

# Phonological Acquisition in Optimality Theory: The Early Stages<sup>1</sup>

Submitted for a forthcoming volume on phonological acquisition and typology,  
edited by René Kager and Wim Zonneveld

## Abstract

Recent experimental work indicates that by the age of ten months, infants have already learned a great deal about the phonotactics (legal sounds and sound sequences) of their language. This learning occurs before infants can utter words or apprehend most phonological alternations. I will show that this early learning stage can be straightforwardly modeled with Optimality Theory. Specifically, the Markedness and Faithfulness constraints can be ranked so as to characterize the phonotactics, even when no information about morphology or phonological alternations is yet available. I will also show how later on, the information acquired in infancy can help the child in coming to grips with the alternation pattern. I also propose a procedure for undoing the learning errors that are likely to occur at the earliest stages.

There are two specific formal proposals. One is a constraint ranking algorithm, based closely on Tesar and Smolensky's Constraint Demotion, which mimics the early, "phonotactics only" form of learning seen in infants. I illustrate the algorithm's effectiveness by having it learn the phonotactic pattern of a simplified language modeled on Korean. The other proposal is that there are three distinct default rankings for phonological constraints: low for ordinary Faithfulness (used in learning phonotactics); low for Faithfulness to adult forms (in the child's own production system); and high for output-to-output correspondence constraints.

---

<sup>1</sup> I would like to thank the participants in a Spring 1998 seminar at UCLA, Adam Albright, Sun-Ah Jun, Patricia Keating, Charles Reiss, audience members in Utrecht and San Diego where this paper was given as a talk, and  
your name here: \_\_\_\_\_ for helpful comments.

# Phonological Acquisition in Optimality Theory: The Early Stages

## 1. Introduction

The study of phonological acquisition at the very earliest stages is making notable progress. Virtuosic experimental work accessing the linguistic knowledge of infants has yielded extraordinary findings demonstrating the precocity of some aspects of acquisition. Moreover, phonologists now possess an important resource, Optimality Theory (Prince and Smolensky 1993), which permits theorizing to relate more closely to the findings of experimental work. The purpose of this paper is to outline one way in which the experimental and theoretical research lines can be brought more closely together. The central idea is that current phonological theory can, without essential distortion, be assigned an architecture that conforms closely to the process of acquisition as it is observed in children. I conclude with a speculative, though reasonably comprehensive, picture of how phonological acquisition might proceed.

## 2. Empirical Focus

To avoid confusion, I will try to make clear that my view of what “phonological acquisition” involves may be broader than the reader is predisposed to expect.

When we study how very young children learn language, we can follow two divergent paths. One is to examine what children say, the other is to develop methods that can determine what children understand or perceive. The reason these two methods are so different is that (by universal consensus of researchers) acquisition is always more advanced in the domain of perception than in production: children often cannot utter things that they are able to perceive and understand.

A fairly standard view of children’s productions (e.g. Smith 1973) is that the internalized representations that guide children are fairly accurate,<sup>2</sup> and that the child carries out her own personal phonological mapping (Kiparsky and Menn 1975) which reduces the complex forms she has internalized to something that can be more easily executed within her limited articulatory capacities. The study of this mapping is a major research area; for some recent contributions see Levelt (1994), Fikkert (1994), Gnanadesikan (1995), Pater (1996), Boersma (1998), and various papers in this volume.<sup>3</sup>

---

<sup>2</sup> Things are of course more complicated than this cursory statement can indicate; see for instance the classic study of Macken (1980), the literature review in Vihman (1996, Ch. 7), and Pater (this volume). What is crucial here is only that perception have a wide lead over production.

<sup>3</sup> Hale and Reiss (1998) likewise take the child’s output mapping to be separate from her phonological system per se. However, they go further in claiming that the child’s mapping is utterly haphazard, indeed the result of the child’s “body” rather than her “mind”. I cannot agree with this view, which strikes me as an extraordinary denigration of research in child phonology. To respond to two specific contentions:

But it is also important to consider the other side of the matter: we need a clear characterization and analysis of the child's **internalized conception of the adult language**. As just noted, this will often be richer and more intricate than can be detected from the child's own speech. Indeed, the limiting case is the existence (see below) of language-particular phonological knowledge in children who cannot say anything at all. This paper focuses especially on the latter area, which can perhaps be fairly described as neglected by phonologists.

To clarify what I mean by "internalized conception of the adult language," consider the classic example of *blick* [blɪk] vs. \**bnick* [bnɪk] (Chomsky and Halle 1965). Speakers of English immediately recognize that *blick* is non-existent but possible, whereas *bnick* is both non-existent and ill-formed; it *could not* be a word of English. This is a purely passive form of linguistic knowledge, and could in principle be learned by an infant before she ever was able to talk. As we will see shortly, there is experimental evidence that this is more or less exactly what happens.

