1 Observations and questions about epenthesis

(1) Epenthetic vowels
   a. Observations
      • Epenthetic vowels are usually drawn from a small set — most commonly i or a, more rarely e or a. These same vowels are also common in phonemic systems.
      • Sometimes the choice of epenthetic vowel is determined by context; the epenthetic vowel may copy a nearby vowel (e.g., Makassarese) or it may be assimilated to an adjacent consonant (e.g., epenthize i, except u after a labial p/b/m). These same contextual processes are also very common in phonology generally, affecting non-epenthetic vowels.
   b. Questions
      • What does the empty-nucleus/FILL\textsuperscript{Nuc} theory of nucleus epenthesis appear to say about these observations? What about the Correspondence/DEP-V theory of vowel epenthesis?

(2) Epenthetic consonants
   a. Observations
      • Epenthetic consonants are usually drawn from a small set — most commonly ?, more rarely t or h. These same consonants are also common in phonemic systems.
      • Sometimes the choice of epenthetic consonant is determined by context — the epenthetic consonant may be assimilated to an adjacent vowel (e.g., /ua/ → [uwa], /ia/ → [iya]) or an adjacent consonant (this happens in Lardil, for instance; see P&S 1993/2004, Ch. 7).
   b. Questions
      • What does the empty-onset/FILL\textsuperscript{Ons} theory of onset epenthesis appear to say about these observations? What about the Correspondence/DEP-C theory of consonant epenthesis?

(3) Other phonological participation of epenthetic segments
   a. Observation
      • Often, an epenthetic segment is observed to trigger phonological alternations in just the same way as an underlying segment of the same type does; e.g., both underlying and epenthetic i may be observed to palatalize a preceding consonant: /pati/ → [paʧi], /bat/ → [baʧi].
   b. Questions
      • What does the empty-nucleus/FILL\textsuperscript{Nuc} theory of nucleus epenthesis appear to say about this observation? What about the Correspondence/DEP-V theory of vowel epenthesis?

(4) Additional question
   • Under the unparsed segment/PARSE theory of deletion, should a deleted segment be visible or invisible to phonological processes? For instance, if t palatalizes when immediately followed by i, what do we expect to happen in pat(\text{k})i?

(5) In the PARSE/FILL model, the quality of epenthetic segments is not under phonological control; it is attributed to a separate component which interprets the output of the phonology by, among other things, spelling-out empty syllabic nodes as particular segments. In the Correspondence model, epenthetic segments are literally present in output candidates, and so their quality (i.e., the choice of what segment to epenthize) is at least potentially under phonological control. This handout addresses the results and benefits of the Correspondence model in this regard.
2 Basic Theory

OT asserts that a grammar is a ranking of the constraints of UG, and so every constraint is present in the grammar of every language. We have so far been mostly concerned with situations where some markedness constraint \( M \) crucially dominates some relevant faithfulness constraint \( F \) in a language \( L \):

\[
M \gg F \quad \text{in } L
\]

With this ranking, \( M \) has some control over the outputs in \( L \), and so \( M \) leads to alternations, overt restrictions on \( L \)'s surface structures, and so on. For instance, if \( M \) is NOCODA and \( F \) is PARSE/MAX, then \( L \) will not have codas. In situations like (6), we say that \( M \) is inventory-defining for \( L \). The inventory of \( L \) is the set of all segment-types, with their distributional restrictions, syllable-types, and other structures permitted by \( L \). Configurations that violate \( M \) are not found in \( L \)'s inventory.\(^1\) But UG contains many other markedness constraints. Consider a markedness constraint \( M' \) which is crucially dominated by all relevant faithfulness constraints \( F \) in \( L \):

\[
F \gg M' \quad \text{in } L
\]

\( M' \) will not be inventory-defining for \( L \). Observed configurations in \( L \)'s inventory will violate \( M' \) with impunity, because of the ranking (2). Does this mean that \( M' \) is completely inactive in \( L \)? In parametric theories, this question does not arise; the parameter corresponding to \( M' \) is turned off and so it could not have any activity anywhere in \( L \). But as you know from some of our previous discussions, it is actually rather hard to ensure that a constraint is totally inactive in OT. Indeed, activity by \( M' \) plays a central role in an OT mode of explanation called the emergence of the unmarked (TETU).\(^2\)

