Bridging the gap between receptive and productive development with minimally violable constraints

Joe Pater, University of Massachusetts, Amherst

1. Introduction

It is a commonplace observation that there are gaps between perception and production in child phonology; that children often appear to have receptively acquired a segmental or prosodic contrast that is neutralised in production (see §2 below for a review of some evidence). The developmental lag of productive behind receptive ability raises a fundamental question: Does perception or production reflect the state of phonological competence in the child's emerging linguistic system? The answer given in this paper is that the child's knowledge of phonology is displayed in both domains, and that linguistic competence can be sufficiently variegated to handle gaps between their development.

Studies of phonological acquisition usually take children's productions as the data to be accounted for by phonological analysis, and studies of child phonology in Optimality Theory usually follow this tradition (see the introduction to this volume for references to this literature; cf. Hale and Reiss 1998, Hayes this volume). However, there do exist data on the development of perception that appear to demand a phonological treatment. The experimental tasks used in recent research in infant speech perception tap not only phonetic discrimination, but also the ability to store contrasts in memory, and to link these phonetic contrasts with meaning distinctions. This research suggests that the representations accessed in these tasks develop gradually, in ways quite parallel to the later growth in complexity of the structures employed in production. The theoretical challenge that arises is to account for the parallels in development across these domains, while at the same time allowing for synchronic differences in whether structures have been acquired productively and receptively.

Optimality Theory is well suited to meet this challenge because its constraints are minimally violable. The activity of a constraint is not an all or nothing affair; rather, a constraint is violated only to meet the demands of a higher ranked constraint. I assume that the role of the phonological grammar in perception is to regulate the complexity, or markedness, of representations that are constructed, and/or accessed, on the basis of the acoustic signal. This is done by having markedness constraints evaluate candidate lexical representations for the perceived surface string (though cf. section 4.2 for an alternative approach).
Parallels in development across perception and production occur because the same markedness constraints apply in both domains.

A gap between receptive and productive competence arises when the perceptual representations are more marked than the representations evinced in production. I propose to capture this by invoking perception specific faithfulness constraints. When such a constraint ranks above the relevant markedness constraint, complexity of perceptual representation along the dimension regulated by these constraints is permitted. In production however, the perception faithfulness constraint does not apply, so that the markedness constraint will be obeyed, and the unmarked structure will emerge.

In what follows, I first discuss evidence of segmental and prosodic phonological development in infant speech perception (section 2), and then account for this perceptual development, and its relation to the later development of production, in terms of the proposed model (section 3). Section 4 compares this proposal with three other ways of accounting for perception/production differences in terms of properties of the phonological grammar: the Optimality Theoretic precedent to the present model in Smolensky (1996), a 'mixed model' that contains some attributes of the Smolensky (1996) proposal and some of the one presented in section 3, and finally, the dual–lexicon model proposed in the acquisition literature reviewed in Menn and Matthei (1992).

2. The development of receptive competence

Gaps between perception and production are almost a logically necessary aspect of the acquisition process. Linguistic input must of course be perceived and given structure before that structure can then be applied in producing new utterances (except for unmarked structures that are available in the initial state). However, such gaps might not exist if grammatical development is in fact monolithic, that is, if the receptive acquisition of a structure entails that it can be immediately employed in production. And even given empirical evidence of perception/production gaps, one might explain them away by claiming that restrictions on receptive and productive ability are entirely separate, so that only one, or the other, is controlled by the grammar.

The aim of this section is to provide some arguments from the study of phonological acquisition against these positions. We will see that early receptive phonology is subject to restrictions strikingly similar to those governing later production, which argues against treating them as products of entirely separate linguistic subsystems. We will also see that restrictions on the complexity of the structures employed in perception are overcome well before the parallel restrictions are overcome in production, thus showing that grammatical development is not monolithic across these domains.
Before proceeding any further, an important point of definition must be addressed: what counts as evidence of phonological acquisition in the domain of perception? Given infants' precocious ability to categorically discriminate between almost any of the segmental contrasts of the world's languages (Eimas, Siqueland, Jusczyk, and Vigorito 1971, et seq.; see Jusczyk 1997 for an overview), one might conclude that there is precious little to study in terms of perceptual phonological acquisition. But as amongst others Barton (1980), Werker (1995), Brown and Matthews (1997) and Jusczyk (1997) point out, the tasks employed in these early infant speech perception studies do not require the subjects to use the contrasts either in forming a sound-meaning pairing, or in differentiating sound patterns stored in memory. Both of these characteristics seem to be implicated in a standard definition of what it means for an individual (or a language) to have a phonological contrast. Here, I assume that either storage in memory, or pairing with meaning, brings a perceived contrast into the realm of the phonological. In section 3.2, I discuss how the proposed model can deal with differences between what can be stored in memory, and what can be paired with meaning.

Of course, tasks that tap the ability of 6– to 18–month–old infants to use language in these ways are difficult to develop and implement, especially given infants' limited ability and/or inclination to produce meaningful speech, or even to respond to it gesturally. However, Jusczyk (1997) and Werker (1995), as well as Werker and Stager (2000), review an impressive body of recent research that has overcome many of these methodological hurdles. Here I will discuss just a few of these studies. Subsection 2.1 addresses the development of segmental structure, while 2.2 focuses on prosodic development.

2.1 Segmental constraints in early perception
A number of studies have investigated the development of what Barton (1980) calls 'phonemic' perception, that is, the ability to pair a segmental phonetic contrast with a meaning difference. Experiments of this type usually require the subjects to choose between two objects forming a minimal pair (e.g. between a ball and a doll). The first of these experiments was conducted in the 1940's by Schvachkin (1948/1973), who proposed that the ability to distinguish various contrasts emerges in a relatively fixed sequence. Barton (1976, 1980) provides an excellent review of research in this tradition (see also now Brown and Matthews 1997, as well as Hallé and Boysson-Bardies 1996 for a somewhat different approach). Barton makes the important point that task demands may well cause an underestimation of children's abilities to lexically represent contrasts. He further emphasises the potential for word frequency to act as a confound in the determination of sequences of acquisition. Therefore, claims about the age at which particular contrasts emerge, as well as about precise developmental
sequences, have to be viewed with some caution. However, these studies do converge in suggesting that phonemic perception may well develop gradually, and that the ability to perceive a contrast does not entail the ability to use it to distinguish minimal pairs.