### 3. Some Results from the Acquisition Literature

To start, I will conduct a rapid and cursory summary of various results from the experimental literature in phonological acquisition. All of these results are likely to be modified by current or future research, but I think a useful general trend can be identified.

Before presenting these results, it is worth first mentioning that they were made possible by the development of an extraordinary level of expertise in designing experiments that can obtain evidence about what infants know. Here is a very brief review. At birth, infants can provide information about what interests them in their surroundings when they vary the rate of sucking on an electronically-monitored rubber nipple. Older babies can turn their heads in the direction they choose, and thus can indicate what interests them in experiments that involve conditioning and reinforcement (though the reinforcement is generally not food, as it is for experimental animals, but stimulation, in the form of a pop-up toy). Crucially, such studies have developed methods that ensure that the observations (e.g. "Did the baby turn her head rightward?") are legitimate and do not reflect wishful thinking on the part of the observer. In addition, experimentalists rely on the testimony of many babies and do careful statistical significance testing before any claims are made on the basis of the results.

---

1) The free variation and near-neutralizations seen in the child's output (Hale and Reiss, 669) are common in adult phonology, too. Whatever is developed as a suitable account of these phenomena (and progress is being made) is likely to yield insight into children's phonology as well.

2) Claimed differences between children's constraints and adults' (see Hale and Reiss, (18a)) can be understood once we see constraints (or at least, many of them) as grammaticized principles that address phonetic problems. Since children employ different articulatory strategies (such as favoring jaw movement over articulator movement), they develop different (but overall, rather similar) constraint inventories.

### 3.1 Abilities Present at Birth: Inherent Auditory Boundaries

Eimas et al. (1971) raised the intriguing possibility that there might exist innate “feature detectors.” Neonates apparently best perceive distinctions along the acoustic Voice Onset Time continuum that match those characteristically used in human languages. This remarkable result was later rendered perhaps somewhat less exciting when similar perceptual abilities were located in nonlinguistic species, in particular chinchillas (Kuhl and Miller 1975, 1978) and macaques (Kuhl and Padden 1982, 1983). These later results forced a more modest interpretation of the Eimas findings, of a rather functionalist character (Kuhl and Miller 1975, Keating 1984): human languages tend to place their phoneme boundaries at locations where they are readily distinguished by the mammalian auditory apparatus.

### 3.2 Language-Specific Knowledge at Six Months: Perceptual Magnets

Six-month-old infants apparently know few if any words. Thus, whatever language learning they are doing must take place in the absence of a lexicon—plainly, a major handicap! Nevertheless, the work of Kuhl (1991, 1995) shows that six-month-olds have already made a certain sort of progress toward attaining the ambient phonological system, which plausibly serves them well during the following months, as they acquire the ability to recognize words.

Kuhl’s work demonstrates what she calls a “perceptual magnet” effect: when six-month-olds listen to various acoustic continua (such as synthesized vowels varying in F2), they discriminate tokens relatively *poorly* when token pairs lie close to the phonetic norms for the ambient language’s categories; and relatively *well* when the token pairs lie midway between phonemic norms. This result is somewhat like the familiar pattern of categorical perception (e.g. Fodor, Bever, and Garrett 1974), but in a more sophisticated, gradientized form. Kuhl’s term “perceptual magnet” refers to the phonetic category center, which acts like a magnet in causing closely neighboring tokens to sound more like it than they really are.

Kuhl’s findings were later submitted to theoretical modeling in the work of Guenther and Gjaja (1996). Guenther and Gjaja deployed a neural net model that directly “learned” the set of perceptual magnets found in the input data, relying *solely on facts about token distributions*. That is, if the input set of formant frequencies has a cluster that centers loosely on the phonemic target for (say) [i], the Guenther/Gjaja model would learn a perceptual magnet in this location. The model mimics the behavior of humans with respect to perceptual magnets in various ways, as the authors showed.

As Kuhl (1995) has pointed out, a very appealing aspect of the “perceptual magnet” concept is that it represents a form of information that can be learned before any words are known. In any phonemic system, the phonetic tokens of actual speech are distributed unevenly, being clustered around the phonemic centers. By paying attention to these asymmetries, and by processing them (perhaps in the way Guenther and Gjaja suggest), the child can acquire what I will here call **distributional protocategories**. These protocategories are not themselves phonemes, but as Kuhl points out, they

could in principle serve as discrete building blocks for the later construction of a true phonological system.<sup>4</sup>

### 3.3 The Revolution at 8-10 Months

By about eight months, research suggests, babies start to understand words. This coincides, probably not accidentally, with an extraordinary growth of phonological ability, documented in two research traditions.

