate between the native target and the completely inexperienced L2 value. L2 speakers who use Italian more often exhibit more variability in rhoticity.*

### *I.1.6 The Speech Learning Model*

Flege (2005) states that the purpose of the Speech Learning Model (SLM) is to account for how individuals learn – or fail to learn – to produce and perceive phonetic segments (vowels, consonants) in a second language. The core assumptions of the SLM are as follows:

- L2 learners can, in time, veridically perceive the phonetic properties of L2 speech sounds
- as in L1 learning, L2 learning takes time, and is influenced by the nature of the input
- as in L1, production is guided by perceptual representations stored in LTM
- processes of L1 acquisition, including category formation, remain intact in L2 acquisition
- phonetic elements in L1 and L2 exist in a common phonological space, and influence each other

Predictions:

- the greater the perceived dissimilarity between an L2 sound and a neighboring L1 category, the more likely that a new category will be formed
- category formation for an L2 sound becomes less likely throughout childhood as the L1 representations develop and are elaborated
- when a novel category is not formed in the L2, the L2 sound will be assimilated to an existing L1 category; the merged category will be affected by both L1 and L2 phonetic distributions
- when a novel category is formed in the L2, it may dissimilate from the neighboring L1 category to preserve contrast

Flege (2005) interprets his own and others' work as demonstrating that segmental acquisition is not phonological at all:

- "Both early and late learners can gain access to features not used to contrast L1 phonemes"
    - mentioned in context of schwar study above, but likely intended more generally
- "Their problem seemed to be learning new phonetic segments; in other words, their difficulty seemed to be a problem of phonetic implementation rather than abstract phonological representation"
    - mentioned in the context of Arab speakers acquiring a non-native laryngeal contrast

### *I.1.7 Is this phonology?*

The generative approach has long been to treat segments as partially epiphenomenal -- as bundles of articulatory features whose timing is coordinated partially autonomously. Indeed, the only really compelling evidence I am aware of for a purely segmental level of representation comes from speech errors (identical segments repeated at short distance elicit a large increase in speech errors, e.g. *edited*; no significant increase is seen for segments that differ by a single feature; see review in Goldrick, 2002). However, the constraint-based approach attempts to explain not only alternations but the distribution of static contrasts as falling out from the grammar.

(4) *Richness of the Base (ROTB), Freedom of Analysis, and Lexicon Optimization*

- Marked features expressed only if faithfulness constraints protect them in one or more positions
- Features will never be expressed if both contextual and general markedness constraints outrank the relevant faithfulness constraints
- Lexical representations may contain phonological features that are not expressed in the output
- Learners are free to posit features in their input for which there is no evidence on the surface
- Indeed, learners do exactly this to discover the relative rankings of markedness and faithfulness during acquisition
- However, once the contrast system has been learned, learners only posit contrastive features when there is some chance for them to be expressed

A reminder of how this works:

| (5) /ba/ | Ident[nas] | *$V_{nasal}$ | *$V_{oral}$N |
|---|---|---|---|
| ► [ba] | | | |
| [bã] | * | * | |

| /bã/ | Ident[nas] | *$V_{nasal}$ | *$V_{oral}$N |
|---|---|---|---|
| [ba] | * | | |
| [bã] | | * | |

| /ban/ | Ident[nas] | *$V_{nasal}$ | *$V_{oral}$N |
|---|---|---|---|
| [ban] | | | * |
| [bãn] | * | * | |

| /bãn/ | Ident[nas] | *$V_{nasal}$ | *$V_{oral}$N |
|---|---|---|---|
| [ban] | * | | * |
| [bãn] | | * | |

Other rankings:

*$V_{nasal}$ >> {*$V_{oral}$N, Ident[nas]}
  *no nasal vowels, ever (Spanish)*
*$V_{oral}$N >> Ident[nas] >> *$V_{nasal}$
  *contextual neutralization*
  *contrast non-prenasally, only $V_{nasal}$ prenasally*

ROTB -- all segments may be spec'd as [+nasal] or [-nasal]; here just look at pre-nasal and pre-pausal vowels

Let *$V_{oral}$N >> *$V_{nasal}$ >> IDENT[NAS]
  /ba/ → [ba]
  /bã/ → [ba]
  /ban/ → [bãn]
  /bãn/ → [bãn]
Surface pattern observed (English):
  [ba], *[bã]
  [bãn], *[ban]

Compare with rule-based analysis:
  no underlying nasality
  allophonic nasalization

Constraint-based view:
  vowels can be [+/- nasal]
  constraints drive expression

In this language contextual markdn's is most important and faithfulness is least important: allophonic nasal'n.

Infant stage: observes allophonic distribution [ba], [bãn]; {*$V_{oral}$N, *$V_{nasal}$} >> ID[NAS]
  nasalization occurs on surface, so *$V_{nasal}$ is demoted into next stratum (still above IDENT[NAS])
  learner pushes /bã/ and /ban/ through the grammar, deriving [ba] and [bãn]
  this is the right grammar -- derives all and only the right SR types
  nasalization is not contrastive pre-nasally or non-pre-nasally -- so do not bother with it in URs
Adult stage: observes L2 SR [bã]
  assume the listener has overcome L1 perceptual biases, correctly perceives nasality
  phonologically, must move ID[NAS] to a higher position in the grammar

So here is the crux:

- Empirical evidence (and SLM) suggest it is possible (but not guaranteed) for listeners to form new segmental categories and/or incorporate new features
- According to this view, failure to acquire native-like production and perception arises from L1 interference at the *phonetic* level, not the phonological one
- ROTB says the grammar is the thing which is supposed to determine the surface pattern of contrasts

Are these views compatible? **Yes, they are very similar.**

- Both agree it is possible for L2 speakers to acquire lexical representations that are native-like
- Both agree, not all features that occur in lexical representations are always faithfully expressed
- Both depend on a not-yet-fully-specified theory of the discrete-continuous mapping
- SLM significantly more precise about when L2's will fail to acquire a non-native contrast
- ROTB is significantly more precise about positional licensing

### *I.1.8 Conclusion (Goldilocksian)*

Non-native segmental acquisition is both phonological and phonetic.
L2 speakers normally have little trouble with the phonological end of non-native segments.
Rather, the problem seems to be in acquiring the phonetic implementation

*Thought questions*:

- how do we handle cases like Japanese [l]~[ɹ], where the contrast is generally not acquired?
- what about Korean, where [l] and [ɻ] both occur, but allophonically?
- what about native segments in non-native positions, e.g. Americans learning Tagalog [ŋa]?

## I.2 Non-native phonotactics

We have seen above that L2 speakers appear to acquire rather veridical lexical representations in their non-native language. Although there are cases of absolute failure to learn non-native categories, the general pattern is that non-native segments are assimilated to the nearest native category, with new category formation if/when necessary. Though we do not have a good predictive model of the phonology-phonetics interface as of yet, several approaches converge on the idea that L2 speakers' trouble with non-native segments largely arises at the phonetic level (perception: tuning to the relevant acoustic cues; production: perhaps mastering the coordination of the relevant articulatory gestures), where there is considerable L1 interference. In short, learning new segments is a matter of putting them into the UR, and learning the appropriate phonetics. Phonology is involved, but not problematic (except for the broader problem of the phonology-phonetics/discrete-continuous mapping). What about non-native phonotactics?

### *I.2.1 Background: The Emergence of the Unmarked (TETU)*

The Emergence of the Unmarked is an emergent property of constraint-based phonology. Recall that competition is decided by the highest-ranked constraint for which all violations are not equal. This means that, in general, low-ranked constraints don't have a lot of say. TETU refers to cases in which lower-ranked constraints turn out to matter, after all. One such is epenthesis:

| (6) | /kɪs+z/ | *SibSib | Max-C | Dep-V | *Low | *Back | *Front | *High |
|---|---|---|---|---|---|---|---|---|
| | kɪsz | *! | | | | | | |
| | kɪz | | *! | | | | | |
| | kɪsaz | | | * | *! | | | |
| | kɪsuz | | | * | | *! | | * |
| | kɪsoz | | | * | | *! | | |
| | kɪsɛz | | | * | | | *! | |
| → | kɪsəz | | | * | | | | |

The epenthetic vowel in most languages tends to be the shortest one (Parlato-Oliveira, 2010), with comparatively unmarked vowel qualities. An alternative formalization to (6) would be to replace the context-free markedness constraints with a family of *Map(∅, *x*) constraints. This has the pleasing effect of unifying the major faithfulness families (Dep-*x*, Max-*x*, Id[F] → *Map), at the cost of potentially exploding the actual number of constraints. Similar remarks apply to epenthetic consonants, which tend to be [ʔ], [t], or [n]. Glottal stops are sometimes analyzed as placeless; coronal, obstruent, and nasal are comparatively unmarked. TETU arises in other places (such as what neutralization is to in weak environments) but the key point is: *low-ranked con's get their say when everything else is equal*.

### *I.2.2 Case study: Mandarin speakers acquiring English as an L2*

Broselow, Chen, & Wang (1998) presented data first collected in Wang (1995), presented in two different 'slices' (Table 1; Tables 2-4). In this experiment, native Mandarin listeners were presented with an English nonce word (e.g. [vig]) and asked to repeat it out loud. Their response was transcribed and coded. Categories included 'correct' [vig], 'epenthesis' [vigɤ], 'deletion' [vi], and 'devoicing' [vik].

Note that Mandarin has much more restrictive syllable phonotactics than English. The only coda consonants permitted in standard Mandarin are [n] and [ŋ] (the Beijing dialects also permit [ɹ]). Mandarin does not have a voicing contrast per se as all obstruents are voiceless (though onset plosives contrast in aspiration). Thus, strong L1 interference might be expected in English/L2 forms with obstruent codas, especially voiced obstruents.

Wang (1995) found that indeed, 88/90 English forms with voiced codas were repaired somehow. However, the preferred repair depended on the number of syllables in the form: *when the input was monosyllabic, epenthesis was preferred; but when the input was disyllabic, deletion was preferred*.