The experiments reported in Stager and Werker (1997) and Werker and Stager (2000) provide a particularly compelling demonstration that pairing sound with meaning can lead to a decrease in sensitivity to phonetic detail. They employ a version of the habituation/dishabituation procedure developed by Werker, Cohen, Lloyd, Casasola and Stager (1998). Infants are presented with a novel brightly colored moving object on a video screen, while simultaneously hearing a novel syllable being spoken by a recorded female voice over a loudspeaker. Nonce sound/meaning pairings are used to control for possible effects of familiarity. Each trial lasts 14 seconds, and includes 7 repetitions of the syllable. Trials are repeated until habituation, which is reached when the mean looking time across two sequential trials falls below 65% of the mean of the first four trials. Following habituation, the first test trial presents the moving object paired with a different syllable, which is then followed by another trial in which the pairing reverts to the same as in the habituation phase. The order of these two test trials is counterbalanced across subjects. Mean looking time is compared between the trials in which the sound/meaning pairing is switched from the habituation phase, and those in which it is the same. If infants notice the change in the aural stimulus, looking time should be longer in the switch trials than in the same trials.

Results from this procedure with 14-month-olds are presented in (1). When the syllables differ only in the place of articulation of the initial consonant (i.e. [br] vs. [dr]), looking time in the same and switch trials does not differ significantly (Experiment 1); the error bars indicate Standard Error. When the syllables differ more greatly ([llr] vs. [nim]), looking time in the switch trials is significantly higher (Experiment 2). To control for the possibility that the infants simply did not perceive the consonantal place distinction in Experiment 1, a further experiment was run in which the moving object on the video monitor was replaced by a checkerboard pattern. Since an unbounded display of this type is not likely to be perceived as a nameable object at this stage of development, this removes meaning from the task, and turns it into a pure test of perception. In this case, the infants did respond to a switch in the aural stimulus, as shown by the significantly higher looking time in switch than same trials in Experiment 3 (overall lower looking time in this experiment is likely due to the difference in visual displays).
This experiment provides an elegant demonstration that infants can perceive a contrast (Experiment 3), yet still be unable to pair it with meaning (Experiment 1). Furthermore, by employing looking time as the dependent variable, task complexity is minimised, especially in comparison with the paradigm initiated by Schvachkin (1948/1973). For other experimental methodologies that use looking time as a measure of "phonemic" discrimination, see Swingley et al. (1999), Swingley and Fernald (2000), Hoskins and Golinkoff (2000), as well as Werker et. al. (2000) for discussion of the differences between the outcomes of the looking time studies.

Cast in terms of linguistic constructs, this pattern of results leads to the conclusion that at this stage of development, the consonantal place distinction is not encoded in lexical representations, though it is present in phonetic representations. I use a 'C' to denote a representation that does not distinguish between labial and coronal place of articulation, since there is no evidence for exactly how this consonant should be represented (it could be labial, coronal, or lack place specification entirely).

(2) Phonetic/Surface Representations: [br] [dr]
Lexical/Underlying Representations: /Ci/ /Ci/
Experiment 3 shows that at some level of representation, 14-month-olds do encode the place distinction, since they are capable of perceiving it. Here this level is taken to be the Phonetic/Surface level. In Experiment 1, however, when sound is paired with meaning, 14-month-olds ignore the place difference. This leads to the conclusion that in lexical representations in which sound is linked with meaning the place distinction is absent. In Experiment 2, the syllables are distinguished along some dimension other than consonantal place which is presumably encoded in lexical representations at this stage; this experiment does not allow us to determine which of the many differences between [lɛf] and [nim] are lexically encoded.

An alternative account of these results might be that [bɪ] and [dɪ] fail to be given lexical representation because they violate English phonotactics: lexical words in English are minimally bimoraic (i.e. *[bɪ] vs. [bɪɪ] and [bɪ]). Evidence against this alternative comes from a follow-up study conducted by Pater et al. (1998) (see also Pater et al. 2001). Using the same methodology as Stager and Werker (1997), they examined whether 14-month-olds would notice a switch in voicing from [bm] to [pnm], as well as a combined place and voice switch [dm] vs. [pnm], when the presentation of the sounds was paired with the moving colourful object. They found that the infants showed no evidence of noticing the switch, which indicates that the sub-minimal status of the [br]/[dr] pair in Stager and Werker (1997) was not at issue. These results also suggest that the voice distinction, along with place, is typically absent from lexical representations at 14 months of age.

One might accept these results as indicating a developmental difference between what is perceived, and what is lexically encoded, and question whether there is in fact any later difference between what is lexically encoded and what is produced. For segmental contrast, there is a significant body of evidence, both anecdotal and experimental, that lexical encoding of contrast does precede production (see section 2.2 on prosody). Anecdotally, there are reports that when children reduce many segmental distinctions, they respond appropriately to minimal pairs containing such distinctions (see e.g. Smith 1973, Menn and Matthei 1992). These anecdotal reports are confirmed by several experimental studies, including Edwards (1974), Barton (1976), Strange and Broen (1980), and Velleman (1988). While assessing perception and production of a contrast in children of this age encounters a number of methodological hurdles, the weight of the evidence supports the genuineness of the comprehension/production gap. In fact, as the survey in Vihman (1996:56) emphasises, the main research question in this area is not whether perception precedes production, but whether some neutralizations of adult contrasts in children’s production are in fact due to misperception. This research seems to indicate that for some perceptually difficult contrasts (e.g. [θ]/[f]) this is the case, but that for most contrasts, accurate lexical
representations do exist for distinctions that are lost in production. And when the
difficult-to-perceive contrasts are acquired, they are also accurately represented
lexically before being produced (Velleman 1988: 233).

Though many details remain to be filled in through further experimentation, the picture that emerges from the extant research is that the
lexical representations that children initially construct are reduced in segmental complexity compared to what they can perceive. And at a later stage, when these lexical representations are enriched, production representations remain to be elaborated. What remains to be determined is the extent to which the restrictions on the segmental complexity of early lexical representations exactly mirror the ones that apply to early productions. In the development of prosody, however, there does exist evidence for the activity of quite similar restrictions on complexity across perception and production.

2.2 Prosodic constraints in early comprehension
In production, children learning English, as well as the prosodically similar Dutch, often go through a stage in which words are limited in size to a single initially stressed disyllable, that is, a trochaic foot (Smith 1973, Ingram 1974, Allen and Hawkins 1978, Echols and Newport 1992, Fee 1992, Fikkert 1994, Gerken 1994, Wijnen et al. 1994, Demuth 1995, Pater 1997). This stage is illustrated in the following data from Trevor (Compton and Streeter 1977, Pater 1997), in which initially stressed disyllables are produced intact, while finally stressed disyllables lose their stressless initial syllable (Trevor was learning American English, which has retained gallic final stress in *garage*):

(2)  

a. [gæ:beðʒ] garbage (1;10.5)  [wæ:dr]  rabbit (1;9.2)  
b. [gæ:dʒ] garage (1;10.5)  [wæ:f]  giraffe (1;9.1)

In terms of prosodic phonology, the operative restriction on word size is one that allows words to consist only of a single trochaic foot. This restriction is obeyed by the structures in (3a) and (3b), which correspond to Trevor's productions of the words in (2a) and (2b) respectively. Accurate production of the targets in (2b) would require either the creation of an iambic (right-headed) foot (3c), or the adjunction of the initial syllable to the Word level (3d.), either of which would exceed the 'Word=Trochaic Foot' limitation.