(1) Werker and Tees (1984) have shown that at this age, babies start to resemble adult speakers in having difficulty in discriminating phonetically similar pairs that do not form a phonemic opposition in their language. What is a loss in phonetic ability is, of course, a gain in phonological ability: the infant is learning to focus her attention on precisely those distinctions which are useful, in the sense that they distinguish words in the target language. This effect has been demonstrated by Werker and Tees for retroflex/alveolar contrasts in Hindi and for uvular/velar contrasts in Nthlakampx.<sup>5</sup>

(2) At more or less the same time, infants start to acquire knowledge of the legal segments and sequences of their language, as Jusczyk et al. (1993, 1994) and Friederici and Wessels (1993) have shown: in carefully monitored experimental situations, babies of this age come to react differently to legal phoneme sequences in their native languages than to illegal or near-illegal ones.<sup>6</sup>

All of the scholars just cited are commendably cautious in making any claims about whether their experiments show that 10-month-old babies can be said to possess a “phonology,” though it is quite clear that what they are acquiring is language-specific. In a speculative vein, however, let us suppose that infants really *are* acquiring phonology, and ponder what might be done to characterize in a formal theory what a 10-month-old has already learned.

## 4. Phonological Knowledge

To clarify this task, it will help to review received wisdom about what kinds of phonological knowledge are possessed by adult speakers. Note that we are speaking here only of *unconscious* knowledge, deduced by the analyst from linguistic behavior and from experimental evidence. Overt, metalinguistic knowledge is ignored here throughout.

---

<sup>4</sup> Thus, for example, some distributional protcategories may turn out to be strongly differentiated allophones of the same phoneme, which are only later united into a single categories as the child learns words and discovers that the protcategories have a predictable distribution.

<sup>5</sup> Best et al. (1988) have shown that English-learning babies do not have difficulty in discriminating a click contrast of Zulu. This is probably unsurprising, given that adult monolinguals can also discriminate contrasts that are not phonemic for them when the phonetic cues are extremely salient. A further relevant factor is that English has no existing phonemes that could be confused with clicks and would distort their perception.

<sup>6</sup> Examples: Dutch \*[rtum], English ?[ji:dʒ]. Many of the sequences used in Jusczyk et al.'s experiment violate formalizable phonotactic restrictions that are exceptionless in English; the others are sufficiently rare that they could in principle be describable as ill formed, from the point of view of the restricted data available to the infant.

There are basically three kinds of phonological knowledge. For each, I will review how such knowledge is currently described formally in Optimality Theory, the approach to phonology assumed here.<sup>7</sup>

#### 4.1 Contrast

To start, phonological knowledge includes knowledge of the system of contrasts: the speaker of French tacitly knows that [b] and [p], which differ minimally in voicing, contrast in French; that is, they can distinguish words such as /bu/ ‘end’ vs. /pu/ ‘louse’. Korean also possesses [b] and [p], but the speaker of Korean tacitly knows that they are contextually predictable variants. Specifically, as shown by Jun (1996), [b] is the allophone of /p/ occurring between voiced sounds when non-initial in the Accentual Phrase.

In Optimality Theory, knowledge of the system of contrasts is reflected in the language-specific rankings (prioritizations) of conflicting constraints. For example, in French the Faithfulness constraint of the IDENT family that governs voicing outranks the various Markedness constraints that govern the default distribution of voicing. This permits representations that differ in voicing to arise in the output of the grammar. In Korean, the opposite ranking holds; thus *even if* Korean had underlying forms that differed in voicing, the grammar would alter their voicing to the phonological defaults; thus no contrast could ever occur in actual speech.<sup>8</sup>

It will be important to bear in mind that in mainstream Optimality Theory, constraint ranking is the *only* way that knowledge of contrast is grammatically encoded: there is no such thing as a (theoretically primitive) “phoneme inventory”, or restrictions on the nature of underlying forms. The experience of analysts applying Optimality Theory to diverse languages shows that such theoretical entities would perform functions that are already carried out adequately by constraint ranking, and they are accordingly dispensed with.

#### 4.2 Legal Structures

The second aspect of phonological knowledge is the **set of legal structures**: specifically, the legal sequencing of phonemes, as well as the structures involved in higher-level prosodic phenomena such as syllables, stress, and tone. The classic example of [blik] vs. \*[bnik] noted above illustrates this sort of knowledge: [blik] constitutes a legal structure of English, and \*[bnik] an illegal one, though neither actually exists. For brevity, I will use the somewhat archaic term **phonotactics** to cover this sort of knowledge: a speaker who knows the phonotactics of a language knows its legal sequences and structures.

---

<sup>7</sup> For reasons of space, I cannot provide a summary of Optimality Theory, now the common currency of a great deal of phonological research. A clear and thoughtful introduction is provided in the textbook of Kager (1999).

<sup>8</sup> And in fact, it is plausible to suppose that Korean learners would never uselessly internalize underlying representations with contrastive voicing, since the distinction could never be realized.