3 Exemplification

Suppose that UG contains the following constraints:

\[
\begin{align*}
(8) & \quad \text{a. Markedness} \\
& \quad *[-\text{bk},+\text{rd}] \equiv [-\text{back}] \text{ iff } [-\text{round}] / [+\text{round}] \text{ iff } [+\text{back}] \text{ (i.e., } *\ddot{u}, *\ddot{o}, \text{ etc.)} \\
& \quad \text{b. Faithfulness} \\
& \quad \text{IDENT(rd)} = \text{if } s_i \not\in s_o \text{, then } s_i \text{ and } s_o \text{ have identical values for } [+\text{round}]. \\
& \quad \text{IDENT(bk)} = \text{if } s_i \not\in s_o \text{, then } s_i \text{ and } s_o \text{ have identical values for } [+\text{back}].
\end{align*}
\]

Ranked as in (9), these constraints characterize a language that bars front rounded vowels from its inventory, such as English or Spanish or Japanese. Ranked as in (10), however, they characterize a language that has front rounded vowels in its inventory, such as French or German or Turkish.

\[
\begin{align*}
(9) & \quad \{ *[-\text{bk},+\text{rd}], \text{IDENT(bk)} \} \gg \text{IDENT(rd)} \quad \text{or} \quad \{ *[-\text{bk},+\text{rd}], \text{IDENT(rd)} \} \gg \text{IDENT(bk)} \\
(10) & \quad \{ \text{IDENT(rd)}, \text{IDENT(bk)} \} \gg *[-\text{bk},+\text{rd}]
\end{align*}
\]

It may seem that in the latter case the constraint *[-bk,+rd] is completely inactive in (10), but that is incorrect. Suppose the same language has a process of vowel epenthesis because of the ranking NOCODA \( \gg \) DEP. An epenthetic segment has no underlying correspondent, and so IDENT(rd)(bk) are irrelevant to it. *[-bk,+rd] is free to emerge in that situation, ensuring that the epenthetic vowel will not be front rounded — even though front rounded vowels exist in the language as a whole!

This is “the emergence of the unmarked” or TETU. A markedness constraint that is crucially dominated by a faithfulness constraint is violated generally in the language, but it emerges whenever that faithfulness constraint is irrelevant to the optimization at hand. Featural faithfulness constraints are

---

\(^1\) Unless some higher-ranking constraint mitigates the force of \( M \) in some situation; e.g., NOCODA \( \gg \) MAX in Diola Fogny, forbidding syllables with codas generally, but ALIGN-R \( \gg \) NOCODA, permitting syllables with codas in morpheme-final position.

usually irrelevant in cases of segmental epenthesis, and so even low-ranking featural markedness constraints can be emergent in determining which segment is epenthesized.

- This model predicts that an epenthetic segment must always be chosen from the inventory of the language as a whole. Why?
- This model predicts that restrictions on epenthetic segments in one language will hold for the whole inventory in another language. Why?
- Infixed in Tagalog (infix um with consonant-initial stems; otherwise prefix) is also a case of the emergence of the unmarked. Why?

### 4 Typology

Certain strong cross-linguistic tendencies can be observed concerning the quality of epenthetic segments, be they consonants (11) or vowels (12). These can be divided into two large classes: non-contextually determined (a) and contextually determined (b).

#### (11) Epenthetic consonants

- **Non-contextually determined**
  - Except when their identity is determined by nearby context, epenthetic consonants are usually ʔ, or more rarely a coronal like t, n, or r.

- **Contextually determined**
  - The identity of epenthetic consonants may be determined by nearby context. This is usually a kind of assimilatory effect, as in Lardil, where the epenthetic consonant can even be k, in kaŋka, but only because of the preceding ŋ. A more common example of this type involves epenthetic glides. A typical pattern, seen in many languages (e.g., Tamil), is to get epenthetic homorganic glides after non-low vowels but epenthetic ʔ after a:
    
    
    ```
    /i+a/ → iya   /u+a/ → uwa
    /e+a/ → eya   /o+a/ → owa
    /a+a/ → aʔa
    ```

#### (12) Epenthetic vowels

- **Non-contextually determined**
  - Except when their identity is determined by nearby context, epenthetic vowels are always [−round], and if [−low], [−back] as well. The most common epenthetic vowels are probably i (Palestinian Arabic, Yoruba) and o (Hebrew, Moroccan Arabic, Hindi). Others include e (Spanish) and a (Saudi Arabic, Lardil, Axininca Campa).