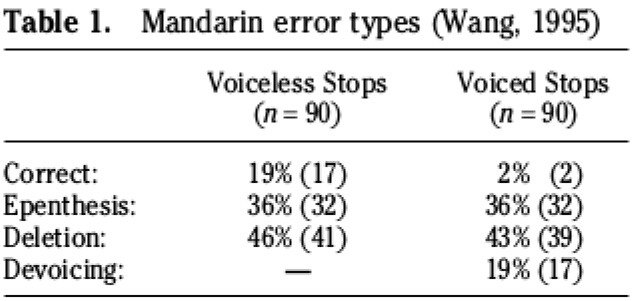


**Table 1.** Mandarin error types (Wang, 1995)

| | Voiceless Stops ($n = 90$) | Voiced Stops ($n = 90$) |
|---|---|---|
| Correct: | 19% (17) | 2% (2) |
| Epenthesis: | 36% (32) | 36% (32) |
| Deletion: | 46% (41) | 43% (39) |
| Devoicing: | — | 19% (17) |

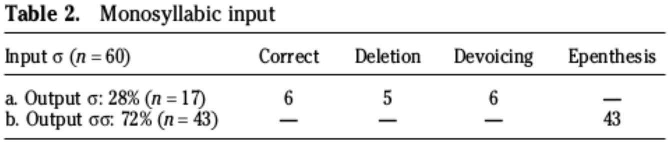


**Table 2.** Monosyllabic input

| Input σ ($n = 60$) | Correct | Deletion | Devoicing | Epenthesis |
|---|---|---|---|---|
| a. Output σ: 28% ($n = 17$) | 6 | 5 | 6 | — |
| b. Output σσ: 72% ($n = 43$) | — | — | — | 43 |

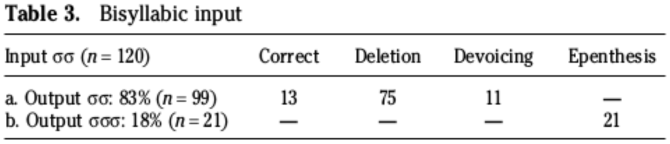


**Table 3.** Bisyllabic input

| Input σσ ($n = 120$) | Correct | Deletion | Devoicing | Epenthesis |
|---|---|---|---|---|
| a. Output σσ: 83% ($n = 99$) | 13 | 75 | 11 | — |
| b. Output σσσ: 18% ($n = 21$) | — | — | — | 21 |

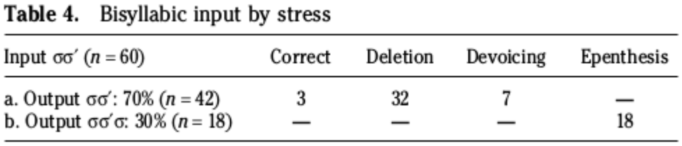


**Table 4.** Bisyllabic input by stress

| Input σσ′ ($n = 60$) | Correct | Deletion | Devoicing | Epenthesis |
|---|---|---|---|---|
| a. Output σσ′: 70% ($n = 42$) | 3 | 32 | 7 | — |
| b. Output σσ′σ: 30% ($n = 18$) | — | — | — | 18 |

A second finding was that there is considerable variation, between and presumably within speakers. We will put that aside for the moment, and come back to it in the next section.

Broselow et al. (1998) pointed out that to account for the variable patterns of repair in Wang (1995) using SPE-style rules, the analyst would be forced to posit a very special kind of 'interphonology':

(7) Rule-based interphonology of Mandarin-English speakers

| | |
|---|---|
| *Word-Final Stop Deletion* | [-son,-cont] → ∅ / __# |
| *Word-Final Epenthesis* | ∅ → ɤ / [-son,-cont]__# |
| *Word-Final Devoicing* | [-son,-cont] → [-vce] / __# |

What makes these rules so special (aside from the fact that application is both variable, and largely mutually exclusive) is that *none of these processes are attested in English phonology,* ***or*** *Mandarin phonology*. English has a limited form of stop deletion, in the form of the well-studied *t/d*-deletion; but that process is limited to <u>coronal</u> stops which occur word-finally in a consonant <u>cluster</u>. Wang's items only contained singleton coda stops, and at all 3 major places of articulation. English also has a limited form of vowel epenthesis, but it only occurs in environments derived by inflectional morphology, between nearly identical segments (inter-sibilant epenthesis, inter-coronal stop epenthesis). English may also be alleged to have some kind of variable, phonetic devoicing process, but this is only licensed when the preceding vowel is lengthened. Mandarin has none of these processes; while it surely exhibits conversational reduction processes like what have been documented in English/Dutch/French, the main alternations that occurs in this language are tonal (tone 3 sandhi, second tone neutralization in reduplication and certain compounds). The existence of such 'interphonologies' is a major puzzle under the rule-based approach to phonology, since the 'rational' thing to do when going from a non-alternating language (Mandarin) to a slightly-alternating language (English) is to slowly and gradually learn to apply only the rules in your new language; rather than wildly inventing new rules that are not present in either language and applying them with abandon.

Broselow et al.'s insight was that Wang's participants' behavior is completely sensible from a constraint-based framework. The interference that is present is not easily statable in terms of *processes*, but is very straightforward to state in terms of *output well-formedness*. Mandarin doesn't allow obstruent codas, and it doesn't allow voiced obstruents in general. The repair processes themselves are not attested in either language, because Mandarin no longer has lexical representations that require these repairs. What is being transferred, then, is the relatively high prioritization of the markedness constraints that enforce Mandarin syllable structure. According to the standard logic of OT, when an input form violates a markedness constraint that is higher-ranked than the relevant faithfulness constraints, some kind of repair must be applied. The choice of repair is dictated by which faithfulness constraint is lowest-ranked, such that violating it can repair the markedness violation.

Since epenthesis is sometimes chosen, Dep-V must be a low-ranked (but still active) constraint; and since deletion is sometimes chosen, the same holds for Max-C. There is then still the puzzle of why epenthesis is preferred for monosyllables, but deletion is preferred for disyllables. Broselow et al. point out that this pattern tends to yield bisyllabic outputs:

(8) a. *monosyllabic input* /vig/ → [vigɤ] *disyllabic (CVCV) output*
b. *disyllabic input* /vilig/ → [vili] *disyllabic (CVCV) output*

Broselow formalize the relevant markedness constraint as WdBin ('Penalize PrWds that do not consist of exactly 2 syllables'). For support, they appeal to the mild but noticable tendency for Mandarin 'words' to be bisyllabic (e.g. *xiǎng* can be used to mean 'want/believe/think/desire' but *xiǎng yào* is strongly preferred for the 'want' meaning, even though the *yao* element is contributing little or nothing semantically). We might also appeal to the concept of Minimal Prosodic Word, which requires that content words be minimally binary (moraically, or syllabically); however these two are formally different in that WdBin is violated by forms with 1 *and* more than 2 syllables, while MinPrWd is only violated by forms with 1 syllable. Broselow et al. conservatively appeal to two additional constraints, NoObsCoda and NoVcdObsCoda. The relevant tableaux are below:

(9) Tableaux for repairs of maximally ill-formed inputs in L2 phonology

| /vig/ | Dep-V | Max-C | WdBin | NoVcdObsCoda | NoObsCoda | Id[vcd] |
|---|---|---|---|---|---|---|
| [vig] | | | * | * | * | |
| [vigɤ] | * | | | | | |
| [vi] | | * | | | | |
| [vik] | | | * | | * | * |

| /vilig/ | Dep-V | Max-C | WdBin | NoVcdObsCoda | NoObsCoda | Id[vcd] |
|---|---|---|---|---|---|---|
| [vilig] | | | | * | * | |
| [viligɤ] | * | | * | | | |
| [vili] | | * | | | | |
| [vilik] | | | | | * | * |

Broselow, Chen, & Wang conclude by giving the constraint rankings that can generate each output that is observed in Wang's data. They also go on to discuss (but not formalize) some tendencies in this data that correlate with stress. Namely, their bisyllabic forms varied in whether the input was trochaic or iambic (as well as whether the coda consonant was voiced or voiceless). They find that when the coda consonant is in a stressed syllable, preservation is far more likely (epenthesis especially, but also devoicing), which they attribute to un-formalized perceptual salience.

### *I.2.3 Handling variation*

At the time that Broselow et al. published, there were no good tools for handling variation. However, in the last decade we have made remarkable progress in handling variation. An especially promising formalism is MaxEntHG (Goldwater & Johnson, 2003; Hayes & Wilson, 2008; et seq.). MaxEntHG is a constraint-based framework, very similar to OT, except the EVAL component works as follows:

- constraints have weights, rather than absolute ranks
- the Harmony of an input-output pair is the weighted sum of its constraint violations
- the probability of an input-output pair is proportional to the exponential of its Harmony

An especially appealing aspect of MaxEntHG is that, given a set of finite tableaux set and the frequencies with which each candidate is observed, there is a guaranteed optimal set of constraint weights, and it can be computed efficiently (Della Pietra et al., 1997). This means that we can treat a constraint set as a *predictive model* of variation, and estimate constraint weights from observed data.

To do this with Broselow et al.'s data, we begin by getting the counts for each input-output pair. Working backwards from Tables 2-4, we can infer the following:

| shape | voiced | voiceless | correct | deletion | devoicing | epenthesis |
|---|---|---|---|---|---|---|
| σ́ | *n*=30 | *n*=30 | 6 | 5 | 6 | 43 |
| σ́σ | *n*=30 | *n*=30 | 10 | 43 | 4 | 3 |
| σσ́ | *n*=30 | *n*=30 | 3 | 32 | 7 | 18 |

| | | |
|---|---|---|
| 2 | 17 | correct |
| 39 | 41 | deletion |
| 17 | -- | devoicing |
| 32 | 32 | epenthesis |

Regrettably, Broselow et al. did not provide exact counts broken down by both prosodic shape *and* voicing of the final consonant. I have estimated them assuming conditional independence, below.

(10) Inferred counts from Broselow et al., (1998)

| | correct | deletion | devoicing | epenthesis |
|---|---|---|---|---|
| σ́, voiced | 6*2/19 =**0** | 5*39/80=**2** | 6*17/17=**6** | 43*32/64=**22** |
| σ́, voiceless | 6*17/19=**6** | 5*41/80=**3** | 6*0/17=**0** | 43*32/64=**21** |
| σ́σ, voiced | 10*2/19=**2** | 43*39/80=**21** | 4*17/17=**4** | 3*32/64=**2** |
| σ́σ, voiceless | 10*17/19=**8** | 43*41/80=**22** | 4*0/17=**0** | 3*32/64=**1** |
| σσ́, voiced | 3*2/19=**0** | 32*39/80=**16** | 7*17/17=**7** | 18*32/32=**9** |
| σσ́, voiceless | 3*17/19=**3** | 32*41/80=**16** | 7*0/17=**0** | 18*32/32=**9** |

I know of three implementations of MaxEntHG, all developed by UCLA faculty:

- **OT-soft** (http://www.linguistics.ucla.edu/people/hayes/otsoft/)
    - *pros*: lots of flexibility, including older learning algorithms for ranked Eval
    - *cons*: only runs on Windows
- **MaxEnt** Grammar Tool (http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/)
    - *pros*: platform-independent (JRE); simple GUI
    - *cons*: requires you to install Java Runtime Environment and get it working properly
- **PhoMEnt** (https://github.com/rdaland/PhoMEnt)
    - *pros*: platform-independent (Python 2.7)
    - *cons*: command-line only