In Optimality Theory, the phonotactics of a language is, just like the system of contrasts, defined exclusively by constraint ranking. In particular, the legal sequences are those for which the Faithfulness constraints that protect them outrank the Markedness constraints that forbid them. As with contrast, theorists have found no reason to invoke any other mechanism than constraint ranking in defining the phonotactics.

### 4.3 Alternation

The third and remaining kind of phonological knowledge is knowledge of the **pattern of alternation**: the differing realizations of the same morpheme in various phonological contexts. To give a commonplace example, the plural ending of English alternates: in neutral contexts it is realized as [z], as in [kænz]; but it is realized as [s] when it follows a voiceless consonant: *cats* [kæts].

The [s] realization is related to the phonotactics in an important way: English does not tolerate final sequences like [tz], in which a voiced obstruent follows a voiceless one. This relationship between phonotactics and alternation is commonplace, and we will return to it below.

### 4.4 Interpreting the Acquisition Literature

Turning now to the acquisition results reviewed earlier, I suggest the following tentative interpretations of them in Optimality-theoretic terms.

**System of contrasts**: the evidence gathered by Werker and her colleagues indicates, at least tentatively, that by the time infants are eight to ten months old, they have gained considerable knowledge of the correct ranking of IDENT constraints with respect to the relevant Markedness constraints, which in OT establishes what is phonemic.

**Phonotactics**: the work of Jusczyk and others has demonstrated, tentatively, that by time babies are eight to ten months old, they have considerable knowledge of the constraint rankings (often Markedness constraints vs. MAX and DEP) that determine the legal phonotactic patterns of their language.

**Pattern of alternation**: ????. I leave question marks for this case, because my literature search has yielded little evidence for just when infants/young children command patterns of alternation. In fact, I believe much interesting work could be done in this area. The next section outlines some findings that seem relevant.

## 5. The Acquisition Timetable for Morphology and Alternation

Learning alternations demands that one have first learned morphology. It makes no sense to say that a morpheme alternates if the learner hasn't yet learned to detect that morpheme as a component substring of the words she knows. If we have good evidence that a child does not know a morpheme, then we can be fairly sure that she doesn't know its pattern of alternation.

It is often feasible to show that a child does not command a particular morpheme. For example, Smith (1973, 17) was able to show that his son Amahl did not command plurals by the following observation: “[At 2;2] Amahl had no contrast anywhere between singular and plural, e.g. [wʌt] and [wi:t] were in free variation for both *foot* and *feet*.” Given this, we can hardly suppose that Amahl had made sense of the alternation pattern ([z]/[s]/[əz]) of the English plural suffix; and indeed, there is evidence (Smith 1973, 17) that Amahl wrongly construed the data as involving an optional process of phonological /z/ deletion.

Note that the age of two years and two months arrives a very long time (as children’s lives go) after ten months. It is thus likely, I think, that Amahl went through a long period in which he tacitly knew that English words cannot end in heterovoiced obstruent sequences, but was in no position to make use of this knowledge to help him with the plural allomorphy seen in *dogs* [dɔgz] and *cats* [kæts].

Some morphology seems to be learned considerably later than this. An extreme case is the non-concatenative morphology of Modern Hebrew, which is rendered particularly difficult by historical changes that rendered the system opaque in various areas. According to Berman’s (1985) study, children learning Modern Hebrew fail to achieve productive command over some parts of the non-concatenative morphology before they reach four to five years of age.

Berko’s (1958) famous “Wug”-testing study, in which children were asked to inflect novel stems like *wug*, also provides support for the view the morphophonemic acquisition happens relatively late. Specifically, quite a few of Berko’s subjects, particularly the four-year-olds, did rather badly on their Wug tests. It seems clear that many of them did not possess full, active command over the patterns of alternation in English inflectional suffixes. Much the same holds true for the children described in a similar study by Baker and Derwing (1982), as well as studies reviewed by Derwing and Baker (1986, 330-331).

The earliest evidence I have seen for command of morphology is correct usage of the Turkish accusative suffix [-a] ~ [-e] at 15 months, documented by Aksu-Koç and Slobin (1985). In principle, knowledge might come earlier, since all evidence I have seen in the literature involves active production by the child rather than experimental tests of perceptual knowledge. Children who can’t talk obviously cannot demonstrate active command over a morpheme.

To sum up this somewhat inconclusive picture: we earlier asked what is the *relative timing* of the acquisition of the three general areas of phonological knowledge—system of contrasts, phonotactics, and pattern of alternation. For the first two, it appears that acquisition is precocious, with much progress made by ten months. For the third, the data are skimpy, and there seems to be quite a bit of variation between morphological processes. Certainly, we can say that there are at least *some* morphological processes which are acquired long after the system of contrasts and legal structures is firmly in place, and it seems a reasonable guess that in general, the learning of patterns of alternation lags the learning of the contrast and phonotactic systems.

