Perception in Optimality Theory: The frugality of the base

Patricia Donegan and Paul Lassettre
University of Hawai‘i at Manoa

One thread of the Optimality Theory (OT, Prince and Smolensky 1993) literature on acquisition analyzes how children accurately perceive sounds they cannot yet accurately pronounce (Gnanadesikan 1995, Smolensky 1996). The elegant OT analysis uses a single grammar for perception and production. However, the OT account of perception seems to predict that adults will perceive phonetic distinctions with high accuracy, regardless of what those adults can produce. Any adult who has ever studied an unfamiliar language knows that this is not true: distinctions that are perfectly clear to speakers of one language can be incomprehensible to others.

Smolensky (1996) notes this, and, in passing, proposes that adults cannot reliably perceive some phonetic distinctions because they cannot use those distinctions to retrieve different lexical items. If speakers do not use some feature distinctively in their own language, that feature will be underspecified in lexical representations. Perceptual mergers occur when phonetically distinct items match the same underspecified lexical entry.

We agree that perceptual mergers result from mapping different surface structures onto a single representation. However, we argue that underspecification does not answer the original objection, and propose an alternate model.

In the standard model, structural well-formedness constraints evaluate only surface representations. In our model, perceptually grounded structural constraints evaluate the underlying representation as well. The result is a grammar that maps perceptually homophonous surface representations onto identical underlying representations. Our model accounts for the facts of both child and adult phonological perception in terms of constraint ranking, without recourse to a third layer of representation.


Smolensky 1996 presents a model of linguistic perception, building on previous work such as Prince and Smolensky (1993:175–196) and Gnanadesikan (1995). Smolensky’s model accounts for the widely-noted discrepancy between linguistic production and perception in young children, who are able to hear the difference between sounds that they cannot pronounce distinctly, using a single grammar for both production and perception.

The model makes crucial use of two representations, UNDERLYING REPRESENTATION (UR) and SURFACE REPRESENTATION (SR). Production is modeled as mapping a given UR onto an unknown SR, and perception is modeled as mapping a given SR onto an unknown UR. The discrepancy between children’s production and perception corresponds in the model to a more accurate mapping from SR to UR than from UR to SR.
An optimality theoretic grammar evaluates candidates for whichever representation is to be discovered. Two types of constraints are used, **Structural Constraints** and **Faithfulness Constraints**. Structural constraints evaluate the well-formedness of SRs only. Faithfulness constraints evaluate URs and SRs with respect to each other; cases where UR and SR differ are less optimal than where UR and SR are alike. Thus, the evaluation function does not treat UR and SR equally: UR is evaluated only with respect to its faithfulness to SR, while SR is evaluated with respect both to its faithfulness to UR and to its inherent well-formedness. Moreover, in the early stages of acquisition, structural constraints are assumed to outrank faithfulness constraints (Tesar and Smolensky 1996). In Smolensky’s model, the unequal ranking of the two types of constraint, and their unequal treatment of the two representations, underlies the discrepancy between perception and production.

Tableau 1 shows the grammar of a child who pronounces *fish* as [fis], but rejects the same pronunciation by adults, according to our interpretation of Smolensky’s model, and using our own notation. The same child pronounces *kiss* as [kis], and accepts that pronunciation by adults. Candidates are shown as ordered pairs, in which the known member of the pair is fixed and the unknown varies freely. We interpret faithfulness as UR-SR correspondence (McCarthy and Prince 1994, McCarthy 1995). MAX /feature/ is violated when the specified feature in UR does not have an identical correspondent in SR, and DEP [feature] is violated when the specified feature in SR does not have an identical correspondent in UR.

In production, the URs /fis/ and /kis/ are given. Candidates are ordered pairs of UR and SR in which the SR varies freely. The candidate set contains all possible SRs but just one type of UR. The highest ranked of the three constraints shown is a structural constraint, which evaluates the SR. The harmonic ordering of the candidate set will therefore be affected by the well-formedness of the SR, and the winning candidates are [kis] and the unfaithful but well-formed [fis].