Luckily, these implementations all use the same format, which is well-documented in all cases, and is very close to the tableau format that you are accustomed to from OT. Here are two tableaux in the format you are used to:

(11) Mandarin L2 tableaux

| /vig/ | Dep-V | Max-C | WdBin | NoVcdObsCoda | NoObsCoda | Id[vcd] |
|---|---|---|---|---|---|---|
| [vig] | | | * | * | * | |
| [vi] | | * | | | | |
| [vik] | | | * | | * | * |
| [vigɤ] | * | | | | | |

| /vilig/ | Dep-V | Max-C | WdBin | NoVcdObsCoda | NoObsCoda | Id[vcd] |
|---|---|---|---|---|---|---|
| [vilig] | | | | * | * | |
| [vili] | | * | | | | |
| [vilik] | | | | | * | * |
| [viligɤ] | * | | * | | | |

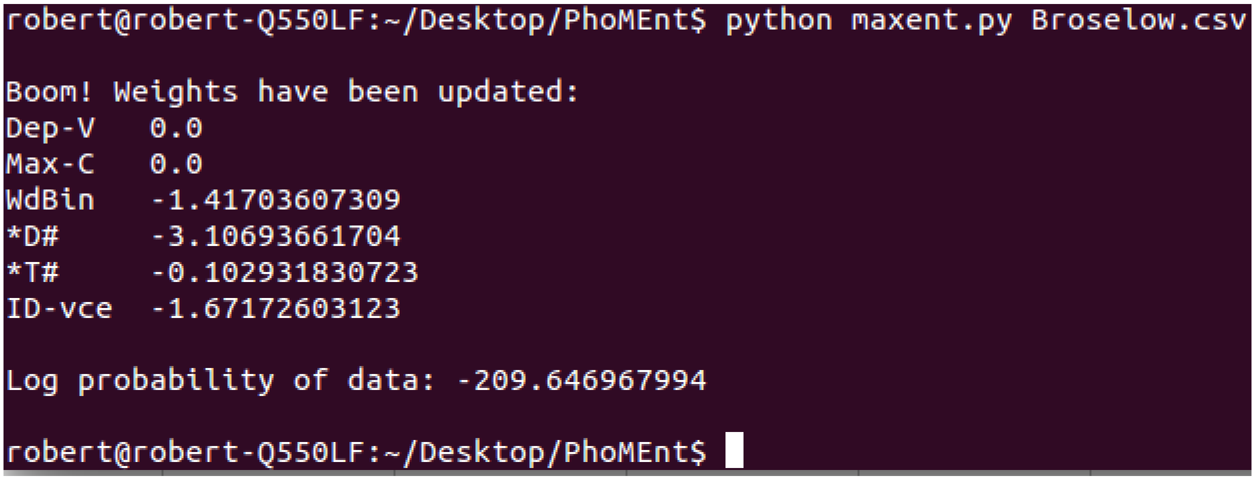
And now, here are the tableaux transmogrified into what the MaxEntHG implementations need:

```
                     Dep-V  Max-C  WdBin  NoVcdObsCoda  NoObsCoda     Id[vcd]
                     Dep-V  Max-C  WdBin  *D#    *T#    ID-vce
vig    vig                         1      1      1
       vi     2             1      1             1
       vik    6                    1                    1
       vigO   22     1
vilig  vilig  2                           1      1
       vili   37            1                    1
       vilik  11                                        1
       viligO 11     1             1
```

(Note that the above conflates both stress patterns, which is completely faithful to Broselow's analysis.) There are several things to note:

- it is always recommended to open tableau files in spreadsheets rather than text editors, because the cell structure is indicated by tabs, which show up consistently in spreadsheets but not in text editors
- in general, every field is separated by exactly one tab; zero values may be omitted
- the very first line begins with 3 tabs, and then consists of tab-separated long constraint names
- the second line also begins with 3 tabs, and then consists of tab-separated short names
    - it is a good idea to choose abbreviations that are <7 characters, so violations are vertically aligned with the short constraint names in the raw text file
- the next line lists the first *input* (/in/$_1$), followed by the first *output* for the first input ([out]$_{1,1}$), followed by the observed *count* of times that the pair (/in/$_1$, [out]$_{1,1}$) was observed, followed by a tab-separated list of the *constraint violation vector* for the pair (in the same order as the constraint names, obviously)
- the next line includes a blank for the input field, which means that this output has the same input as the preceding line; otherwise it has the same format
- zero counts and zero violations can be omitted (blank tab)

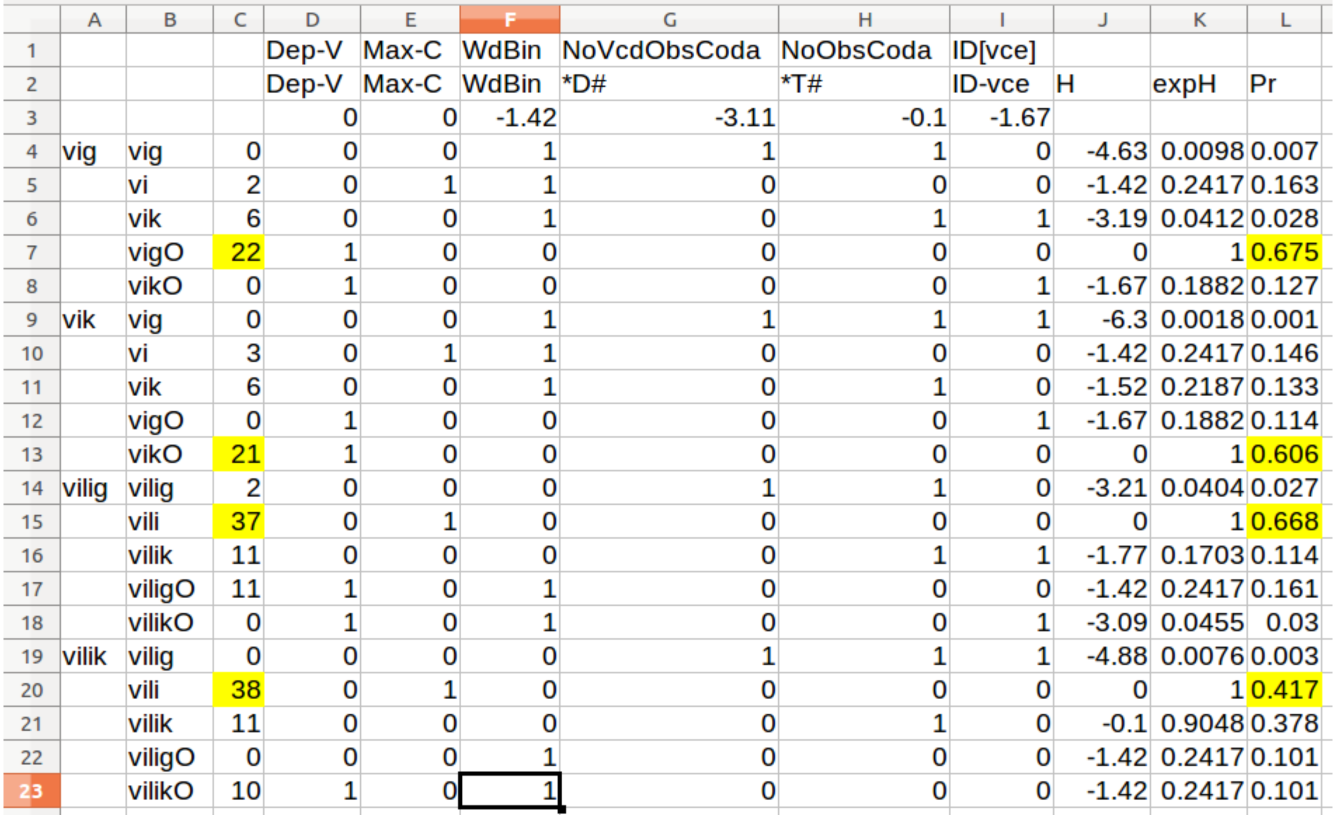
Note that we have chosed two representative inputs (/vig/, /vilig/) to represent all of the voiced items in Wang's data. This is because *all of the voiced items are identical to these from the perspective of the constraints Broselow et al. propose*. All that matters is whether the nonce form is monosyllabic or disyllabic, whether its word-final obstruent is voiced or voiceless, and whether the repair is deletion/epenthesis/devoicing/no-repair. Having set all this up, we can now plug it in to one of the tools above. I used PhoMEnt, since I run Linux and it is the minimal effort to get running if SciPy is already installed (and since my team wrote it, I know exactly how it works):



```
robert@robert-Q550LF:~/Desktop/PhoMEnt$ python maxent.py Broselow.csv

Boom! Weights have been updated:
Dep-V   0.0
Max-C   0.0
WdBin   -1.41703607309
*D#     -3.10693661704
*T#     -0.102931830723
ID-vce  -1.67172603123

Log probability of data: -209.646967994

robert@robert-Q550LF:~/Desktop/PhoMEnt$ 
```

Lo! The magic black box (in this case, literally) has returned weights for the constraints. Inspecting these, we can see that the most important constraint is NoObsVcdCoda (nickname *D#), followed by WdBin and Id[vce]. The remaining constraints are practically or actually inactive.

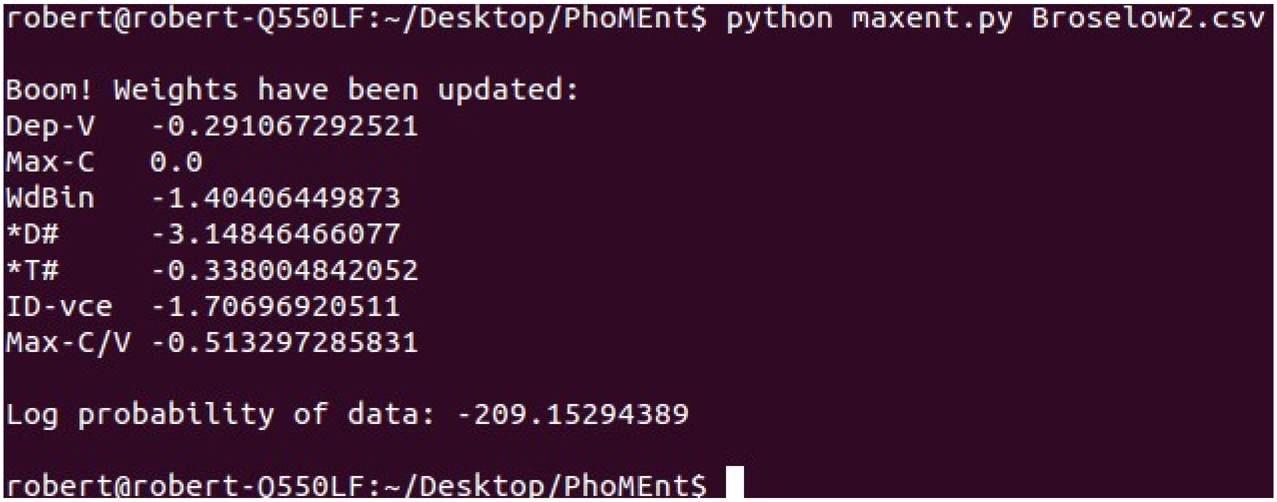
It is somewhat suspicious that the weights for both Max-C and Dep-V are 0, since these are the constraints that are crucially violated by the *next-best competitors* in both the monosyllabic and disyllabic cases. One would imagine that both of these constraints should get high or moderate weight, so as to render the deletion candidate harmonically bounded in the monosyllabic/voiced tableau, and the epenthesis candidate harmonically bounded in the disyllabic/voiced tableau. Since this does not occur, one wonders what probabilities the model is actually assigning. This can be checked fairly easily in Excel (once one has already coded up the relevant tableaux):