- **Contextually determined**
  - The identity of epenthetic vowels may be determined by nearby context. Adjacent consonants may affect the quality of the epenthetic vowel, or it may simply copy some or all of the features of a vowel in an adjoining syllable (Winnebago, Makassarese):
    
    ```
    /iC/ → iCi  /uC/ → uCu
    /eC/ → eCe  /oC/ → oCo
    /aC/ → aCa
    ```

The theory presented here entails certain claims about the relation between (non-contextual) epenthesis and inventories. Because of ranking permutation, the following correlations should hold:

- Any universally-observed restriction on the quality of an epenthetic segment entails a similar implicational universal in segmental inventories.
- Any implicational universal about segmental inventories entails a similar universal restriction on the quality of epenthetic segments, *mutatis mutandis*.
Non-contextual determination of epenthetic consonants is unmarkedness in the classic Jakobsonian sense. For Jakobson, certain types of segments are unmarked non-contextually, and it is these segments that are the first to appear in the developing grammars of children, the last to depart in the degrading grammars of aphasics, and the most commonly found in the languages of the world. Contextual determination of epenthetic consonants is unmarkedness in the extended sense used in OT. Where Jakobson was evaluating segmental inventories in the abstract, OT evaluates candidate words, which bring with them a context in which the epenthetic consonant finds itself.

Below I’ll explain how both the contextual versus non-contextual types are obtained in OT; right now I want to address a different typological question about the non-contextual type. The argument presented in this handout entails certain claims about the relation between (non-contextual) epenthesis and inventories. In principle, because of ranking permutation, the following correlations should hold:

- Any universally-observed restriction on the quality of an epenthetic segment entails a similar implicational universal in segmental inventories. For instance, front rounded vowels are never epenthesized; therefore the presence of front rounded vowels in an inventory should imply the presence of front unrounded and/or back rounded vowels of the same height in that inventory (and not vice-versa).
- Any implicational universal about segmental inventories entails a similar universal restriction on the quality of epenthetic segments, mutatis mutandis.

5 Applications

I’ll present two examples here, one with context-free and the other with context-sensitive determination of the quality of the epenthetic segment. In both cases, I assume that the constraint(s) and ranking(s) compelling DEP violation has already been established.

Palestinian Arabic has a three-vowel vowel inventory: i, u, and a. This inventory is defined by a set of undominated markedness constraints that crucially dominate some appropriate faithfulness constraints. These markedness constraints include *[-hi, -lo] “[-high] iff [+low] / [-low] iff [+high]”, which rules out all mid vowels, and other constraints dealing with rounding and backing. From derivations like /dārs/ → daris ‘lesson’, we see that i is epenthesized. The vowel i, then, must be less marked under the language-particular ranking of Palestinian than either u or a:

\[(13) \ *\text{LOW} >> *\text{HIGH}\]

<table>
<thead>
<tr>
<th>Input: /d₁a₂r₃s₄/</th>
<th>*LOW</th>
<th>*HIGH</th>
<th>*ROUND</th>
<th>*SPREAD (?)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. \ d₁a₂r₃i s₄</td>
<td>*</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. \ d₁a₂r₃a s₄</td>
<td>**!</td>
<td></td>
<td></td>
<td>**</td>
</tr>
<tr>
<td>c. \ d₁a₂r₃u s₄</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
</tbody>
</table>

Observe that I have assigned marks for both vowels, the underlying one and the epenthetic one. The marks for the underlying vowel are shared by all the candidates considered here, so they can be cancelled out, having no effect on the outcome. Thus, Palestinian Arabic epenthesizes a high unrounded vowel rather than a low or rounded one because a high unrounded vowel is less marked in the literal sense that it receives the lowest-ranking (or fewest) marks for these constraints (Smolensky 1993). Another vowel would violate higher-ranking constraints, and so it could not be optimal.