In perception, the two SRs [fis] and [kis] are given. As above, candidates
are ordered pairs, but now it is the SR that is fixed. The candidate set includes all possible URs but just one type of SR. All candidates incur equal violations of the structural constraints, and therefore the well-formedness of the SR does not affect the harmonic ordering of the candidate set. Faithfulness constraints, although ranked low, are the only constraints that affect the harmonic ordering of the candidate set, and the winning candidate is the one where the UR and the SR are most alike.

The child’s ability to hear the difference between \[f\] and \[s\] is reflected in the model by the faithful mapping from \([fi]\) and \([ks]\) to /fi/ and /ks/. The same child’s inability to pronounce /\(S\)/ distinctively from /\(s\)/ is reflected by the same grammar’s mapping /fi/ onto /fi/.

2. Perception and richness of the base.

Richness of the base is a term by which Prince and Smolensky (1993:191) refer to the basic OT assumption that the candidate set is unlimited. The implication is that systematic gaps and omissions in a speaker’s outputs cannot be explained by gaps and omissions in the input, but must be explained by the phonology itself, in the guise of the optimality theoretic evaluation function. Gaps and omissions in a language’s lexicon reflect the surface patterns, not the other way around. Richness of the base is a reiteration of the fundamental OT claim that the ‘locus of explanatory action’ (Prince and Smolensky 1993:3) is the ranked constraint hierarchy.

For example, Standard Hawaiian has no closed syllables. This is not explained by the fact that Standard Hawaiian has no underlying strings that could be parsed into closed syllables — has no underlying consonant that is not followed by a vowel. Rather, Standard Hawaiian phonology is such that every possible underlying string, including strings that contain such things as consonant clusters and final Cs, will be parsed into surface open syllables. The explanation lies in the constraint ranking, not in the inputs. Any lack of underlying consonant clusters and final consonants is merely a projection of the surface pattern.

Again, Korean has no surface contrast between \([l]\) and \([r]\). The distribution of these liquids is predictable: \([l]\) occurs syllable finally and \([r]\) occurs elsewhere. This is explained if Korean phonology neutralizes /\(l\)/ and /\(r\)/ in all environments. The lack of an underlying contrast between the two liquids is due to their surface neutralization.

However, as Smolensky notes, the neutralization of Korean \([l]\) and \([r]\) is not just neutralization in production. Korean speakers also neutralize \([l]\) and \([r]\) in perception, tending to perceive English bear and bail, for example, as homophones. The OT model with faithful mapping from SR to UR predicts that these sounds will be perceptually as distinct for Korean speakers as \([f]\) and \([s]\) (or \([l]\) and \([r]\)) are for the child acquiring English.

In a footnote, Smolensky proposes that this effect is due to lexical retrieval: There are underspecified LEXICAL REPRESENTATIONS (LRS). Features that
Tableau 2. Korean: faithful URs, homophony at lexical retrieval.

are not used distinctively in a given language are absent from the LR. In lexical retrieval, a UR is matched to an LR. Because of underspecification, some distinct URs may match the same LR. Failure to perceive a distinction is reflected in this model as matching single LR. This is the case with [l] and [r] in Korean: Korean speakers hear bear and bail as homophones because Korean speakers cannot match them with distinct LRs (tableau 2).

This model is indirectly faithful to richness of the base. Perception of homophony — gaps and omissions in the perceptual pattern — is explained by gaps and omissions — underspecified features — in the LR. This is not in the spirit of the strongest interpretation of richness of the base, although gaps and omissions in the LR are ultimately explained in terms of the constraint hierarchy, by way of patterns in SRS.


The underspecified LR model does not account for all perceptual phenomena. For example, under this model the LRs of many American English dialects must include nasal vowels. However, speakers of such dialects do not perceive the large number of contrasts entailed by this analysis, or at least not more reliably than Korean speakers distinguish [l] and [r].

American English speakers ordinarily pronounce say, sate, sane, and saint as [sei], [seɪt], [sɛɪn], and [sɛɪt]. There does not exist any fifth lexical item not homophonous with say, sate, sane, or saint and ordinarily pronounced [sɛɪnt], [sɛi], [sɛin], or [sɛint]. No such lexical item is possible. The phonology determines how nasalized vowels are distributed on the surface, by mapping the eight logically possible URs /sei/, /seɪt/, /sɛɪn/, /sɛɪnt/, /sɛɪ/, /sɛɪnt/, and /sɛɪnt/, onto the four attested SRSs [sei], [seɪt], [sɛɪn], and [sɛɪt].