A moment's thought indicates why this is a plausible conclusion: for the child to learn a morphological process, she must presumably learn an actual *paradigm* that manifests it (e.g., for English plurals, a set of singular-plural pairs). But the learning of contrasts and phonotactics can take place<sup>9</sup> when the child merely possesses a more-or-less random inventory of words. We thus should expect the learning of alternations to be delayed.

## 6. The Appropriateness of Optimality Theory

I will now argue that current Optimality theoretic approaches are particularly well adapted to modeling the course of acquisition as it is laid out above.

Optimality Theory has been widely adopted by phonologists in part because it solves (or certainly appears to solve) the long-standing problem of **conspiracies**. Early theories of phonology were heavily focused on accounting for alternation, with large banks of phonological rules arranged to derive the allomorphs of the morphemes.<sup>10</sup> It was noticed by Kisseberth (1970) and subsequent work that this alternation-driven approach characteristically missed crucial generalizations about phonologies, generalizations that were characteristically statable as constraints. These include bans on consonant clusters, adjacent stresses, onsetless syllables, and so on. The rules posited in the phonology of the 60's through 80's were said to "conspire" to achieve these surface generalizations; but the generalizations themselves never appeared in the actual analysis. Two decades of research following Kisseberth's article addressed, but never fully solved, the "conspiracy problem."

In Optimality Theory, the treatment of alternation is subordinated to the general characterization of phonotactics in the language. OT delegates the problem of deriving output forms to an entirely general procedure, and dispenses with rules. Under this approach, the conspiracy problem disappears, since the rules that formerly "conspired" are absent, and the target of the conspiracy is itself the core of the analysis.<sup>11</sup>

This theoretical architecture is strongly reminiscent, I think, of the acquisitional sequence laid out in sections 3 and 5 above. In OT, knowledge of contrast and phonotactics is logically prior to knowledge of alternations; and in the acquisition sequence, knowledge of contrast and phonotactics are (at least usually) acquired prior to knowledge of alternations.

More important, I believe that prior knowledge of phonotactics would actually *facilitate* the acquisition of alternation for the child. The reason is that most alternation is directly driven by the need

---

<sup>9</sup> With an important exception, the focus of section 8.4 below.

<sup>10</sup> Indeed, the classical training materials for generative phonologists were problem sets in which the student was (only) required to reduce a pattern of alternation to rule; contrast and phonotactics were relatively neglected. This imbalance is now being corrected under OT, which forces the student to construct a considerably more complete answer.

<sup>11</sup> The personal phonologies created by toddlers (mapping adult surface forms to simplified child outputs) are also conspiratorial, as has been pointed out forcefully by Menn (1983). This is a major rationale for current efforts to use Optimality Theory to model these output systems.

for morphologically-derived sequences to conform to the phonotactics—that is, most alternation is conspiratorial.

To follow up on an earlier example: the English plural suffix is [z] in neutral environments (e.g. *cans* [kænz]) but [əz] after sibilants (*judges* [dʒʌdʒəz], *benches* [bɛntʃəz]) and [s] after voiceless sounds other than sibilants: *cats* [kæts]. The allomorphs [əz] and [s] can be traced directly to patterns of English phonotactics, patterns that can be learned prior to any morphological knowledge. Specifically, English words cannot end in sibilant sequences (hence \*[dʒʌdʒz]), nor can they end in a sequence of the type *voiceless obstruent* + *voiced obstruent* (hence \*[kætz]). Note that these phonotactic constraints hold true in general, and not just of plurals; English has no words of any sort that end in \*[dʒz] or \*[tz]. It is easy to imagine that knowledge of these phonotactic principles, acquired early on, would aid the child in recognizing that [əz] and [s] are allomorphic variants of [z]: [əz] and [s] are minimal alterations of [z] that conform to the phonotactic principles.

To put it slightly more generally: a child who has already achieved a good notion of the phonotactics of her language need not, in general, seek *structural descriptions* to cover cases of regular phonological alternation. These structural descriptions are already implicit in the child's internalized knowledge of phonotactics. All that is necessary is to locate the crucial structural change—or, more precisely, the Faithfulness constraint that must be ranked lower in order for underlying forms to be altered to fit the phonotactics. By way of contrast, earlier rule-based approaches require the learner to find both structural description and change for every alternation, with no help from phonotactic knowledge.

The “Wug”-testing study of Berko (1958) suggests that children actually do make practical use of their phonotactic knowledge in learning alternations. Among the various errors Berko's young subjects made, errors that violate English phonotactics, such as \*[wʌɡz] or \*[ɡʌtʃs] (Berko, pp. 162-163) were quite rare. This observation was confirmed in more detail in the later work of Baker and Derwing (1982). In the view adopted here, the greater degree of reliability young children show in this area follows because they have already learned the phonological constraints that ban the illegal sequences.

Summing up, it would appear that the OT answer to the conspiracy problem is more than just a gain in analytical generality; it is the basis of a plausible acquisition strategy.