(3)  

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>/</td>
<td></td>
<td></td>
</tr>
<tr>
<td>'σ σ</td>
<td>'σ</td>
<td>σ σ</td>
<td>σ σ</td>
<td></td>
</tr>
</tbody>
</table>
Jusczyk (1997: 186) notes a striking similarity between such restrictions on early child productions and ones that make their mark on early perception, as evidenced in two studies. Jusczyk et al. (1999; cited as in prep. in Jusczyk 1997) used a version of Headturn Preference Procedure to examine children's early word recognition capabilities (see also Smolensky et al. this volume for explanation of this methodology). In the experiments I will discuss here, the subjects are first familiarised with a pair of disyllabic words produced in isolation, with either trochaic or iambic stress (e.g. "kingdom" and "hamlet" or "guitar" and "device"). They then listen to four passages of connected speech, two of which contain six instances of the familiarised words, with the other two passages lacking the familiarised words, but containing six instances of a non-familiarised word of the same prosodic shape. The dependent variable is the infants' orientation time towards speaker from which the passage is played; orientation time is compared between the sentences containing the familiarised words and those that don't.

Jusczyk et al. (1999) tested two groups of subjects: 7 1/2-month-olds and 10-month-olds. The younger subjects (4a.) listened longer to passages containing the familiarised words when the targets were trochaic. However infants at this age showed no preference for passages containing familiarised words when the targets were iambic. The older group (4b.) did listen longer to the familiarised words when they were iambic.

(4) Results from Jusczyk et al. 1999

<table>
<thead>
<tr>
<th></th>
<th>Familiarised</th>
<th>Unfamiliar</th>
<th>Significant Difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.5 months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trochees</td>
<td>7.92 s (SD 2.77)</td>
<td>6.80 s (SD 2.71)</td>
<td>Yes (t(23) = 2.67, p &lt; .02)</td>
</tr>
<tr>
<td>Iambs</td>
<td>7.19 s (SD 2.08)</td>
<td>7.55 s (SD 2.30)</td>
<td>No (t(23) = -1.19, p &lt; .25)</td>
</tr>
</tbody>
</table>

b. 10 months

<table>
<thead>
<tr>
<th></th>
<th>Familiarised</th>
<th>Unfamiliar</th>
<th>Significant Difference?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iambs</td>
<td>7.48 s (SD 2.18)</td>
<td>6.17 s (SD 2.17)</td>
<td>Yes (t(23) = 4.19, p &lt; .0005)</td>
</tr>
</tbody>
</table>

The results of this study suggest that trochees are receptively acquired before iambics. Further evidence for this comes from another study using the Headturn Preference Procedure, in which Jusczyk et al. (1993) show that 9-month-old English infants listen longer to lists of initially stressed disyllables than to lists containing finally stressed disyllables.

Here we have a considerable age gap between the reductions in complexity evidenced in perception (about 7.5 - 9 months of age) and those seen in production (around 18-24 months). Although Echols and Newport (1992) do propose that truncations in production are due to the initial syllable being missed in perception, considerable evidence supports the position that children do
lexically encode the initial unstressed syllable (Fikkert 1994, Gerken 1994, Paradis, Petitclerc and Genesee 1996, Jusczyk 1997: 186). We can conclude, then, that there is a genuine comprehension-production gap when children are producing truncated words.

The challenge therefore, is to capture the parallels between receptive and productive development (initially simplified representations growing in complexity), while allowing for differences (relatively rich receptive representations at the same time as reduced production). In what follows, I show that minimally violable constraints allow the construction of a model that does precisely that. Since we have relatively clear evidence of perception/production parallels in the constraints governing in prosodic complexity, the present proposal will be illustrated using that example.

3. Minimal violation across perception and production

3.1 The proposal
For simplicity's sake, I will here employ a single markedness constraint as the driving force behind truncation: WORDSIZE. I do emphasise, though, that the effects of this constraint can be reduced to the interaction of more basic prosodic requirements; see Pater (1997) for discussion. The admittedly stipulative WORDSIZE constraint is given in (5).

(5) WORDSIZE
"A word is made up of a single trochee"

In Pater (1997), truncation results from the satisfaction of this (set of) constraint(s), at the expense of faithfulness, that is, of preservation of lexical structure. The violated Faithfulness constraint is MAX (McCarthy and Prince 1999):

(6) MAX
Every segment of the input has a correspondent in the output.

The effect of ranking WORDSIZE above MAX is shown in (7):

(7) WORDSIZE >> MAX: Production

<table>
<thead>
<tr>
<th>L: gorádzb</th>
<th>WORDSIZE</th>
<th>MAX</th>
</tr>
</thead>
<tbody>
<tr>
<td>S₁ [[gádzb]₢]₃₁₄₃</td>
<td>* *</td>
<td></td>
</tr>
<tr>
<td>S₂ [[[g[rádzb]₢]₃]₃₁₄₃</td>
<td>*</td>
<td>!</td>
</tr>
</tbody>
</table>

-9-
This tableau evaluates two candidate output surface forms, $S_1$ and $S_2$, for the input lexical form $L$. Bracketing represents prosodic structure; the structure of $S_1$ corresponds to (3b), and $S_2$ to (3d). The failed candidate $S_2$, with an initial syllable outside of foot structure corresponds to the adult prosodification (Kager 1989, Pater 2000). It violates two of the stipulations of the WORDSIZE constraint, and is thus marked with a violation. Candidate $S_1$ violates none of the structural conditions encoded in WORDSIZE, though it does violate MAX twice due to the loss of two Input segments (the [g] is preserved rather than the [r] because of independent sonority-based constraints; see Pater 1997). It would be possible to satisfy WORDSIZE and MAX by placing stress on the initial syllable (e.g. gárádʒ), but this would violate other constraints not under consideration here (i.e. stress faithfulness; Pater 1997). With a ranking of WORDSIZE above MAX the truncated candidate is optimal; in the adult grammar the ranking would be reversed, and the parsing of $S_2$ would be chosen.