Some SRSs differ only by a nasalized vowel, and are not neutralized in perception: sate [seɪt] and saint [sɛɪnt]. debt [det] and dent [dɛt], lip [lɪp] and limp [lɪp], duck [dʌk] and dunk [dʌk], and so on. Underlying representations are faithful to SRSs, and LRs are not distinct from URs. Vowel nasalization is specified in LR, or it is underspecified: there is no other option. If vowel nasalization were underspecified in LR, debt and dent and pairs like them would be perceived as homo-
phones (tableau 3). Therefore, vowel nasalization must be specified in LR to account for these minimal pairs.

If this were correct, American English speakers would reliably distinguish the eight srs [sei], [seit], [sein], [sé’ét], [séët], [séën], and [séënt], whether or not they can produce them. We do not believe American English speakers can distinguish all eight any more reliably than Korean speakers can distinguish bear and bail. No scheme of lexical underspecification will condition a realistic pattern of perceptual mergers and distinctions.

The standard OT model predicted that adult perception would be highly accurate, more accurate than adult production. The underspecified LR model was intended to account for cases where this prediction was incorrect. Although it is pleasingly simple, and it accounts for Korean perception of liquids, the underspecified LR model cannot account for more complicated examples of the same type, such as American English perception of vowel nasalization. It does not, therefore, solve the problem it was intended to solve.

4. Systematically unfaithful URS.

With a modification to the way some structural constraints evaluate representations, the nasalization paradox is resolved. We recognize two classes of structural constraints, which can be distinguished formally and functionally, and also by their phonetic grounding. One of these classes of constraint should evaluate UR as well as SR, while the other class should continue to evaluate only SR. There is no need for a separate lexical representation.

One class we call SEGMENTAL WELL-FORMEDNESS (SEGWF) constraints, although we might equally well term them ENHANCEMENT or FORTITIVE constraints. Formally, these constraints refer to simultaneous occurrences, or co-

\[
\begin{array}{ll}
\text{a. } & \langle /\text{det}/,[\text{det}] \rangle \\
\text{b. } & \langle /\text{dët}/,[\text{det}] \rangle \\
\text{c. } & \langle /\text{dët}/,[\text{dët}] \rangle \\
\text{d. } & \langle /\text{dënt}/,[\text{dët}] \rangle \\
\text{e. } & \langle /\text{dënt}/,[\text{dënt}] \rangle \\
\text{f. } & \langle /\text{dënt}/,[\text{dënt}] \rangle \\
\end{array}
\]

\[
\begin{align*}
/d\text{e}_{(-\text{nas})}t/ & \Rightarrow d\text{e}_{(\emptyset \text{nas})}t \\
/d\text{e}_{(+\text{nas})}t/ & \Rightarrow d\text{e}_{(\emptyset \text{nas})}t
\end{align*}
\]
occurrences, of distinctive features, although their exact expression depends on
the phonological theory used. They may be expressed as constraints on licensing
or feature geometry. Our notation is shown in (1).

Empirical evidence for SEGWF constraints comes from context-free phonological phenomena, including crosslinguistic patterning of phoneme inventories, unconditioned historical sound shifts, and context-free stylistic variation, as observed in, say, hyperarticulated speech. SEGWF constraints are phonetically grounded in perception, although this does not rule out articulatory grounding as well. That is, using a method such as Hayes’s inductive grounding (Hayes 1996), these constraints emerge from perceptual distinctness rather than articulatory difficulty.

The other class we call SEQUENTIAL WELL-FORMEDNESS (SEQWF) constraints, although another possible name might be LENITIVE constraints. Formally, these constraints refer to sequences of features (2). Empirical evidence for these constraints comes from context-sensitive phonological phenomena such as assimilation and reduction, and conditioned historical sound shifts. These constraints are grounded in articulation, and function to increase articulatory ease.

(2) *[featureₖ][featureₖ]

SEGWF constraints evaluate URs, while SEQWF constraints evaluate SRs only. The result is that in perception URs are not fully faithful to given SRs, although the mappings from SR to UR and from UR to SR are still asymmetrical. Perception of contrast is modeled as mapping of distinct SRs onto distinct URs, as before. Perception of homophony, however, is modeled as mapping of distinct SRs directly onto the same UR.