## 7. Learning Phonotactics in Optimality Theory

Let us assume, then, that it is appropriate to tailor phonological theory to match acquisition order, letting the prior acquisition of phonotactics aid in the later acquisition of alternations. What I want to focus on at this point is: how might we model the stage occurring at ten months, where the child's knowledge is solely or mostly phonotactic knowledge?

There is now a research tradition within which this question can be explicitly addressed. Its goal is to develop algorithms that, given input data and constraint inventories, can locate appropriate constraint rankings, and thus “learn” phonological systems. Work in this tradition includes Tesar and Smolensky

(1993, 1996, 1998), Pulleyblank and Turkel (to appear), Boersma (1997, 1998), and Boersma and Hayes (in progress).

Constraint ranking algorithms have characteristically attempted to learn whole grammars at a time. But further progress might be possible by taking incremental steps, paralleling those taken by real children. In the present case, the goal is to develop what I will call a **pure phonotactic learner**, defined as follows:

A **pure phonotactic learner** is an algorithm that, given (only) a set of words that are well-formed in a language, creates a grammar that distinguishes well-formed from ill-formed phonological sequences.

Following a commonplace notion in learnability, I will stipulate that a pure phonotactic learner must make no use of negative evidence. That is, while it can be given a long and variegated sequence of examples showing what is well-formed, it can never be overtly told what is ill-formed. This is surely a realistic requirement in the present case.

The rankings that a pure phonotactic learner learns can be tested in the following way: we feed **hypothetical** underlying forms, including illegal ones, to a grammar that respects the rankings that have been learned. If the rankings are correct, the grammar will act as a filter: it will alter any illegal form to something similar which is legal, but it will allow legal forms to persist unaltered. This idea is based on the discussion in Prince and Smolensky (1993, 175).<sup>12</sup>

An intriguing aspect of pure phonotactic learning is that, as far as I can tell, the notion of underlying representation would play no significant role. Specifically, if we consider the two primary purposes to which underlying forms have been put, neither is applicable.

First, in earlier theories of phonology, underlying representations were deemed necessary in order to depict the inventory of contrasting phonological units. As noted above (section 4), the shift to OT renders such a function unnecessary; this was shown by Smolensky (1993) and Kirchner (1997). Both authors show that in OT, the notion of possible contrast is fully encoded in the system of constraint rankings, and that reference to underlying forms is not needed to characterize contrast.

Second, underlying forms are posited as a means of establishing a unifying basis for the set of allomorphs of a morpheme: the allomorphs resemble one another, and diverge in systematic fashion, because each is derived from a unique underlying representation. This second is likewise not needed in

---

<sup>12</sup> The psycholinguistically inclined reader should not scoff at the idea of a grammar being required to rule out hypothetical illegal forms. To the contrary, I think such ability is quite crucial. The real-life connection is speech perception: given the characteristic unclarity and ambiguity of the acoustic input, it is very likely that the human speech perception apparatus considers large numbers of possibilities for what it is hearing. To the extent that some of these possibilities are phonotactically impossible, they can be ruled out even before the hard work of searching the lexicon for a good match-up is undertaken. Thus, many “hypothetical illegal forms” are quite real: they are candidate analyses generated by the speech perception apparatus.

pure phonotactic learning: our (somewhat idealized) assumption is that we are dealing with a stage at which words are not yet parsed into morphemes. In such a system, there are no alternations, so there is no need for underlying forms to account for them.<sup>13</sup>

With both functions of underlying forms dispensed with in the present context, we can suppose that underlying representations are the same as surface representations;<sup>14</sup> this follows the principle of Lexicon Optimization of Prince and Smolensky (1993). In principle, this should help: acquisition can proceed, at least for the moment, without the need to explore the vast set of possible underlying representations corresponding to each surface form. As always with learning, it is a good idea to keep the size of the hypothesis space under control.

### 7.1 Constraint Ranking in Tesar and Smolensky's Model

In trying to design a pure phonotactic learner, I took as my starting point the Constraint Demotion algorithm of Tesar and Smolensky (1993, 1996, 1998). When applied to conventional problems of analysis, Constraint Demotion arrives quite efficiently (in binomial time) at suitable constraint rankings. Constraint Demotion serves here as the base algorithm, to be augmented to form a pure phonotactic learner. The expository tasks at hand are first to review Constraint Demotion, then to show that, without modification, it is not suited to the task of pure phonotactic learning. The version of Constraint Demotion I will review here is the simplest one, namely the "batch" version described in Tesar and Smolensky (1993).

Constraint Demotion is provided with: (1) a set of paired underlying and surface representations; (2) an appropriate set of ill-formed rival outputs for each underlying form, assumed to be provided by the GEN function;<sup>15</sup> (3) an appropriate set of Markedness and Faithfulness constraints; and (4) violation data: the number of times each winning or rival candidate violates each constraint. From this, it finds a ranking (should one exist) that generates of the correct output for each underlying form.