|  | A | B | C | D | E | F | G | H | I | J | K | L |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 |  |  |  | Dep-V | Max-C | WdBin | NoVcdObsCoda | NoObsCoda | ID[vce] |  |  |  |
| 2 |  |  |  | Dep-V | Max-C | WdBin | *D# | *T# | ID-vce | H | expH | Pr |
| 3 |  |  |  | 0 | 0 | -1.42 | -3.11 | -0.1 | -1.67 |  |  |  |
| 4 | vig | vig | 0 | 0 | 0 | 1 | 1 | 1 | 0 | -4.63 | 0.0098 | 0.007 |
| 5 |  | vi | 2 | 0 | 1 | 1 | 0 | 0 | 0 | -1.42 | 0.2417 | 0.163 |
| 6 |  | vik | 6 | 0 | 0 | 1 | 0 | 1 | 1 | -3.19 | 0.0412 | 0.028 |
| 7 |  | vigO | 22 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.675 |
| 8 |  | vikO | 0 | 1 | 0 | 0 | 0 | 0 | 1 | -1.67 | 0.1882 | 0.127 |
| 9 | vik | vig | 0 | 0 | 0 | 1 | 1 | 1 | 1 | -6.3 | 0.0018 | 0.001 |
| 10 |  | vi | 3 | 0 | 1 | 1 | 0 | 0 | 0 | -1.42 | 0.2417 | 0.146 |
| 11 |  | vik | 6 | 0 | 0 | 1 | 0 | 1 | 0 | -1.52 | 0.2187 | 0.133 |
| 12 |  | vigO | 0 | 1 | 0 | 0 | 0 | 0 | 1 | -1.67 | 0.1882 | 0.114 |
| 13 |  | vikO | 21 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.606 |
| 14 | vilig | vilig | 2 | 0 | 0 | 0 | 1 | 1 | 0 | -3.21 | 0.0404 | 0.027 |
| 15 |  | vili | 37 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0.668 |
| 16 |  | vilik | 11 | 0 | 0 | 0 | 0 | 1 | 1 | -1.77 | 0.1703 | 0.114 |
| 17 |  | viligO | 11 | 1 | 0 | 1 | 0 | 0 | 0 | -1.42 | 0.2417 | 0.161 |
| 18 |  | vilikO | 0 | 1 | 0 | 1 | 0 | 0 | 1 | -3.09 | 0.0455 | 0.03 |
| 19 | vilik | vilig | 0 | 0 | 0 | 0 | 1 | 1 | 1 | -4.88 | 0.0076 | 0.003 |
| 20 |  | vili | 38 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0.417 |
| 21 |  | vilik | 11 | 0 | 0 | 0 | 0 | 1 | 0 | -0.1 | 0.9048 | 0.378 |
| 22 |  | viligO | 0 | 0 | 0 | 1 | 0 | 0 | 0 | -1.42 | 0.2417 | 0.101 |
| 23 |  | vilikO | 10 | 1 | 0 | 1 | 0 | 0 | 0 | -1.42 | 0.2417 | 0.101 |

(Highlighted counts/probabilities are the most frequent ones in the observed data.) In the bisyllabic data, deletion is preferred fairly strongly over devoicing. Cranking the anti-deletion constraint's weight up would over-penalize /vilig/-[vili]; this can be compensated for by also cranking the anti-epenthesis and anti-devoicing constraints but this in turn completely over-penalizes /vig/-[vik] and under-penalizes /vilik/-[vilik]. In short, the constraints we have do not seem adequate to the complexity of the data.

Broselow et al. admit as much. Their data show that the preference for deletion versus epenthesis is strongly conditioned by whether the final vowel is stressed, and their discussion speculates that deletion from a stressed syllable is less perceptually tolerable. We might then modify the analysis by splitting our data into the full set of categories that Broselow and colleagues supply, and adding a context-specific Max-C/V́_# constraint.



```
robert@robert-Q550LF:~/Desktop/PhoMEnt$ python maxent.py Broselow2.csv

Boom! Weights have been updated:
Dep-V   -0.291067292521
Max-C   0.0
WdBin   -1.40406449873
*D#     -3.14846466077
*T#     -0.338004842052
ID-vce  -1.70696920511
Max-C/V -0.513297285831

Log probability of data: -209.15294389

robert@robert-Q550LF:~/Desktop/PhoMEnt$ 
```

This helps, modestly, although note that it does not significantly improve the data probability.

### *I.2.4 More instances of this technique*

Daland & Norrmann (accepted) applied this technique to model prothesis in Mexican Spanish speakers acquiring English as an L2. The basic phenomenon has long been thought to be driven by syllable-structure constraints (Carlisle, 1991; *et seq.*), as shown in (11):

| (12) | /spid/ *speed* | *[$_\sigma$ sC...] | Onset | Dep-e |
|---|---|---|---|---|
| | spid | *! | | |
| ☞ | es.pid | | * | * |

Norrmann collected /sC/ targets in a variety of contexts (post-pausal, post-consonantal, post-vocalic), varying the modality (reading out loud, repeating after her). There were various interesting things about the data, which we will pass over here.

**Table 5. Left: Raw counts of prothetic vs. faithful outputs from non-native speakers, pooled across speaker and cluster (repetition indicated after task, e.g. "read-2" indicates the second repetition of targets in the reading task). Columns arranged in decreasing order of prothesis rate, i.e. increasing order of native-likeness. Right: Constraint violation profiles (see following section for details).**

| **env't** | **candidate** | **read-2** | **read-1** | **spoken-2** | **spoken-1** | | *[$_\sigma$ sC...] | Onset | Dep-e |
|---|---|---|---|---|---|---|---|---|---|
| C#sC | faithful | 19 | 23 | 31 | 46 | | 1 | | |
| | prothesis | 49 | 43 | 20 | 18 | | | | 1 |
| V#sC | faithful | 33 | 40 | 36 | 55 | | 1 | | |
| | prothesis | 36 | 26 | 16 | 13 | | | 1 | 1 |

Daland found that the very simple 3-constraint grammar shown in (12) did the best job of predicting the variation in the data. When the weights were fit to the observed data, they by-and-large appeared to represent interpolation between the L1 (native/Spanish/source) weights and the target L2-appropriate (non-native/English) weights:

**Table 6. Constraint weights found by the MaxEnt Grammar Tool.**

| | Spanish | read-2 | read-1 | spoken-2 | spoken-1 | English |
|---|---|---|---|---|---|---|
| *[$_\sigma$ sC...] | 12.32 | 6.90 | 6.60 | 5.61 | 5.15 | 0.0 |
| ONSET | 0 | 0.85 | 1.04 | 0.36 | 0.49 | 1.07 |
| DEP-E | 5.63 | 5.96 | 5.98 | 6.06 | 6.1 | 7.37 |

Giavazzi et al. (submitted) apply a similar methodology to modeling generalization in the French adjective system by patients with Huntington's disease (the masculine is related to the feminine on largely phonological grounds, but there are four different regular patterns).

### *I.2.5 Summary and Conclusions*

To a first approximation, L2 segmental acquisition appears to be problematic mostly in terms of mastering the phonetic implementation of novel contrasts. Learning that there *should* be a new contrastive feature is common; learning how to realize it appropriately across contexts is not. Somewhat surprisingly, this view is consistent with tenets of constraint-based phonology (assuming it is easy to re-rank markedness constraints regulating expression of contrast). However, this view loads most of the explanation onto the phonology-phonetics mapping, which we do not have an adequate predictive theory of yet (*target-like URs, source-like phonetic implementation*).

To the extent that we can distinguish L2 segmental acquisition from L2 phonotactics, we can model L2 phonotactics very similarly (*target-like URs, source-like phonotactic grammar*). The most accurate predictive models are obtained by interpolating between the source and target grammars, using a weighted constraint framework. It is slightly different in that we must assume source weightings are effortful to change.

### *I.2.6 Exercise*

1. Install OTSoft, the MaxEnt Grammar Tool, or PhoMEnt. (Recommend OTSoft if you have Windows, PhoMEnt if you already have SciPy installed, and the MaxEnt Grammar Tool otherwise).
2. Here are the counts from Daland & Norrmann's Table 5, collapsed across condition:

| | faithful | prothesis |
|---|---|---|
| C#sC | 119 | 130 |
| V#sC | 164 | 91 |

Prepare an input file in the format described above, using these counts and the constraints/violations from Daland & Norrmann's Table 5.

3. Obtain constraint weights from the program. ***OTSoft***: You will have to specify MaxEntHG learning. ***MaxEnt Grammar Tool***: has sensible defaults. ***PhoMEnt***: The default value for the regularizing prior is too high for a small set of data like this; you will probably want to turn it down, like this: `python maxent.py -L 0.01` *`[yourInputFilename]`*
4. Add an additional tableau with UR /fr/ and the same two candidate types (faithful, prothetic). Make the counts be 1000 for [fr] and 0 for [efr]. Now get the weights. Are they different from the weights you got in 3? If so, why and which grammar do you think is the more psychologically accurate one? If not, why not and what does this tell you about the omission of losing candidates from tableaux?

## PART II: LOANWORD PHONOLOGY

### II.1 Overview

If you only read one thing on loanword adaptation it should be

- Kang, Yoonjung (2010). The emergence of phonological adaptation from phonetic adaptation: English loanwords in Korean. *Phonology 27*, 225-253.