Tableau 4 shows how this resolves the English vowel nasalization paradox. The only SEGWF constraint is *V[+nas], which is violated by any nasal vowel. This is sufficient to change the optimal UR given [dēt] from /dēt/ to /dent/, given the ranking DEP [+nas], *V[+nas] ≫ MAX /c/. Thus [det] and [dēt] are associated with distinct URs without allowing underlyingly nasalized vowels in English.

<table>
<thead>
<tr>
<th></th>
<th>DEP [+nas]</th>
<th>MAX /c/</th>
<th>MAX [+nas]</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>⟨/dēt/,[det]⟩</td>
<td><img src="image1.png" alt="image" /></td>
<td>*!</td>
</tr>
<tr>
<td>b.</td>
<td>⟨/dēt/,[det]⟩</td>
<td><img src="image2.png" alt="image" /></td>
<td>*</td>
</tr>
<tr>
<td>c.</td>
<td>⟨/dēt/,[dēt]⟩</td>
<td>*!</td>
<td><img src="image3.png" alt="image" /></td>
</tr>
<tr>
<td>d.</td>
<td>⟨/dēt/,[dēt]⟩</td>
<td><img src="image4.png" alt="image" /></td>
<td>**!</td>
</tr>
<tr>
<td>e.</td>
<td>⟨/dent/,[dent]⟩</td>
<td><img src="image5.png" alt="image" /></td>
<td><img src="image6.png" alt="image" /></td>
</tr>
<tr>
<td>f.</td>
<td>⟨/dent/,[dēt]⟩</td>
<td><img src="image7.png" alt="image" /></td>
<td><img src="image8.png" alt="image" /></td>
</tr>
</tbody>
</table>

Tableau 4: English Perception 2
constraint ranking will lead to all surface occurrences of \( \overline{\text{V}} \) being perceived as \( /\overline{\text{VN}}/ \) by English speakers. This model is capable of handling production and perception of nasalized vowels in English simply and without paradox.

The Korean case is resolved in the same way. \textit{Bear} [ber] and \textit{bail} [bel] are mapped directly onto the same \textit{UR}. Like Smolensky, we model the perception of homophony as mapping onto the same representation. Unlike Smolensky, we derive that mapping directly from an optimality theoretic constraint hierarchy evaluating a candidate set. That phonology neutralizes liquids in both perception and production.

It should be noted that our model allows the possibility of children learning English who cannot hear the difference between \([\text{ʃ}]\) and \([\text{s}]\). If structural constraints outrank faithfulness initially, children will need to rerank constraints in order to perceive the distinction. This seems to be empirically contradicted by numerous reports that very young children are capable of perceiving any phonetic distinction whatsoever. However, such children are very young, under nine months, while the children at the early stages of language discussed by Smolensky 1996 and Gnanadesikan 1995 are over eighteen months. It is not necessary to assume that perception in the two groups of children works exactly the same.

Our model is of linguistic perception, but we assume that both representations, \textit{SR} as well as \textit{UR}, are percepts. The surface representation has already undergone significant cognitive processing, and is not a representation of the acoustic wave. Rather, it is a representation of what is heard. We suggest that very young children perceive all possible phonetic distinctions because they treat...
linguistic stimuli similar to other sounds. They respond only to differences in SR. Adults may be conditioned to do the same thing. Korean speakers may well judge that bear and bail do not sound exactly alike even while unable to apprehend the difference linguistically.

5. Conclusion: Frugality of the base.

Gaps, omissions, and other patterns in URs are not mere projections of surface patterns. The input to perception must be just as unlimited as the input to production: richness of the base is bidirectional. Gaps and omissions in perception can be explained by the grammar itself, as the result of the interaction of faithfulness and perceptually grounded structural constraints. It is for this reason that we say the base also is frugal.

References.

Hayes, Bruce. 1996. Phonetically driven phonology: The role of Optimality Theory and inductive grounding. ROA. UCLA.
McCarthy, John, and Alan Prince. 1994. Faithfulness and reduplicative identity. ROA. University of Massachusetts, Amherst, and Rutgers University.