A term that useful in understanding Constraint Demotion is **crucially violated**: a constraint is crucially violated if a winning candidate violates it more times than one of its competing rivals. The leading idea of Constraint Demotion is to demote those constraints that are crucially violated to a

---

<sup>13</sup> Plainly, there is a potential debt to pay here when we consider languages that have elaborate systems of alternation at the phrasal level; for example Kivunjo Chaga (McHugh 1986) or Toba Batak (Hayes 1986). Here, one strategy that might work well would be for the child to focus on very short utterances, where the effects of phrasal phonology would be at a minimum.

<sup>14</sup> I am grateful to Daniel Albrow for suggesting this as a basis for pure phonotactic learning. A similar claim is often made with regard to the child's production system; that is, the input to the production system is the output of the adult system.

<sup>15</sup> The GEN function is clearly the most idealized, and perhaps the most controversial, aspect of OT. For versions of OT that are in a sense "GEN-less", see Ellison (1994), Eisner (1997), and Albrow (1997). In the present context, what is needed is simply a set of phonetically neighboring forms, which let us say in effect "this word is pronounced *this* way, and not these other, similar ways."

position *just low enough* in the hierarchy so that, in the candidate-winnowing process that determines the outputs of an OT grammar, winners will never lose out to rivals.

The detailed workings of the batch version of Constraint Demotion can be summarized as follows:

(1) Find all constraints that are not crucially violated. Place them in a “stratum,” a set of constraints assumed to occur together at the top of the ranking hierarchy.

(2) Where a rival candidate violates a constraint in a newly-established stratum more times than the winner does, it may be considered to be “**explained**”: the winnowing procedure of OT is guaranteed at this point never to select the rival in preference to the winner. Thus, as soon as a rival candidate is explained in this sense, it must be removed from the learning data set, as nothing more can be inferred from it.

(3) Of the constraints that have not yet been placed in a stratum, find those which are not crucially violated in the remaining data. Place them in the next stratum of constraints.

(4) Cull out explained rivals again, as in (2).

(5) Repeat steps (3) and (4) ad libitum, until all the constraints have been assigned to a stratum.

The result (when ranking is successful) is the placement of every constraint in a stratum. As Tesar and Smolensky show (in a formal proof), any ranking of the constraints that respects the stratal hierarchy (so that any constraint in a higher stratum is ranked above any constraint in a lower stratum) will derive only winning candidates.

Sometimes, step (3) of the algorithm yields no constraints at all. In such cases, it turns out, there *is no ranking* of the constraints that will generate the observed set of winners. Thus, Constraint Demotion has the ability to detect failed constraint sets.

The Constraint Demotion algorithm is, in my opinion, an excellent contribution, which opens many avenues to the study of phonological learning. However, it is not suited to the task of pure phonotactic learning, which I will now demonstrate with a simple example.

## 7.2 “Pseudo-Korean”: Basic Pattern and Constraints

Imagine a language in which stops contrast for aspiration; thus /ptk/ and /p<sup>h</sup>t<sup>h</sup>k<sup>h</sup>/ form separate phonemic series and are attested in minimal pairs, such as [tal] ‘moon’ vs. [t<sup>h</sup>al] ‘mask’. Assume further that, while /p<sup>h</sup>t<sup>h</sup>k<sup>h</sup>/ show no significant allophonic variation, /ptk/ are voiced to [bdg] when intervocalic: thus [ke] ‘dog’ but [i ge] ‘this dog’. Assume that the voicing pattern is allophonic; thus [bdg] occur only as the voiced allophones of /ptk/, and never in other positions. Lastly, assume that in final and preconsonantal position, aspiration is neutralized, so that the only legal stops are the voiceless unaspirated [ptk]. Thus while [tʃip<sup>h</sup>i] ‘straw-nom.’ and [tʃibi] ‘house-nom.’ show the phonemic

contrast between /p<sup>h</sup>/ and the [b]-allophone of /p/, this contrast is neutralized to plain [p] in final position, so that unsuffixed [tʃɪp] is in fact ambiguous between ‘straw’ and ‘house’.

This phonological arrangement is essentially what we see in Korean, which is the source of the examples just given. Such arrangements are cross-linguistically quite characteristic. A number of languages voice their unaspirated stops intervocalically (Keating, Linker and Huffman 1983), and it is common for languages to suspend contrasts for laryngeal features in positions other than prevocalic (Steriade 1997). I will call the hypothetical example language “Pseudo-Korean”, since all the phenomena of Pseudo-Korean occur in Korean, but Pseudo-Korean has only a small subset of the Korean phenomena.

A suitable set of constraints for analyzing the Pseudo-Korean pattern is given below.

### 7.2.1 *Markedness Constraints*

(1) \*[-SON, +VOICE]

The default, normal state of obstruents is voiceless, for aerodynamic reasons laid out in Ohala (1983) and Westbury and Keating (1986). The constraint above encodes this phonetic tendency as a grammatical principle.