Kang (2010) quotes Haugen (1951) and Paradis & LaCharite (2008, 2009) as follows:

> *Haugen (1950: 216-217) observes that in the initial stage of loanword adaptation (his ‘pre-bilingual period’) words are adapted ‘with great irregularity in the phonetic results’; he refers to such a state as ‘ERRATIC SUBSTITUION’. A subsequent stage (his ‘period of adult bilingualism’) exhibits ‘a more SYSTEMATIC SUBSTITUTION, in which the same [native language] phoneme is consistently employed for new [input language] loans’. In a diachronic study of English loanwords in Quebec French, Paradis & LaCharite (2008, 2009) found that the rate of non-phonological adaptations decreased from 17.5% (70 out of 401 cases) in Old Quebec French (late 19th–early 20th century Quebec City) to 9.5% in recent Quebec City French and to 7.1% in recent Montreal French. In particular, the rate of ‘phonetic approximation’ decreased from 5.5% in Old Quebec French to 1.1% and 0.4% in recent Quebec City French and recent Montreal French respectively. Here, we can observe a diachronic trend from ‘more variable’ to ‘less variable’, and from ‘less phonological’ to ‘more phonological’.*

Kang's work (2010, 2012) involves careful investigation of the diachronic development of loanwords, which is both difficult and valuable (given the outsized role loanword phonology has played in theoretical debates on the relation between loanword phonology and speech perception). Kang (2010) studies the adaptation of English palato-alveolar obstruents /ʃ t͡ʃ d͡ʒ/, and the development indeed appears to follow exactly along the lines of what Haugen (1951) states.

In this sub-unit, we will begin with the 'phonological account' of loanword adaptation, promulgated most forcefully by Paradis & LaCharite. This accounts holds that loanword adaptation is driven by fluent bilinguals, who possess target-like mastery of the source language. The phonological account holds that adaptation occurs at the level of the UR, i.e. each source phoneme is translated to the perceptually closest target phoneme, and then the target phonology is applied, subject to the Preservation Principle (which we will describe in more detail shortly). The alternative, 'perceptual account', advocated for most forcefully by Peperkamp and colleagues, holds that adaptation takes place at roughly the level of the SR. For example, a native Korean adapter hearing English *abstract form* would be forced to parse the SR according to the Korean grammar, assigning the best-matching UR. These contrasting theories are schematized below:

| (13) | English source | | Korean borrowing | |
|---|---|---|---|---|
| a. | /æbstɹækt fɔ˞m/ | ~~> | /æpstlækt pʰom/ | phonological account |
| | ↓ | | ↓ | |
| | [ˈæːb̥st̪ɹæk ˌfɔ˞m] | | [æp̚s*ɨ̥tʰɨ̥ɾæk̚ pʰom] | |
| b. | /æbstɹækt fɔ˞m/ | | /æpsɨtɨlæk pʰom/ | perceptual account |
| | ↓ | | ↑ | |
| | [ˈæːb̥st̪ɹæk ˌfɔ˞m] | ~~> | [æp̚s*ɨ̥tʰɨ̥ɾæk̚ pʰom] | |

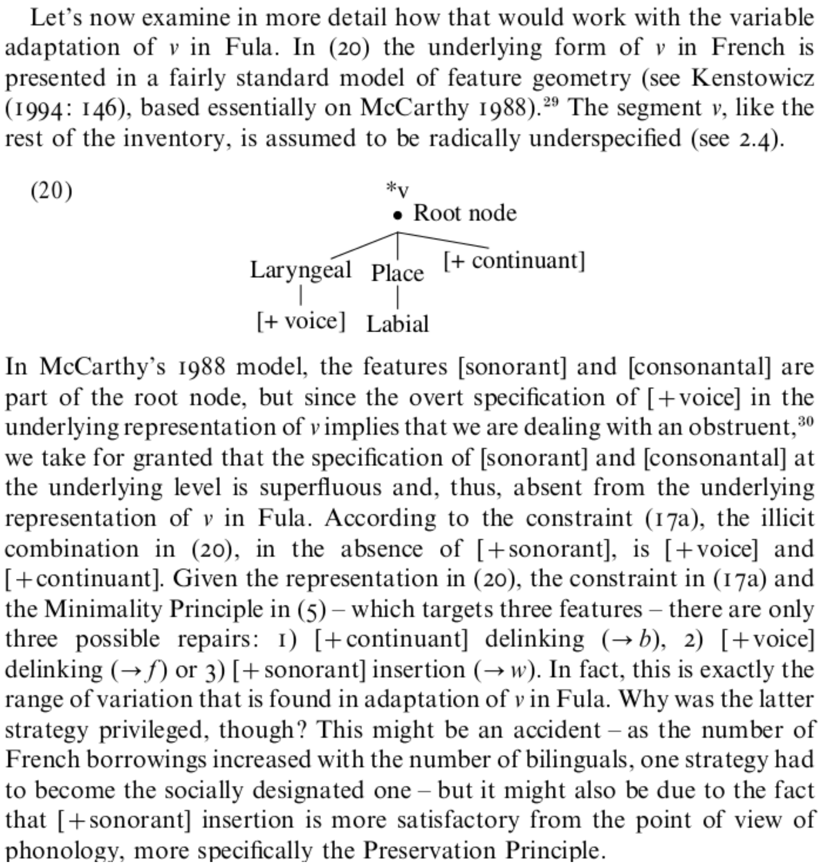
An important point to note is that both theories make the same predictions for a wide variety of phonological configurations. Korean has proven to be a very important language in loanword phonology, owing to a variety of morphological, phonological, and historical properties. In the end, I will argue for a Goldilocksian theory of loanword phonology as well: perceptual adaptation can force certain adaptations (epenthesis), but not all loanword repairs are forced by perception, therefore the residue of non-native adaptations *must* be attributed to phonological adaptation.

## II.2 The Phonological Account

In its simplest form, the phonological account holds that loanword adaptation takes place at the level of the lexical representation. Phonemes in the source UR are translated to the articulatorily closest phonemes in the borrowing language, and then normal borrower phonology applies.

(14) /kɹɪsmas/$_{\text{English}}$ ~~> /kɻismas/$_{\text{Japanese}}$
↓ ↓
[ˈkɹɪsməs] [kɯɻisɯ̥masɯ̥]

The interesting thing about loanword phonology from this point of view is that it yields lexical representations that violate markedness constraints which are otherwise unviolated in the language. For example, Japanese only allows C(G) onsets and nasal or geminate codas. Thus, the form /kɻismas/$_{\text{Japanese}}$ contains 3 distinct phonotactic violations (ill-formed onset cluster, ill-formed medial cluster, ill-formed word-final coda). As evident from the pronounced form, Japanese chooses to repair all three of these violations with epenthesis. In this language, the epenthetic vowel is usually [ɯ]. We can straightforwardly analyze this in a constraint-based framework by putting DEP-V as the lowest-priority faithfulness constraint. Vowel devoicing is analyzed as a postlexical process in which voiced vowels are dispreferred between voiceless consonants, and word-finally after voiceless consonants.[1]



Let's now examine in more detail how that would work with the variable adaptation of *v* in Fula. In (20) the underlying form of *v* in French is presented in a fairly standard model of feature geometry (see Kenstowicz (1994: 146), based essentially on McCarthy 1988).[29] The segment *v*, like the rest of the inventory, is assumed to be radically underspecified (see 2.4).

(20)

In McCarthy's 1988 model, the features [sonorant] and [consonantal] are part of the root node, but since the overt specification of [+voice] in the underlying representation of *v* implies that we are dealing with an obstruent,[30] we take for granted that the specification of [sonorant] and [consonantal] at the underlying level is superfluous and, thus, absent from the underlying representation of *v* in Fula. According to the constraint (17a), the illicit combination in (20), in the absence of [+sonorant], is [+voice] and [+continuant]. Given the representation in (20), the constraint in (17a) and the Minimality Principle in (5) – which targets three features – there are only three possible repairs: 1) [+continuant] delinking (→*b*), 2) [+voice] delinking (→*f*) or 3) [+sonorant] insertion (→*w*). In fact, this is exactly the range of variation that is found in adaptation of *v* in Fula. Why was the latter strategy privileged, though? This might be an accident – as the number of French borrowings increased with the number of bilinguals, one strategy had to become the socially designated one – but it might also be due to the fact that [+sonorant] insertion is more satisfactory from the point of view of phonology, more specifically the Preservation Principle.

Paradis and LaCharite argue for the phonological account of loanword adaptation in a variety of publications, chiefly based on corpus data from English loanwords in Quebec French. Their position is roughly as follows:

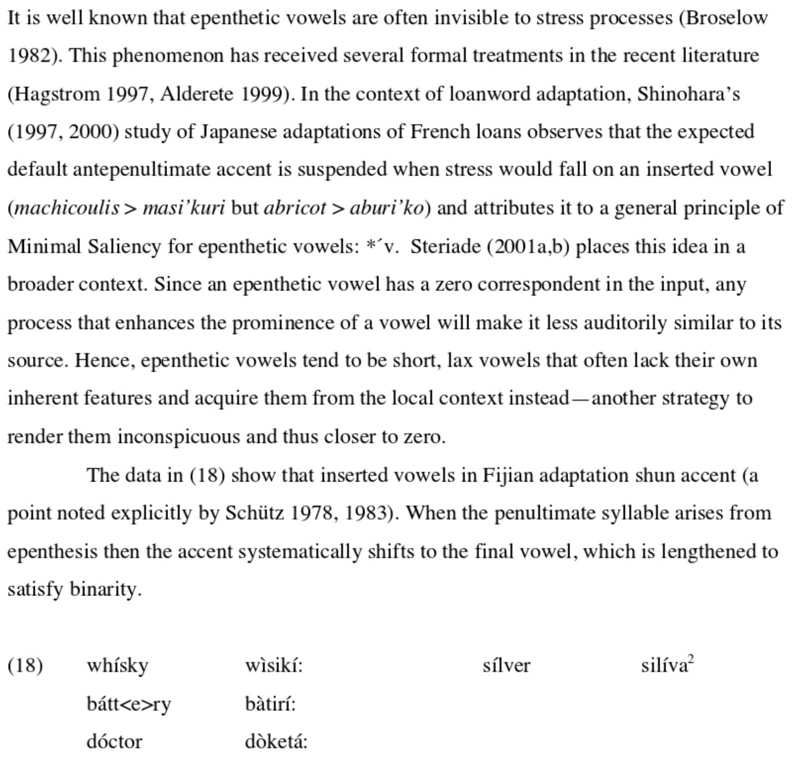
- loanwords are typically adapted by fluent bilinguals
- fluent bilinguals know the phonology of both languages
- when adapting, bilinguals respect phonological contrast to the extent that can be accommodated in the borrowing language
- epenthesis is the default repair strategy (Preservation Principle), but deletion occurs
- when the articulatory modification of the default adaptation would be too great, delete instead (Threshold Principle)

Variability in adaptation is a major challenge for this theory. The excerpt above from Paradis & LaCharite (1997)'s discussion of the adaptation of French [v] into Fula (a language spoken in Senegal).

---

[1] According to this rule, the second epenthetic *u*, after *Christ-*, should not be devoiced. Presumably devoicing is licensed here because *Christmas* is adapted as two components, /*Christ*/ + /*mas*/. See Oh (2012) for evidence that complex words are adapted component-by-component in Korean.