---


So then, why doesn’t every vowel turn into i, if i is such a great vowel — why don’t we get /dars/ → *diris? The answer is that i’s unmarkedness is an emergent property. Faithfulness constraints — MAX-V, IDENT(hi), IDENT(lo), etc. — crucially dominate the markedness constraints that favor i.

(14) \{ MAX-V, IDENT(hi/lo) \} >> *LOW

<table>
<thead>
<tr>
<th>/d₁a₂r₃s₄/</th>
<th>MAX-V</th>
<th>IDENT(hi/lo)</th>
<th>*LOW</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. *d₁a₂r₃i s₄</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. d₁i r₃i s₄</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. d₁i₂r₃i s₄</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

A similar argument can be made for \{ MAX-V, IDENT(rd) \} >> *ROUND.

In summary, epenthesis of i in Palestinian Arabic is an instance of the emergence of the unmarked. The markedness constraints *ROUND and *LOW emerge to determine what the epenthetic vowel is, even though they aren’t obeyed in the language generally.

In the context-sensitive epenthesis cases, the quality of the epenthetic segment is predictable from the quality of some other nearby segment. For instance, Winnebago has vowel-copying epenthesis:

(15) Winnebago

<table>
<thead>
<tr>
<th>Underlying</th>
<th>Derived</th>
</tr>
</thead>
<tbody>
<tr>
<td>/hoikwe/</td>
<td>hoikewe</td>
</tr>
<tr>
<td>/krepna/</td>
<td>krepna</td>
</tr>
<tr>
<td>/pras/</td>
<td>paras</td>
</tr>
<tr>
<td>/rupri/</td>
<td>rupri</td>
</tr>
<tr>
<td>/xwuxwu/</td>
<td>xwuxwu</td>
</tr>
</tbody>
</table>

Winnebago has a five-vowel system i, e, a, o, and u, and so the markedness constraints against the vowels in this inventory, such as *[–hi,–lo], *LOW, *HIGH and so on, are all crucially dominated by relevant faithfulness constraints. No matter how these three constraints are ranked with respect to one another, however, they will pick the wrong epenthetic vowel in at least some words. And if they’re crucially unranked, they will give a variable or indeterminate output, picking the wrong epenthetic vowel sometimes in all words. These markedness constraints alone don’t seem to be enough to give us the contextual determination that’s typical of Winnebago.

A key insight is that context-sensitive epenthesis may not be some kind of unique phenomenon; it may just be a special, emergent case of that most common of phonological processes: assimilation (in this case, complete vowel harmony). Suppose there is a constraint which, through crucial domination of faithfulness, causes ordinary, inventory-defining assimilation in language L; put that same constraint in an emergence of the unmarked ranking in L′, and you’ve got context-sensitive epenthesis.
On the other hand, total vowel harmony like this is not terribly well-attested outside of vowel copy epenthesis! An alternative: vowel copy epenthesis could be viewed as breaking (a violation of INTEGRITY; roughly: “a single input segment corresponds to a single output segment”) instead.

Another alternative involves assimilation, but less directly so.\(^5\) Suppose that a single bundle of features can be shared by two separate vowels. Suppose also that the markedness constraints \(*[-hi, -lo]\), \(*LOW\), \(*HIGH\), etc. evaluate each autosegmental feature bundle rather than each individual segment that is associated with that bundle. It’s clear that any independent epenthetic vowel is more marked than any linked one, regardless of the relative ranking of the markedness constraints:

Under this interpretation of markedness, epenthesis by spreading adds no markedness violations. Epenthesis by insertion of a different vowel, on the other hand, \textit{always} adds some marks. This means that epenthesis by spreading should always occur, regardless of the ranking of markedness constraints. That’s obviously not true, so to be typologically complete we would require another constraint that, through high rank, rules out the spreading option in languages with fixed epenthetic vowels and consonants: a constraint against spread structure (or against particular kinds of spread structure like V-to-V across C), banning candidates like (a), will be high-ranking in such languages.

---