I assume that the role of the phonological grammar in the development of receptive competence is to regulate the structure, and therefore the complexity, of the representation(s) used in perception. This can be seen as an integral part of the learning process: by subjecting incoming data to phonological analysis, the grammar is not only placing limits on the complexity of lexical representations, but is also being itself shaped to conform with the regularities of the language. In this view, the grammar intervenes between the incoming data and the stored representation(s), as well as taking its usual place between the stored representation(s) and the concatenated utterance. A diagram of this scenario is presented in (8):

\[
\begin{align*}
\text{Perceived form} & \rightarrow \text{Grammar} & \rightarrow \text{Produced form} \\
gárádʒ & \rightarrow gárádʒ & \rightarrow gárádʒ
\end{align*}
\]

In the scenario in (8), the phonetic form of the word *garage* passes through the grammar twice, once in perception, and once in production, and each time is given a perfectly faithful representation. However, with the ranking $\text{WORDSIZE} >> \text{MAX}$, this would not occur in production, as the initial syllable would be lost. And under the most straightforward interpretation of this scenario, the perception mapping would also yield a reduced output. Such an interpretation interprets the diagram in (8) literally and takes the perceived form as the input to the grammar, and the lexical form as the output. In tableaux for this mapping, I label the input as $S$ to indicate its status as a (perceived) surface form, and the output as $L$, a stored (lexical) form. With $\text{WORDSIZE} >> \text{MAX}$, structures are also reduced on their way to lexical representation.
The tableaux in (7) and (9) are identical, except for the labeling in (9) of the input as the surface form, and of the candidate outputs as lexical forms. I make the standard assumption that structural constraints, like \textsc{wordsize}, apply to output forms. The innovation here is that in perception, the surface form is the input to the grammar, and the lexical form the output; this leads to lexical forms being subject to structural constraints in this mapping. The ranking \textsc{wordsize} $\gg$ \textsc{max} thus yields reduced structures in comprehension, corresponding to a stage in receptive development consistent with the evidence that Jusczyk (1997) discusses for English learning infants at the age of 7.5 to 9 months as outlined in section 2 above.\footnote{2}

It is worth noting that in this account, lexical representations end up being endowed with prosodic structure, since prosodification will be required by the same constraints that force segments to be parsed in surface structure, including those encapsulated in the \textsc{wordsize} constraint. In the generative tradition, predictable stress would typically be absent in the lexicon (cf. Burzio 1996). However, in Optimality Theory, predictability is not accounted for by omitting structure from the lexicon and inserting it in the proper environments by rule. Rather, phonotactic generalizations emerge from the interaction of output markedness constraints and faithfulness constraints (see Prince and Smolensky 1993 on Richness of the Base and Lexicon Optimization). For example, stress is predictable in a given language because lexically specified stress in any other position would be filtered out by dominant markedness constraints (much like iambics are filtered out by the grammar being considered here). Therefore, there is no general imperative to remove predictable information from the lexicon.

The next stage of development that we need to account for is one in which the \textsc{wordsize} limit is overcome in perception, but not in production. This situation, in which a constraint is obeyed in one domain, but not in another, is paralleled in phonological typology in a class of phenomena that Prince (1993) refers as instances of \textit{nonuniformity} of constraint application. Nonuniformity encompasses cases of the 'Emergence of the Unmarked' (McCarthy and Prince 1994), in which a markedness constraint is generally violated in a language, but its effects are seen in a particular domain, as well as what we might refer to as the 'Emergence of the Marked', where a markedness constraint is violated only in a particular environment (see e.g., McCarthy and Prince 1993 on \textit{onset} in Axininca Campa, or McCarthy 2000 on \textit{nocoda} in Rotuman).
Nonuniformity is handled straightforwardly in Optimality Theory: a constraint applies nonuniformly because it conflicts with a higher ranked constraint in some environments. Usually, the constraint that stops a markedness constraint from applying uniformly is a faithfulness constraint that applies only in a subset of the environments targeted by the markedness constraint. For example, in McCarthy and Prince's (1994, 1999) analysis of reduplication, markedness constraints apply to both reduplicants and their bases, as well all other forms in the language. Faithfulness constraints, however, are specified as to whether they apply to the Input-Output mapping, thus targeting bases and other non-reduplicated strings, or to the Base-Reduplicant mapping, in which case they control the extent to which reduplicants resemble their bases. Structural limitations that apply only to reduplicants are derived by having a markedness constraint dominate Base-Reduplicant Faithfulness, but not Input-Output Faithfulness:

(10) Ranking schema for the emergence of the unmarked in reduplication
FAITH(I-O) >> MARKEDNESS >> FAITH(B-R)

For example, in Diyari (McCarthy and Prince 1994), the constraints encapsulated in our WORDSIZE dominate Base-Reduplicant faithfulness, forcing a one-foot maximum on the reduplicated string, but are in turn dominated by Input-Output Faithfulness, allowing bases and other lexical words to exceed the one foot limit. Generalizing beyond reduplication, nonuniformity is often produced by the following ranking (where FAITH(D1) and FAITH(D2) are faithfulness constraints indexed to different domains, and MARKEDNESS is a well-formedness constraint):

(11) Typical ranking schema for Nonuniformity
FAITH(D1) >> MARKEDNESS >> FAITH(D2)


To handle discrepancies between the amount of complexity permitted of structures accessed in perception and production, I propose that Faithfulness constraints can be specified to apply either to the Lexical-to-Surface mapping of production – FAITH(LS), or to the Surface-to-Lexical mapping of perception – FAITH(SL). When the Input to the grammar is a Lexical form, FAITH(LS) constraints are relevant, and when the Input to the grammar is a Surface form, FAITH(SL) is at issue. Thus, we now have two versions of Max, given in (12):
(12) \textit{MAX}(LS)
If the input is a lexical form, every segment of the input has a correspondent in the output

\textit{MAX}(SL)
If the input is a surface form, every segment of the input has a correspondent in the output

When initial unstressed syllables begin to be represented in comprehension, this indicates the promotion of \textit{MAX}(SL) above \textit{WORDSIZE}. With \textit{MAX}(LS) resting beneath \textit{WORDSIZE}, we get the following ranking, which instantiates the nonuniformity schema:

(13) \textit{MAX}(SL) >> \textit{WORDSIZE} >> \textit{MAX}(LS)

The following tableau shows the effect of this ranking in perception

(14) \textit{MAX}(SL) >> \textit{WORDSIZE} >> \textit{MAX}(LS): \textit{perception}