One empirical success of the phonological account is in explaining why epenthetic elements in loanwords appear to avoid accentual defaults. The following excerpt from Kenstowicz (2007) makes this point for accentuation in Fijian loanwords:



It is well known that epenthetic vowels are often invisible to stress processes (Broselow 1982). This phenomenon has received several formal treatments in the recent literature (Hagstrom 1997, Alderete 1999). In the context of loanword adaptation, Shinohara's (1997, 2000) study of Japanese adaptations of French loans observes that the expected default antepenultimate accent is suspended when stress would fall on an inserted vowel (*machicoulis* > *masi'kuri* but *abricot* > *aburi'ko*) and attributes it to a general principle of Minimal Saliency for epenthetic vowels: *´v. Steriade (2001a,b) places this idea in a broader context. Since an epenthetic vowel has a zero correspondent in the input, any process that enhances the prominence of a vowel will make it less auditorily similar to its source. Hence, epenthetic vowels tend to be short, lax vowels that often lack their own inherent features and acquire them from the local context instead—another strategy to render them inconspicuous and thus closer to zero.

The data in (18) show that inserted vowels in Fijian adaptation shun accent (a point noted explicitly by Schütz 1978, 1983). When the penultimate syllable arises from epenthesis then the accent systematically shifts to the final vowel, which is lengthened to satisfy binarity.

| | | | | |
|---|---|---|---|---|
| (18) | whísky | wìsikí: | sílver | silíva[2] |
| | bátt<e>ry | bàtirí: | | |
| | dóctor | dòketá: | | |

## II.3 The Perceptual Account

The perceptual account holds that most or all apparent changes that are introduced during loanword adaptation are driven by speech *perception* rather than speech production. Therefore, most such mutations are predictable based on the phonetics and phonology of the borrowing language. A key consequence of this is that so-called *epenthetic vowels are incorporated into the lexical representation of the borrowed form*:

| | | | |
|---|---|---|---|
| (14) | /kɹɪsmas/$_{\text{English}}$ | | /kɯɻisɯ+masɯ/$_{\text{Japanese}}$ |
| | ↓ | | ↑ |
| | [ˈkɹɪsməs] | ~~> | [kɯɻisɯ̥masɯ̥] |

Another key consequence of this account is that adaptation should be sensitive to phonetic detail, including contextual allophony and sub-phonemic detail. More specifically, the perceptual account holds that loanword adaptation generally occurs so as to *maximize perceptual similarity between the source SR and the borrowing SR, subject to the phonotactic requirements of the borrowing language*. That is, the adapted form is the closest one to the source SR which is legal in the borrowing language.

It turns out that in most cases where this theory and the phonological account both make clear predictions, they are the same predictions and they are correct. For example, the perceptual account has an explanation for the variability of French [v] adaptation into Fula, and also for the avoidance of accentuating 'epenthetic' vowels in Fijian loanwords. (*Exercise*: What are the explanations for these two effects under the perceptual account?) More broadly, both theories predict that vowel epenthesis will be a preferred repair strategy. To distinguish the predictions of these theories, it is necessary to look somewhat far afield. One difference is that the perceptual account predicts greater dependence on context (since adaptation should take place at the level of the allophone in context, rather than the phoneme). Also, the two theories differ in the the perceptual account claims that adaptation takes place during perception, whereas the phonological account assumes relatively veridical lexical representations (implying veridical perception at least in the bilinguals doing the adaptation).

### II.3.1 The *ebzo/ebuzo* study

Dupoux et al. (1999) argue that vowel epenthesis can be driven by native language phonotactics. They compared the performance of native French listeners with native Japanese listeners on a speeded AX discrimination task. The relevant phonological facts are as follows:

- French has a voicing contrast, and it is licensed in medial clusters
- Therefore, word-medially, French has a vowel~zero contrast, but not a vowel length contrast
- Japanese has a voicing contrast, but heterorganic obstruent clusters are illegal
- Therefore, word-medially, Japanese has a short~long vowel length contrast, but not a vowel~zero one

This is schematized in (15), with (~) meaning there is a lexical contrast between the two forms:

| (15) |  | *ebzo* |  | *ebuzo* |  | *ebu:zo* |
|---|---|---|---|---|---|---|
|  | French | √ | ~ | √ |  | * |
|  | Japanese | * |  | √ | ~ | √ |

The results were as follows:

| Discrim. | *ebzo~ebuzo* | *ebuzo~ebu:zo* |
|---|---|---|
| *French* | √ | * |
| *Japanese* | * | √ |

In short, the discrimination results are exactly as predicted according to PAM. Although the Japanese participants were tested in France, they were unable to perceive this contrast in speeded discrimination. The effect has been replicated in a variety of paradigms by a variety of labs, so it is quite robust.

Dupoux and colleagues interpreted the results as evidence that Japanese speakers are experiencing an *auditory hallucination* driven by the phonotactic structure of Japanese. Specifically, since Japanese phonotactics do not allow [b.z] syllable contacts, an actual [b.z] sequence (in which all traces of the vowel itself were digitally excised from the stimuli) must be interpreted as a [bɯ.z] sequence. Note that vowel devoicing is illegal in this context, so it could not be that Japanese listeners were parsing a 'phonological' vowel but not a phonetic one. Followup work has shed some doubt on the details of the explanation as to why Japanese listeners are unable to discriminate *ebzo~ebuzo*: was it really syllable structure, or could it have been linear phonotactic constraints? or could it have been something phonetic? But subsequent work has largely accepted or further supported the interpretation that an illusory vowel is actually perceived by these speakers. This work therefore supports a key prediction of the perceptual account of loanword adaptation, that putative vowel epenthesis may actually take place during perception, so the 'extra' vowel is incorporated into the lexical representation. (*Thought question*: What would Paradis and LaCharite say about Dupoux et al.'s experiment? Do you think Paradis and LaCharite's response would be swayed by the information that most of Dupoux et al.'s participants were highly proficient L2 speakers of French, some with native-sounding accents?)

### II.4 Wait, why are talking about loanwords again?

That's a great question. For Paradis and LaCharite, it is interesting because it provides a source of inputs that fall outside what a language normally provides (that is, a perturbation to the feedback loop between harmonic bounding, Lexicon Optimization, and ROTB). Loanwords are interesting because they provide an opportunity to observe something about the borrowing language's phonology that you could not see otherwise (e.g. TETU effects).

For myself and most of my collaborators, loanword phonology is not really interesting of itself. Rather it is interesting because it sheds light on how the grammar is deployed for natural language tasks, in particular speech perception. To the extent that loanword adaptation is driven by speech perception, it provides a cheap and easily replicable way to study how the native phonological grammar affects speech perception ('cheap' in the comparative sense of analyzing loanword corpora rather than conducting carefully controlled laboratory experiments). The remainder of this lesson will be taken up exploring these issues with a case study, Korean loanword phonology.

## II.5 Korean loanword adaptation

Korean loanword adaptation has played an outsized role in the subfield of loanword adaptation, for a variety of historical and theoretical reasons. Most of the contact between Korean and English has occurred in the 20th century, and the pace of loanword adaptation only really became high during the Korean War. Korean has moderately restrictive phonotactics, and a wide array of interesting alternations. The picture is somewhat complicated by various language changes that are underway or nearly complete in Korean, such as the loss of productive vowel harmony, contrastive vowel length, and a phonetic change in the obstruent series which may or may not be contributing to morphophonological changes.

### II.5.1 A whirlwind tour of Korean phonology

*Phonemes*. Seoul Korean has the monophthongs [a, ɛ, ʌ, o, i, ɨ, ɯ], front and back glides, a single liquid, 3 nasals (labial, coronal, dorsal), coronal and glottal fricatives, and a four-way place contrast in stops (labial, coronal, alveo-palatal, dorsal). The stops are *tense* ([+c.g.,–s.g.]), *aspirated* ([–c.g.,+s.g.]), or *plain/lax* ([–c.g.,–s.g.]). The coronal fricatives may be tense or lax. The clearest phonetic correlate of tenseness is raised F0 on the following vowel (lax stops have lowered F0 on the following vowel, relative to plain stops in English). In present-day Korean, lax and aspirated stops have the same long-lag VOT target while tense stops are short-lag (but still voiceless). The default accentual pattern in Korean is (LH)* (low-high-low-high; no stress *per se*), but this interacts with tenseness: if the word-initial sound is a tense obstruent, the following vowel must be high (H(LH*)).

*Tautosyllabic allophony*. Coronal fricative alveopalatalize before [i]. Alveo-palatal stops are generally affricated. Lax stops are allophonically voiced when they follow a sonorant and precede a glide/vowel. The liquid is a light [l] when in coda position, but a tap/flap [ɻ] when in onset position (geminate [l:]). Word-initial /l/ mutates to [n]. There is extensive coda neutralization: all fricatives and alveoplatal stops reduce to lax [t]. All other stops reduce to plain stops at the same place of articulation. Coda stops are obligatorily unreleased.

*Syllable contact*. Rising sonority across a syllable boundary is *verboten*. Underlying nasal-liquid and liquid-nasal sequences are resolved by total assimilation, usually to a geminate liquid. Stop-nasal sequences are resolved by nasalizing the stop. Place assimilation is optional but standard: labials assimilate before dorsals, while coronals assimilate before labials and dorsals (*Paper topic alert!* allegedly the assimilation pattern is for both stops and nasals -- but I haven't seen any phonetic studies). An obstruent tensifies after any other consonant.

*Morpheme and syllable structure*. Syllables are maximally CGVC. Stems may contain underlying final clusters (e.g. /saks/ 'payment'), which are repaired in the isolation form by deletion ([sak̚] 'payment'). The contrast is preserved with the abundance of vowel-initial morphemes, including especially vowel-initial allomorphs of case particles ([sak̚s*i] 'payment'-NOM).

### II.5.2 Segmental adaptation

For the most part, segmental adaptation is predictable. Some interesting things happen for phonemes which English has but Korean lacks. Voiced stops are most often adapted as lenis stops, especially when they occur word-medially (*Abba* [æbə] ~~$E_2$K~~> /apa/ [aba]); however they are occasionally adapted as tense stops, and less often as aspirated stops. Voiceless stops are almost always adapted as aspirated stops, which is unsurprising from stressed/onset positions, but less obviously a good phonetic match for non-aspirated voiceless stops (*apple* [æpəɫ] ~~E$_2$K~~> /ap$^{h}$ɨl/ [ap$^{h}$ɨl]). Voiceless non-strident fricatives are generally mapped to an aspirated stop at the closest place of articulation ([f] ~~E$_2$K~~> [p$^{h}$]), but there is some variability for [θ]. There is more variability for voiced nonstrident fricatives ([v] and [ð]). The strident fricative [s] is adapted variably as lax [s] or tense [s*], but predictably so (see next section). English [ʃ] is also adapted predictably, as an alveo-palatalized coronal fricative, so a front vowel/glide is epenthesized with it to license the [ɕ]. Word-final voiceless stops are adapted variably, sometimes as word-final plain stops, and sometimes with vowel epenthesis (*pack* /p$^{h}$æk/ ~~E$_2$K~~> [p$^{h}$ɛk̚], [p$^{h}$ɛk$^{h}$ɨ]). Nasals are adapted as expected. English onset *r*'s are adapted into the onset liquid (tapl/flap allophone), while coda *r*'s are deleted (though usually the preceding vowel is then interpreted as [a] or [ʌ], the lowest and longest vowels in Korean). English coda [l]'s are always adapted as coda [l]'s in Korean (with the exception of morphologically complex items like *all in*), while onset [l]'s are variable adapted as onset tap/flaps or geminate [l]'s.

Variable adaptation of the same phoneme is in general a problem for the phonological account. We might attribute all the variable adaptations to an 'early'/'phonetic' stage of adaptation (see Paradis & LaCharite's discussion of the Fula adaptation of French /v/ above). However, this explanation does not really work for variable adaptation that is contextually determined, but categorically so. An example is given in the next subsection.

### II.5.3 Context-specific adaptation 1: word-initial s-stop clusters

English has many words that begin with *sV*, and many others that begin with *sC*. It has been known for some time that these items are adapted differently. The following are from a descriptive corpus assembled by the Korean language academy (NAKL, 1991):

| (16) | | *sV* | | | | *sC* | |
|---|---|---|---|---|---|---|---|
| a. | *satan* | ~~E$_2$K~~> | [s$^{*}$at$^{h}$an] | f. | *star* | ~~E$_2$K~~> | [sɨnɛp̚] |
| b. | *save* | ~~E$_2$K~~> | [s$^{*}$ɛibɨ] | g. | *snap* | ~~E$_2$K~~> | [sɨnɛp̚] |
| c. | *seep* | ~~E$_2$K~~> | [ɕ$^{*}$ip$^{h}$ɨ] | h. | *slow* | ~~E$_2$K~~> | [sɨlːo] |
| d. | *solo* | ~~E$_2$K~~> | [s$^{*}$olːo] | i. | *skip* | ~~E$_2$K~~> | [sɨk$^{h}$ip̚] |
| e. | *super* | ~~E$_2$K~~> | [s$^{*}$up$^{h}$a] | j. | *spin* | ~~E$_2$K~~> | [sɨp$^{h}$in] |

As evident from (16a-e), when a word-initial /s/ is prevocalic, it is adapted as a *tense* [s$^{*}$] in Korean, but when it is preconsonantal (16f-j), it is adapted as a lax [s]. (Epenthesis also applies in 16f-j to repair the ill-formed onset cluster.)

This kind of contextual variation cannot be accounted for by the phonological account. English /s/ is supposed to be a single phoneme, so it should be mapped systematically to Korean /s/, or to Korean /s$^{*}$/. On the other hand, this kind of contextual variability is exactly what is expected under the perceptual account, *provided that* there is independent evidence that English *s / #__C* is perceptually closer to [s] while English *s / #__V* is perceptually closer to Korean [s$^{*}$]. Some suggestive data is given in de Jong & Cho (2012).

### II.5.4 Context-specific adaptation 2: word-final stops

As already shown in (16c,i), English word-final voiceless stops may be variable adapted with or without an epenthetic vowel. Kang (2003) hypothesized that vowel epenthesis in this context was conditioned by the presence/absence of a stop burst release. As noted above, the phonetic grammar of Korean requires that final stops be unreleased. When a final stop burst is released, it is phonetically consistent with a phonological parse with a final vowel that has undergone devoicing.

(17) [[sip̚]] ~~E$_2$K~~> [ɕ$^{*}$ip̚] <-- /s$^{*}$ip/

[[sip]] ~~E$_2$K~~> [ɕ$^{*}$ip$^{h}$ɨ̥] <-- /s$^{*}$ip$^{h}$ɨ/
*~E$_2$K~~> *[ɕ$^{*}$ip] **

Kang investigated the frequency of audible bursts in English source words. The probability of a burst was strongly correlated with the likelihood of epenthesis in the corresponding Korean loanword.

### II.5.5 Context-specific adaptation 3: Let's have a *picnic*

As noted above, medial stop-nasal clusters are illicit in Korean. The native repair is to nasalize the stop. However, when this configuration occurs in simplex source loanwords, it is adapted with vowel epenthesis. In fact, it is more generally the case in this language that illegal native sequences are repaired by assimilation, while illegal loanword sequences are repaired by epenthesis (data from Daland, Oh, & Davidson, under revision)

| (18) | cluster | example | native phonological processes |
|---|---|---|---|
| a. | stop-stop | /supʰ-pʰul/ [sup̚pʰul] 'forest-grass' | laryngeal neutralization |
| b. | stop-nasal | /kuk-min/ [kuŋ̚min] 'nation' | nasality assimilation |
| c. | fric-stop | /pus-tʰoŋ/ [put̚tʰoŋ] 'brush case' | neutralization (laryngeal, manner) |

| | cluster | example (NAKL, 1991) | loanword adaptation pattern |
|---|---|---|---|
| d. | stop-stop | *cha*[pt]*er*$_{English}$ ~> [cʰɛp̚tʰʌ]$_{Korean}$ | faithful (laryngeal neutralization?) |
| e. | stop-nasal | *pi*[kn]*ic*$_{English}$ ~> [pʰikʰɨnik̚]$_{Korean}$ | vowel epenthesis |
| f. | fric-stop | *di*[sk]*ette*$_{English}$ ~> [tisɨkʰɛt̚]$_{Korean}$ | vowel epenthesis |

Note that while voiceless stops in *absolute* coda position are adapted variably (17), voiceless stops in word-*medial* coda position are adapted categorically -- epenthesis before nasals, assimilation before stops.

### II.5.6 How do the theories fare so far?

*Within-context variability* is not well-predicted by the phonological account. There is no story about why one adaptation is chosen sometimes, and another another. However, the perceptual account does have a coherent story on this: listeners are attuned to the phonetic properties of their native language, and therefore attend to phonetic details which are sub-phonemic in the borrowing language (Davidson & Shaw, 2011).

*Between-context variability* is only sometimes predicted by the phonological account. (*Exercise*: Determine whether the phonological account predicts vowel epenthesis in the adaptation of *picnic*, *diskette*, *chapter*, and *star*.) However it is straightforwardly predicted by the perceptual account.

*Within-context categoriality* is only sometimes predicted by the perceptual account. For example, to the extent that medial coda stops are like final coda stops, we should expect variability in word-medial epenthesis with coda stops (*pack* : [pʰɛk̚]~[pʰɛkʰɨ] :: *picnic* : [pʰikʰɨnik̚]~*[pʰiŋnik̚]). Davidson (2006) found that English word-medial coda stops are released about 30% of the time in American English. If loanword adaptation is truly determined only by perception at the moment of adaptation, the rate of vowel epenthesis in medial clusters should be close to 30% (just as the vowel epenthesis rate in word-*final* stops is conditioned by the stop burst release rate). Actually, it is 0% for stop-stop clusters, and 100% for stop-nasal clusters. That is a pretty striking difference. Another example is the adaptation of voiceless stops. They are 100% adapted as aspirated stops (except word-finally, where the contrast between aspirated and lax is neutralized when there is not epenthesis, and therefore unobservable).

In short, each theory does moderately well on predicting general tendencies in the data, but there are particular aspects of loanword adaptation that are troubling for each theory. To get more divergent predictions, we have to go beyond observable loanwords, and into the interesting land of perception.

### II.6 Korean speech perception

According to the perceptual account, *all* (or most?) deviations between the source and borrowed form are driven by perception. Therefore, the loanword adaptation system should transparently mirror the speech perception facts. This yields two crucial predictions.

1. If the Dupoux *ebzo/ebuzo* illusory vowel story is right, then every time you see an epenthetic

vowel in loanwords, you should also get a failure to discriminate the presence/absence of a vowel in the same configuration during online adaptation.

2. If loanword adaptation is actually driven by perceptual similarity (subject to phonotactic legality), then every time that the loanword repair is different than the native repair, it should be because the loanword repair is more perceptually similar to the source form than the native repair is.

Both these predictions can be tested in Korean, with items like *pakna*.

### II.7.1 Daland, Oh, & Davidson, Experiment 1

Daland, Oh, & Davidson (under revision) conducted a cross-linguistic AX discrimination task. Both American English native listeners and Korean native listeners were tested. The *base* stimuli consisted of nonce words like *pakna* -- potential source forms. The *repair* stimuli consisted of logically imaginable repairs of the base stimuli which would render them phonotactically licit (or nearly so) in Korean. Two types of repair were considered: native/assimilation, and loanword/epenthesis. Critical AX trials consisted of a *base*-*repair* pair, e.g. [pʰakna] ~ [pʰakʰɨna] (epenthetic repair) or [pʰakna] ~ [pʰaŋna] (assimilatory repair). In addition to the medial cluster type, the presence/absence of audible frication (including an audible release burst) was manipulated. The results are shown in DOD's Fig. 2:

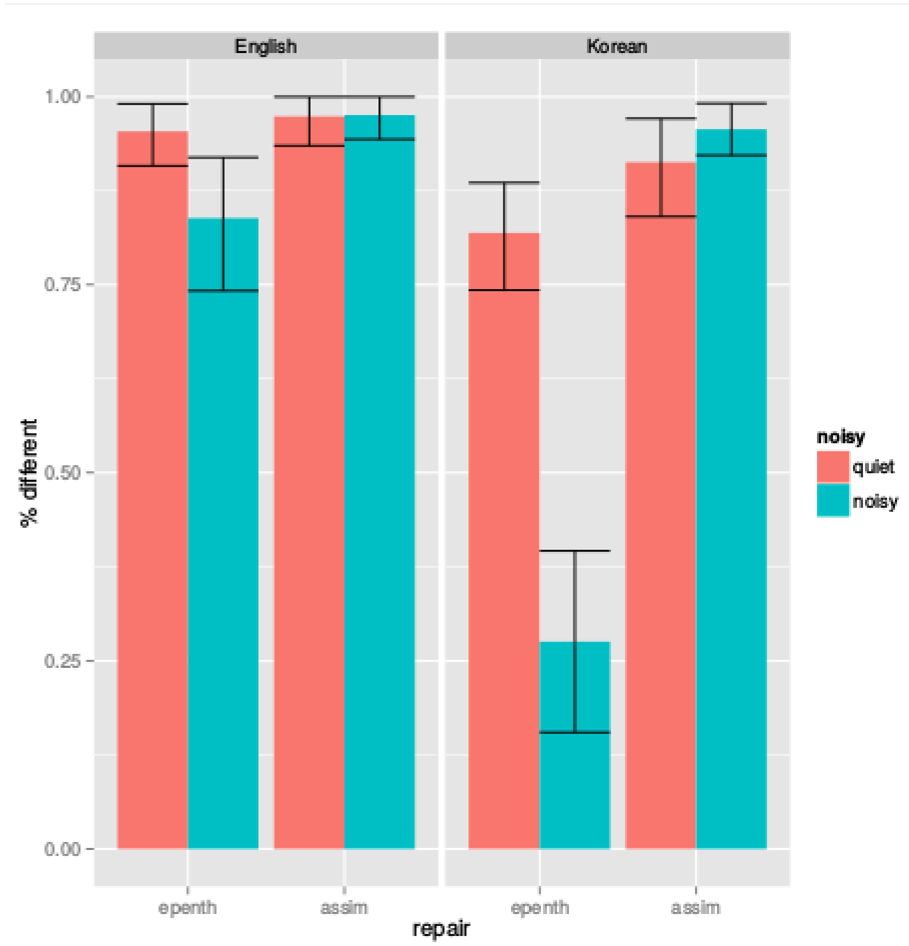


Fig. 2 shows the percentage of 'different' responses for critical stop-nasal trials as a function of listener language, repair type, and presence of an audible release burst.