<table>
<thead>
<tr>
<th>S: [ga\textsuperscript{r}\text{\textael}d\textsuperscript{z}]_{\textit{PwWd}}</th>
<th>\textit{MAX}(SL)</th>
<th>\textit{WORDSIZE}</th>
<th>\textit{MAX}(LS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{L}\textsubscript{1} [([g\text{\textael}d\textsuperscript{z}]<em>{\textit{PwWd}})</em>{\textit{PrWd}}]</td>
<td>**</td>
<td>*</td>
<td>!</td>
</tr>
<tr>
<td>\textit{L}\textsubscript{2} [g\text{\textael} [r\text{\textael}d\textsuperscript{z}]_{\textit{PwWd}}]</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The initial unstressed syllable is represented in the lexical form, in violation of \textit{WORDSIZE}, due to the dominance of \textit{MAX}(SL). In the production mapping, however, \textit{MAX}(SL) does not apply, and we get truncation as the optimal result:

(15) \textit{MAX}(SL) >> \textit{WORDSIZE} >> \textit{MAX}(LS): \textit{production}

<table>
<thead>
<tr>
<th>L: [g\text{\textael} [r\text{\textael}d\textsuperscript{z}]<em>{\textit{PwWd}}]</em>{\textit{PrWd}}</th>
<th>\textit{MAX}(SL)</th>
<th>\textit{WORDSIZE}</th>
<th>\textit{MAX}(LS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{S}\textsubscript{1} [([g\text{\textael}d\textsuperscript{z}]<em>{\textit{PwWd}})</em>{\textit{PrWd}}]</td>
<td>*</td>
<td>**</td>
<td>!</td>
</tr>
<tr>
<td>\textit{S}\textsubscript{2} [g\text{\textael} [r\text{\textael}d\textsuperscript{z}]_{\textit{PwWd}}]</td>
<td>*</td>
<td></td>
<td>!</td>
</tr>
</tbody>
</table>

The pair of tableaux in (14) and (15) represent the stage of development in which children lexically represent initial unstressed syllables, but fail to produce them. The \textit{WORDSIZE} constraint is violated in comprehension to satisfy higher \textit{MAX}(SL), but continues to force truncation in production where \textit{MAX}(SL) fails to apply. The comprehension–production gap is thus dealt with as an instance of minimal constraint violation.

The adult state would be characterised by the reranking of \textit{MAX}(LS) above \textit{WORDSIZE}, as illustrated by (16):
This of course abstracts from the rich interactions in the adult grammar between faithfulness and the structural constraints encapsulated in WORDSIZE (Pater 2000).

Summing up, we have see that the effects of the WORDSIZE constraint change over the course of development due to its ranking with respect to the posited faithfulness constraints:

(17)
Stage 1 - Fully satisfied WORDSIZE: \textbf{WORDSIZE} >> \textbf{MAX(SL), MAX(LS)}
Stage 2 - Minimally violated WORDSIZE: \textbf{MAX(SL)} >> \textbf{WORDSIZE} >> \textbf{MAX(LS)}
Stage 3 - 'Inactive' WORDSIZE: \textbf{MAX(SL), MAX(LS)} >> \textbf{WORDSIZE}

In Stage 1, the structural constraint is satisfied in both perception and production, due to its dominance over both faithfulness constraints. In Stage 2, it is violated only when it conflicts with the requirement of perceived surface structures to be represented lexically, but not in production. And finally, in the last stage, it is violated both in perception and production (though its component constraints do continue to play an active role in the grammar). This account captures the parallels between the restrictions imposed on early perception and on early production by attributing them to a single set of markedness constraints. Differences between the state of development of receptive and productive competence are dealt with in terms of differential rankings of the respective faithfulness constraints with respect to the markedness constraints.

From a learnability perspective, one might ask why the grammar does not immediately converge on the final state (cf. Hale and Reiss 1998). After all, the learner at Stage 1 does have evidence that initial unstressed syllables are permitted in the language, and given an algorithm like Error-Driven Constraint Demotion (Tesar and Smolensky 1998), the WORDSIZE constraint should be demoted immediately so that the grammar produces an output matching the adult form. This suggests, I believe, that real-life acquirers of language either require, or wait for, a certain weight of evidence before adjusting rankings (see Boersma's 1998 Gradual Constraint Demotion algorithm for another approach).
3.2 Further articulating the model

In this section, I will suggest how this model can be further articulated to handle data that are at this point problematic. Using the same headturn preference procedure as in the studies by Jusczyk and colleagues described above, Jusczyk and Aslin (1995) showed that 71/2-month-olds familiarised with a word like 'cup' and 'dog' would listen significantly longer to passages containing those words than to passages containing the minimally different 'tup' and 'bawg'. This result stands in contrast to the evidence from Stager and Werker (1997) that infants at 14 months of age do not lexically represent consonantal place of articulation differences, as well the results of Hallé and Boysson-Bardies (1996).

Jusczyk and Aslin (1995) are careful to point out that though their results do show that the infants are capable of representing relatively fine phonological detail in memory, they not show that the infants are capable of linking these minimal phonological distinctions to meaning differences. The apparent contradiction between the results of Jusczyk and Aslin (1995) and Stager and Werker (1997) can be resolved if we acknowledge the existence of two levels of representation. The level of lexical representation links phonological form to meaning, and structures at this level are accessed in the methodology employed by Werker and colleagues. The surface phonological representation, however, is meaning-free, and this is the level accessed by the methodology used by Jusczyk and colleagues.

Under this view, we would need separate faithfulness constraints for the mapping from perceived acoustic representation to the surface representation (\(\text{FAITH(AS)}\)), as well as the constraints mapping from surface to lexical representation (\(\text{FAITH(SL)}\)). A ranking of \(\text{FAITH(AS)} \gg \text{MARKEDNESS} \gg \text{FAITH(SL)}\) would produce richer surface representations than lexical ones, consistent with the divergence between the Jusczyk and Aslin (1995) and Stager and Werker (1997) results. The posited levels of representations, along with the relevant faithfulness constraints, are illustrated in (18).