The *ebzo/ebuzo* effect was replicated in the Korean/noisy/epenthesis column. When there was an audible release burst, Korean listeners exhibited very poor discrimination of the vowel/zero contrast ([pʰakna] ~ [pʰakʰɨna]: *).

However, when there was *not* an audible release burst, Koreans exhibited excellent discrimination of the same vowel/zero contrast ([pʰak̚na] ~ [pʰakʰɨna]: √)

This indicates that the audible release burst plays a causative role in the illusory vowel effect.

These results support the perceptual account, by indicating that when the appropriate phonetic cues are present, they can *force* perceptual epenthesis. In this case, the presence of an audible release burst or other frication is sufficient to force perceptual epenthesis on its own, as evident from the dramatic decline in discrimination in native Korean listeners. Note further that Koreans exhibit excellent discrimination of [pʰak̚na]~[pʰaŋna], although this is the expected native repair.

However, these results also argue against the strongest interpretation of the perceptual account in prediction 1 above, that all epenthetic vowels in loanword adaptation are caused by perceptual epenthesis. As we saw already, *picnic*-type items are always adapted with epenthesis, regardless of whether the medial burst is released. The data in Fig. 2 shows that Koreans exhibit excellent discrimination of pairs like [pʰak̚na] ~ [pʰakʰɨna], even though the former item is always repaired by epenthesis in loanword adaptation. Since Koreans are able to discriminate these items, perceptual epenthesis must not have occurred. That is, there are cases in which Koreans epenthesize during loanword adaptation, even when they are not forced to do so because of perceptual epenthesis. *Perceptual epenthesis can be forced by particular phonetic cues, but not every epenthetic vowel in loanword adaptation results from perceptual epenthesis.*

### II.7.2 Daland, Oh, & Davidson, Experiment 2

The other prediction to test is that the repair that is actually chosen really is the most perceptually similar one. To test this, DOD presented Korean listeners with ABX triples, and asked them to judge whether X was more similar to A, or B. The A and B items represented epenthetic and assimilatory repairs (counterbalanced, of course), while the X item represented a putative source form. The same items were used as in DOD Exp't 1; trials are schematized below:

| (19) | | | A | B | X |
|---|---|---|---|---|---|
| | a. | [+noisy] | [pʰakʰɨna] | [pʰaŋna] | [pʰakna] |
| | b. | [-noisy] | [pʰakʰɨna] | [pʰaŋna] | [pʰak̚na] |

The results were near-categorical. Koreans judged [+noisy]X items (like 19a) as more perceptually similar to the epenthetic repair, which is consistent with an audible release burst causing an illusory vowel. However, they judged [-noisy] X items (like 19b) as more perceptually similar to the assimilatory repair, even though they actually apply the epenthetic repair in loanword adpatation. This finding falsifies prediction 2: *sometimes speakers apply an epenthetic vowel to loanwords even when they are not forced to by perceptual factors; even when they themselves judge the result as less perceptually similar to the source form than the native repair is*.

## II.8 Conclusions as to the relation between speech perception and loanword adaptation

Collectively, the Korean data point to a Goldilocksian theory of loanword adaptation. As claimed by Haugen (1951) and documented in, e.g., Kang (2010, 2012), loanword adaptation begins as largely 'phonetic'/perceptual. Gradually, as language contact increases, individuals and societies move toward a crystallized system for loanword adaptation which *largely mirrors* the speech perception facts, but eliminates certain kinds of within-category variability, largely in the direction predicted by phonological adaptation. The speech perception system of adapters is partly determined by the native phonotactic structure; for example if the native phonetic grammar bans audible release bursts in stop codas, source stop codas with audible release busts will inevitably trigger the percept of an epenthetic vowel. That is, the perceptual account is right in that sometimes, illusory vowels are perceived at the moment of perception. However, there are many configurations in which loanwords receive an epenthetic vowel, even when laboratory studies suggest there is no perceptual epenthesis. These epenthetic vowels seem to occur in order to prevent the borrowed form from deviating too strongly from the source form, a form of the Preservation Principle. In the case of Korean, this happens exactly in the case where listeners are quite able to discriminate the native/assimilatory repair from an unrepaired form, suggesting perhaps that they are aware that some repairs are driven by production or cognitive factors rather than perception.

**FIN**

## References

1. Best, C. T., McRoberts, G. W., & Sithole, N. T. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345-360.
2. Bradlow, A., Akahane-Yamada, R., Pisoni, D., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. Perception and Psychophysics, 61(5), 977-985.
3. Broselow, E., Chen, S.-I., Wang, C. (1998) The Emergence of the Unmarked in second language phonology. *Studies in Second Language Acquisition, 20*(2), 251-280.
4. Carlisle R (1991a) The influence of environment on vowel epenthesis in Spanish/English interphonology. Applied Linguistics 12(1): 76-95.
5. Daland, R., Oh, M., & Davidson, L. (under revision). On the relation between online perception and loanword adaptation: Evidence from the perception of illicit word-medial clusters. *Natural Language and Linguistic Theory*.
6. Daland & Norrmann (accepted). Phonetic evidence for the resyllabification account of vowel prothesis in Spanish speakers acquiring English [s]-consonant clusters. *Open Linguistics*.
7. Davidson L (2006) Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. Cognitive Science 30(5), 837-862.
8. Davidson, Lisa & Jason Shaw. 2012. Sources of illusion in consonant cluster perception. Journal of Phonetics 40(2). 234-248.
9. Della Pietra, Stephen, Vincent J. Della Pietra, & John D. Lafferty (1997). Inducing features of random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**. 380-393.
10. Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology-human Perception and Performance, 25(6), 1568-1578.
11. Flege, J. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics, 15*, 47-65.
12. Flege, J. E. (2005). Origins and development of the Speech Learning Model. First ASA workshop on L2 Speech Learning, Simon Fraser University, Vancouver, B.C., April 14-15, 2005.
13. Flege, J. E., Schirru, C. & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. Speech Communication 40(4): 467-491.
14. Ganong, W. F. (1980). Phonetic categorization in auditory perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.
15. Maria Giavazzi, Robert Daland, Sharon Peperkamp, Laëtitia Debernard, Stefano Palminteri, Pierre Brugières, Catherine Schramm, Laurent Cleret de Langavant, Anne-Catherine Bachoud-Lévi (submitted). Selecting alternatives: the role of the striatum in language. Nature Neuroscience.
16. Goldrick, M. (2002). Patterns in sound, patterns in mind: Phonological regularities in speech production. PhD dissertation, Johns Hopkins University.
17. Goldwater, Sharon & Mark Johnson (2003). Learning OT constraint rankings using a Maximum Entropy model. In Jennifer Spenador, Anders Eriksson & Osten Dahl (eds.) *Proceedings of the Stockholm Workshop on Variation within Optimality Theory*. Stockholm: Stockholm University. 111-120.
18. Haugen, I. (1951) The analysis of linguistic borrowings. Language, 26, 211-31
19. Hayes, B. & Wilson C. (2008) A maximum entropy model of phonotactics and phonotactic learning. Linguistic Inquiry 39: 379-440.
20. Hillenbrand, J.M., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97, 3099-3111.
21. de Jong, K.J., and M.H. Cho (2012). Loanword phonology and perceptual mapping: Comparing two corpora of Korean contact with English. Language, 88(2). 341-368. doi:10.1353/lan.2012.0035.
22. Kang, Yoonjung. 2003. Perceptual similarity in loanword adaptation: English postvocalic word-final stops in Korean. Phonology 20. 219-273. doi: 10.1017/S0952675703004524
23. Kang, Y. (2010). The emergence of phonological adaptation from phonetic adaptation: English loanwords in Korean. Phonology 27, 225-253.
24. Kang, Y. (2012). The adaptation of English liquids in contemporary Korean: a diachronic study. Catalan Journal of Linguistics 11, 41-63.
25. Kazanina, N., Phillips, C. & Idsardi, W. J. (2006). The influence of meaning on the perception of speech sound contrasts. Proceedings of the National Academy of Sciences, 103, 11381-11386.
26. Kenstowicz, Michael. 2007. Salience and similarity in loanword adaptation: a case study from Fijian. Language Sciences 29(2-3). 316-340.
27. *Kwulipkwukeyenkwuwen [The National Academy for the Korean Language]. 1991. Oylaye sayong siltay cosa:*

*1990 nyendo [Survey of the state of loanword usage: 1990]. Seoul: NAKL.*

28. LaCharité, Carole & Darlene Paradis. 2005. Category preservation and proximity versus phonetic approximation in loanword adaptation. Linguistic Inquiry 36(2). 223-258.
29. Miyawaki, K.; Strange, W.; Verbrugge, R.R.; Liberman, A.M; Jenkins, J.J.; Fujimura, O. (1975). An effect of linguistic experience: the discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18(5),* 331–340.
30. Oh, Mira. 2012. Adaptation of English complex words into Korean. Journal of East Asian Linguistics 21. 267-304.
31. Paradis, Carole & Darlene LaCharité. 1997. Preservation and minimality in loanword adaptation. Journal of Linguistics 33(2). 379-430.
32. Parlato-Oliveira, E., Christophe, A., Hirose, Y. & Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. Journal of the Acoustical Society of America, 127(6), 3738-3748.
33. Pisoni, D. (1977). Identification and discrimination of the relative onset time of two component tone: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61(5), 1352-1361.*
34. Wang, C. (1995). The acquisition of English word-final obstruents by Chinese speakers. Unpublished doctoral dissertation, State University of New York at Stony Brook, NY.