(18) \(\begin{align*} \text{Acoustic Representation} \downarrow \text{FAITH(AS)} \downarrow \\
\text{Surface Representation} \downarrow \text{FAITH(SL)} \downarrow \\
\text{Lexical Representation} \downarrow \text{FAITH(LS)} \downarrow \\
\text{Surface Representation} \end{align*}\)

Each of these levels of representation is elaborated sequentially through development, and the extant literature on segmental perception and production provides a relatively clear outline of this process (see Werker and Pegg 1992,
Jusczyk 1997 for relevant discussion). The acoustic representation is assumed to be relatively richly specified even in the infant's earliest exposure to language; this representation is the one tapped by experiments showing infants' ability to discriminate almost any contrast from amongst the world's languages (see Jusczyk 1997 on the controversy about whether this level of representation is language-specific or not). The surface representation is elaborated once the relevant FAITH(AS) constraints are promoted above the relevant MARKEDNESS constraints after some amount of exposure to the language being learned (19b); this corresponds to the language-specific perception that emerges between 6-9 months. Learning of sound-meaning pairings between around 11-18 months would force the elaboration of lexical representations and the consequent ranking of FAITH(SL) above markedness constraints (19c). Appearance of contrasts in production of meaningful words in the period between about 18-to-24 months would entail the ranking of FAITH(LS) above markedness (19d).³

(19)  a. MARKEDNESS >> FAITH(AS), FAITH(SL), FAITH(LS)

    b. FAITH(AS) >> MARKEDNESS >> FAITH(SL), FAITH(LS)

    c. FAITH(AS), FAITH(SL) >> MARKEDNESS >> FAITH(LS)

    d. FAITH(AS), FAITH(SL), FAITH(LS) >> MARKEDNESS

Here, one might ask whether there is any reason acquisition should follow this particular path. Why couldn't FAITH(LS) be ranked above MARKEDNESS before FAITH(AS) and FAITH(SL), yielding productive competence ahead of receptive? The answer, I believe, lies in the nature of the triggering evidence: in order for the structures that would trigger reranking of a FAITH(LS) constraint over MARKEDNESS to exist in the lexical forms, the analogous FAITH(SL) constraint must already dominate MARKEDNESS.

4. Comparison with other approaches

4.1 Smolensky 1996
The present proposal borrows much from the one in Smolensky (1996) (see also Ohala 1996 for a similar, independently developed approach). There too, a single Optimality Theoretic grammar is invoked in both production and comprehension, and is involved in the latter in evaluating candidate lexical forms for a fixed surface form. However, in Smolensky's proposal, the choice between candidate lexical forms is made solely on the basis of compliance to faithfulness, rather than
because of differences in markedness violations. This is because the choice of lexical form is irrelevant to the assessment of markedness violations.

Taking the case of truncation discussed in the previous section, I provide in (20) examples of competing candidates in the production and comprehension mappings, and an indication of which is optimal under the WORDSIZE >> FAITH ranking.

(20) WORDSIZE >> FAITH
Candidates competing in production:
- a. L: /gærád3/ S: [gærád3] Structure: [g[ê[rád3]Ω]]PmPrWd
- b. L: /gærád3/ S: [gád3] Structure: [g<ê[<rád3]Ω]]PmPrWd

Candidates competing in comprehension:
- c. L: /gád3/ S: [gærád3] Structure: [g[ê[rád3]Ω]]PmPrWd
- d. L: /gærád3/ S: [gærád3] Structure: [g[ê[rád3]Ω]]PmPrWd

In production, the WORDSIZE violation in (20a) is fatal, because of the existence of candidate (20b), which satisfies WORDSIZE at the expense of faithfulness violations, here indicated in the structure in terms of unparsed segments enclosed in ‘< >’ (Smolensky 1996 assumes the Parse/Fill version of Faithfulness of Prince and Smolensky 1993). In comprehension, candidates sharing the surface form [gád3] compete, and a mapping like (20b) is therefore not in the candidate set. Since any candidate with the surface form [gærád3] will necessarily violate WORDSIZE, the decision falls down to the faithfulness constraints. Here a constraint against insertion between lexical and surface form would choose (20d) over (20c); inserted segments are italicised.4 Markedness constraints might choose between possible structural interpretations of a surface string (using this case as an example, they would choose between leaving the initial syllable outside of foot structure and creating an iambic foot; i.e. between 20c. and 20d.). However, the mapping between the underlying and surface form will always be purely faithful, as dictated by the faithfulness constraints.

This model of establishing lexical representations does successfully capture a situation in which comprehension is advanced relative to production, as it is intended to. However, because only faithfulness determines the choice of lexical forms in the comprehension mapping, it is incapable of dealing with a situation in which phonological comprehension is anything less than perfect. As a result, the development of comprehension must be taken to be purely extra-grammatical. Not only is this implausible, but it would entail much duplication of analytic effort and machinery, insofar as early comprehension and early production, and grammars in general, are subject to similar restrictions, as the evidence reviewed in section 2 of this paper suggests they are.5
4.2 A mixed model

In the basic model of the development of receptive competence proposed in section 3.1 above, markedness constraints influence comprehension because the lexical form is taken to be the output that they assess; it denies the standard equivalence between 'lexical form' and 'input', and between 'surface form' and 'output'. It is possible, however, to recast the proposal in such a way as to restore that equivalence, and in so doing, create a model that incorporates some aspects of the proposal in the present paper and some of the one in Smolensky (1996).

The key is in recognizing, as in section 3.2, that the perceived string is not simply equivalent to a phonological surface string; rather, the phonological surface string must be derived from the acoustic-phonetic representation. This could be done by taking the acoustic-phonetic string as input to the grammar, as suggested in 3.2. However, let us instead follow the standard assumption that grammatical inputs are strictly equivalent to lexical forms. The following then, would be potential grammatical mappings for the perceived acoustic-phonetic string [garádʒ]. Note that we are dropping Smolensky's (1996) premise that the surface strings must faithfully represent the phonetic form:

(21) Acoustic-phonetic   Surface   Lexical
    a. garádʒ      [gádʒ]PWF   /gádʒ/
    b. garádʒ      [gərādʒ]PWF   /garádʒ/

Markedness constraints would prefer (21a); this would be the correct outcome for the earliest stages of acquisition, in which comprehension representations are simplified. To make (21b) optimal, it must be the case that there are constraints that conflict with markedness besides ordinary lexical-surface faithfulness, which are fully satisfied in both (21a) and (21b). Here we need to allow at least a limited grammatical status to acoustic-phonetic representations, as the targets of FAITH(AS) constraints. Such constraints are in a sense analogous to Output-Output constraints (see e.g. McCarthy and Prince 1995, 1999, Benua 1997) in demanding identity between the surface form and some non-input form. The rankings that generate the three stages of development would be identical to those proposed in 3.1, with the exception of the substitution of FAITH(AS) for FAITH(SL); faithfulness between lexical and surface forms is now governed by a single set of constraints. Generalizing over markedness and faithfulness constraints, the schema for the stages of acquisition would be as follows:

(22) Stage 1: MARKEDNESS >> FAITH(AS), FAITH(LS)
    Stage 2: FAITH(AS) >> MARKEDNESS >> FAITH(LS)
    Stage 3: FAITH(AS), FAITH(LS) >> MARKEDNESS
Stage 1 is the schema for reduced representations in comprehension and production, Stage 2 for faithful comprehension and reduced production, and Stage 3 for faithful comprehension and production. In (23) I provide a tableau for perception at Stage 1, using the structural constraint WORDSIZE, along with general FAITH constraints that are violated by any change to the segmental makeup of a string:

(23) WORDSIZE >> FAITH(AS), FAITH(LS): perception

\[
\text{Acoustic-phonetic form: } \text{gar\text{á}\text{d}}\text{\text{á}}
\]

<table>
<thead>
<tr>
<th></th>
<th>Acoustic-phonetic form:</th>
<th>perception</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gar\text{á}\text{d}\text{á}</td>
<td></td>
</tr>
<tr>
<td>WORDSIZE</td>
<td>FAITH(AS)</td>
<td>FAITH(LS)</td>
</tr>
<tr>
<td>a. L: /gád\text{á}/ ~ S: [gád\text{á}]_{\text{prWd}}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. L: /gar\text{á}d\text{á}/ ~ S: [gá\text{á}d\text{á}]_{\text{prWd}}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. L: /gá\text{d}\text{á}/ ~ S: [ga\text{\text{á}}\text{d\text{á}}]_{\text{prWd}}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

At the earliest stage of acquisition, words are limited to a single trochaic foot, in comprehension as well as production. Here this limitation is enforced by having the structural constraints apply to the surface phonological form constructed on the basis of the perceived acoustic-phonetic representation. Following McCarthy and Prince (1999), I take the candidates supplied by Gen to be input-output pairs (=lexical~surface). Given by the perceptual system is the form gar\text{á}\text{d}3. Because of the ranking of FAITH(AS) beneath WORDSIZE, the reduced structure (23a) is preferred over the faithful representations of the acoustic-phonetic string in (23b) and (23c). Note that candidate (23c) is harmonically bounded by (23b); no ranking of these constraints will ever make it optimal. This points to the fact that this model cannot produce lexical representations that are simplified relative to surface ones; the consequences of this will be taken up below. For now, the important point is that this model does allow for initially impoverished comprehension, and subsequent development, unlike the proposal in Smolensky (1996).

The subsequent development is captured by the reranking of FAITH(AS) above WORDSIZE; with FAITH(LS) remaining beneath the markedness constraint, we produce the comprehension/production gap:

(24) FAITH(AS) >> WORDSIZE >> FAITH(LS): perception

\[
\text{Acoustic-phonetic form: } \text{gar\text{á}\text{d}}\text{\text{á}}
\]

<table>
<thead>
<tr>
<th></th>
<th>Acoustic-phonetic form:</th>
<th>perception</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gar\text{á}\text{d}\text{á}</td>
<td></td>
</tr>
<tr>
<td>FAITH(AS)</td>
<td>WORDSIZE</td>
<td>FAITH(LS)</td>
</tr>
<tr>
<td>a. L: /gád\text{á}/ ~ S: [gád\text{á}]_{\text{prWd}}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. L: /gar\text{á}d\text{á}/ ~ S: [gá\text{á}d\text{á}]_{\text{prWd}}</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In comprehension, the marked prosodic structure is now created, as in (23b), in order to satisfy the higher ranked FAITH(AS). In production, though, the continued
dominance of \textsc{wordsize} over \textsc{faith}(\textsc{ls}) leads to the emergence of the unmarked structure:

(25) \textsc{faith}(\textsc{as}) >> \textsc{wordsize} >> \textsc{faith}(\textsc{ls}): \textit{production}

\begin{center}
\begin{tabular}{|c|c|c|}
\hline
L: \texttt{/g\textipa{\textacuten\textcedilla\textmd{\textipa{\textacuten\textcedilla}}/} } & \textsc{faith}(\textsc{as}) & \textsc{wordsize} & \textsc{faith}(\textsc{ls}) \\
\hline
a. L: \texttt{/g\textipa{\textacuten\textcedilla\textmd{\textipa{\textacuten\textcedilla}}/} } & & & \\
\hline
\hline
b. L: \texttt{/g\textipa{\textacuten\textcedilla\textmd{\textipa{\textacuten\textcedilla}}/} } & & & \\
\hline
\end{tabular}
\end{center}

In production, the lexical form is given, and the candidate set is thus restricted to mappings that contain it. Candidate (25b) is chosen due to its satisfaction of \textsc{wordsize}; here it satisfies \textsc{faith}(\textsc{as}) vacuously, since working memory contains no trace of the adult pronunciation. Interestingly, this suggests an approach to the greater pronunciation accuracy often observed in imitation: if the adult acoustic-phonetic representation were still in memory, \textsc{faith}(\textsc{as}) would target the surface forms in production, and would favour faithful pronunciations.

This alternative model retains the key element of the proposal presented earlier in this paper: structural constraints influence the outcome of comprehension. In deriving the structure of lexical forms from surface forms without the influence of markedness, and in treating the 'input-output' mapping as equivalent to a 'lexical-surface' mapping it is consistent with Smolensky (1996), and more generally, with standard assumptions in Optimality Theory. However, a possible shortcoming of this compromise proposal comes from the immunity of lexical forms to markedness constraints. Because of this, it is impossible to articulate this model so as to allow for lexical representations that are impoverished relative to surface ones (cf. section 3.2). The choice between this approach, and the one suggested in sections 3.1 and 3.2, seems to hinge on whether restrictions on the complexity of lexical representations are truly phonologically determined, which can only be ascertained through further empirical research (see Pater \textit{et al.} 2001 for some discussion).

4.3 The dual-lexicon model

The present proposal might appear to be simply an Optimality Theoretic recasting of the dual-lexicon model (see e.g. Menn and Matthei 1992). Where that model posits separate representations for production and perception, and different constraints on what is permitted at each level, here we have separate faithfulness constraints for each domain.

There are some important differences, however. First of all, the present model is constrained by having a single set of markedness constraints operating across both domains. There are two advantages to this. The obvious one is that in allowing levels to differ only in ranking of faithfulness constraints, limits are
placed on the extent to which they can diverge (see Benua 1997 and Itô and Mester 1999 for related discussion). However, a more important difference in the present context is that knowledge gained from receptive ranking of markedness constraints is directly transferred to production. For example, when a child learning Dutch or English receptively learns that feet are left-headed trochees rather than right-headed iambs, a constraint favouring the former type of foot would be ranked over one favouring the latter. This ranking of markedness constraints would not have to be relearned in the production grammar (see Fikkert 1994 for evidence of productive imposition of trochaic form).

In addition, this approach avoids a fatal flaw in the dual lexicon model, pointed out by Menn and Matthei (1992:223). They note that there is evidence of processes that apply both within and across words (Donahue 1986, Stemberger 1988, Matthei 1989). In the dual-lexicon model, the stored output representation is the target of production constraints, which leads to an inability to deal with between word processes.

\[(26) \quad \ldots \text{we must assume that the same constraints are operating both within and across words. We might, as Matthei (1989) proposed, move the output constraints from the output lexicon to somewhere else – but where? And if we do that, how do we specify the way that those constraints will interact with the lexicon to create the child's output representations?}\]

In terms of Optimality Theory, the answers to these questions are now obvious: output constraints are in the grammar, and they interact with the lexicon through the mediation of faithfulness constraints.

5. Conclusions

This proposal follows Smolensky (1996) in using Optimality Theory to model developing receptive and productive phonology with a single grammar. Where it differs is in that the grammar constrains early comprehension as well as early production, instead of simply mapping perceived surface forms to identical lexical forms and thus making receptive phonology flawless from the outset of acquisition.

While it might seem obvious that comprehension does in fact develop, current evidence from developmental speech perception further suggests that receptive competence unfolds in a manner parallel to the later development of productive competence, with similar initial unmarked segmental and prosodic structures in the two domains.
These parallels can be captured by having a single set of structural constraints that applies to both the produced surface form as well as the receptively created and accessed lexical form. With these constraints ranked over faithfulness constraints, the initial state of the grammar admits only structurally unmarked representations. Receptive competence develops by the promotion of comprehension specific faithfulness constraints above the structural constraints, as well as the establishment of rankings between the structural constraints. Productive competence is allowed to lag behind, however, by the lower rank of the production specific faithfulness constraints, whose later promotion is indicated by the evidence that relatively complete lexical representations underly simplified productions.

From the perspective of phonological acquisition, one might question whether receptive development is in fact subject to all and only the constraints that are later seen in production. Indeed, a strong version of this proposal would hold exactly that, and it is worth maintaining that position in the absence of evidence to the contrary. It is likely that some structural constraints are comprehension or production specific, but at this point, the extent to which receptive and productive development converge on the same constraints remains an empirical question. This question seems independently interesting, since in attempting to answer it, we will likely forge a deeper understanding of what the shared grammatical core is between domains, as well as where they diverge.

Beyond the domain of first language acquisition there is also growing body of evidence that shows that phonotactic restrictions play a role in adult speech perception (Dupoux et al. 1998, Moreton 1999, 2000). This can be taken as additional support for the central thesis of this paper: that in perception, markedness constraints do intervene between the raw acoustic signal and the lexical representation.

References


Hayes, Bruce (This volume). Phonological acquisition in Optimality Theory: the Early Stages.


Pater, Joe, Christine Stager, and Janet Werker (2001). The Lexical Acquisition of Phonological Contrasts. Ms. University of Massachusetts, Amherst and University of British Columbia.

Prince, Alan (1993). Minimal Violation, handout of talk presented at ROW-1, Rutgers University, New Brunswick, N.J.


Smolensky, P., L. Davidson, and P. Jusczyk (This volume). The Initial and Final States: Theoretical Implications and Experimental Explorations of Richness of the Base.


Werker, Janet, Kathleen Corcoran, Christopher Fennell, and Christine Stager (2000). Infants’ ability to learn phonetically similar words: Effects of age and vocabulary size. Ms. University of British Columbia.


Notes

* Portions of this paper were presented at From Speech Perception to Word Learning: A Workshop, at the University of British Columbia, as well as in colloquia at Utrecht University, the Max Planck Institute for Psycholinguistics, Nijmegen, University of Groningen, SUNY Stonybrook, and the University of Massachusetts, Amherst. Special thanks to Janet Werker for inviting me to speak on these issues at the workshop, and for much helpful subsequent discussion, as well as to Paul Smolensky for extensive comments on an earlier version of this paper. Conversations with Ellen Broselow, Angela Carpenter, Suzanne Curtin, Della Chambless, Anne Cutler, Lyn Frazier, Bruce Hayes, Peter Jusczyk, René Kager, John Kingston, Clara Levelt, John McCarthy, Steve Parker, Doug Pulleyblank, Chris Stager and Wim Zonneveld also helped clarify my thoughts on the topic. This research was supported by SSHRC postdoctoral fellowship 756-96-05211 and SSHRC research grant 410-98-1595, for which I am grateful.

1 One might instead posit a more indirect link between performance and grammar, as suggested by the linking hypotheses of Smolensky et. al. (this volume). There do not seem to be any obvious barriers to implementing the present proposal in such terms; one would simply need a means for the rank of faithfulness constraints to vary depending on the use to which the grammar is being put.

2 Interestingly, for one account of Jusczyk et al.’s (1999) findings to go through, we must assume that the onset is at least sometimes borrowed from the initial syllable. Recall that when the 7.5-month-olds are trained on guitar, they do not listen longer to guitar passages. We might assume that in familiarization, a lexical form is created which is then matched against the passage. If the infants stored [tar], then this would match perfectly with the stressed portion of guitar. And indeed, when 7.5-month-olds are trained on tar, they do listen longer to guitar passages. If, however, [gar] is the lexicalized form, then it does not match the stressed syllable of guitar (due perhaps to FAITH-based preference for dorsal over coronal). I should emphasize, however, that my strategy in this paper is not to provide such a direct account of the experimental evidence. Rather I make the minimal inference from these experiments that trochaic forms are acquired before iambic ones, and then I provide a grammatical account of the stage of development at which only trochees are present.

3 Insofar as babbling is taken to interact with the grammar, one might hypothesize that constraints mapping perceived surface representations to produced surface representations would be at issue.
Under a Parse/Fill theory, this candidate could not even be generated, since inserted segments are empty syllabic nodes, rather than featurally specified phones (unless phonetic interpretation happened to supply a [ə] and [r] as the default vowel and consonant respectively). This would be another difficulty in producing reductions in structure in lexical structures under the account presented in Smolensky (1996).

Similar difficulties also face the alternative to Smolensky’s model presented in Hale and Reiss (1998), where an initial ranking of Faithfulness over Markedness is posited to yield initially rich lexical representations. In this proposal though, not only is impossible to portray the development of receptive competence, but the development of production is also explicitly claimed to be extra-grammatical. One is therefore led to wonder what empirical domain in acquisition would be considered under the provenance of grammar; while Hale and Reiss suggest that perceptual experiments might be used to tap phonological competence, it is not clear whether any of those reviewed in the present paper would be considered suitable.