(2) \*[+VOICE][-VOICE][+VOICE]                      (abbreviation: \*[+V][-V][+V])

This constraint bans voiceless segments surrounded by voiced ones. The teleology of the constraint is presumably articulatory: forms that obey this constraint need not execute the laryngeal gestures needed to turn off voicing in a circumvoiced environment. For evidence bearing on this point from an aerodynamic model, see Westbury and Keating (1986).

With two constraints in hand, we may consider their ranking. All else being equal, where \*[-SON, +VOICE] dominates \*[+V][-V][+V], obstruents will be voiceless everywhere; Keating et al. (1983) note that this is the pattern found in Hawaiian and various other languages. Under the opposite ranking, obstruents are voiced in voiced surroundings but voiceless elsewhere. This ranking prevails in (Pseudo-) Korean.

(3) \*[+SPREAD GLOTTIS]                                      (abbr. \*ASPIRATION)

This constraint, too, has an articulatory teleology: aspiration involves a glottal abduction gesture of considerable magnitude.

(4) \*[+VOICE, +SPREAD GLOTTIS] (abbr. \*D<sup>h</sup>)

Voicing and aspiration are inherently not very compatible, and indeed most languages lack voiced aspirates. Note that \*D<sup>h</sup> bans a subset (a particularly difficult subset) of the cases banned by \*ASPIRATION.

Let us consider ranking again. In pseudo-Korean, the pattern of intervocalic voicing reflects a three-way ranking among the Markedness constraints. Aspirated stops cannot be voiced intervocalically, because (as we will show more carefully below)  $*D^h$  outranks  $*[+V][-V][+V]$ . Adding this to the ranking we saw before, we may now record the three-way ranking  $*D^h \gg * [+V][-V][+V] \gg *[-SON, +VOICE]$ . This is one of the ranking sets that will have to be learned.

### 7.2.2 Faithfulness Constraints

#### (5) IDENT(ASP) / \_\_\_ V

This constraint is based on the work of Steriade (1997), who shows that aspiration and other laryngeal contrasts gravitate cross-linguistically to prevocalic position. In Steriade's view, this has an acoustic explanation: vowels provide a clear "backdrop" against which aspiration and other laryngeal phenomena can be perceived; and languages characteristically limit their phonemic contrasts to locations where perceptibility is maximized.<sup>16</sup>

#### (6) IDENT(ASP)

This is the general, context-free constraint for Faithfulness in aspiration.

The type of aspiration pattern a language will allow depends on the ranking of \*ASP in the hierarchy: if \*ASP is on top, then aspiration will be missing entirely (as in French); if \*ASP is outranked by IDENT(ASP) / \_\_\_ V, then aspiration will occur only prevocalically (as in Korean and Pseudo-Korean); and if \*ASP is at the bottom, then aspiration will be possible in all positions (as in Hindi).

#### (7) IDENT(VOICE) / \_\_\_ V

#### (8) IDENT(VOICE)

These two constraints work just like the analogous constraints for aspiration, though (7) turns out to be ranked differently in Pseudo-Korean.

### 7.3 Pseudo-Korean: Candidates

The challenge provided in the Pseudo-Korean ranking problem is to arrive at a ranking that generates only phonotactically legal forms. To provide a reasonable test, I developed a large set of Pseudo-Korean forms, with numerous rival candidates for each. The full set may be downloaded;<sup>17</sup> a representative subset is given below:

---

<sup>16</sup> There is a current open research issue in OT: whether contextual information properly belongs within the Markedness constraints or the Faithfulness constraints. For useful argumentation on this point, see Zoll (1998). The account given here places context in the Faithfulness constraints; I have also tried a parallel simulation using the opposite strategy, and obtained very similar results.

<sup>17</sup> <http://humnet.ucla.edu/humnet/linguistics/people/hayes/>

Input	Winning Output	Rivals
/ta/	[ta]	*[t <sup>h</sup> a], *[da], *[d <sup>h</sup> a]
/ada/	[ada]	*[ata], *[at <sup>h</sup> a], *[ad <sup>h</sup> a]
/t <sup>h</sup> a/	[t <sup>h</sup> a]	*[ta], *[da], *[d <sup>h</sup> a]
/at <sup>h</sup> a/	[at <sup>h</sup> a]	*[ata], *[ada], *[ad <sup>h</sup> a]
/at/	[at]	*[ad], *[at <sup>h</sup> ], *[ad <sup>h</sup> ]
/tada/	[tada]	*[dada], *[t <sup>h</sup> ada], *[d <sup>h</sup> ada], *[tata], *[tat <sup>h</sup> a], *[tad <sup>h</sup> a]
/tat <sup>h</sup> a/	[tat <sup>h</sup> a]	*[dat <sup>h</sup> a], *[d <sup>h</sup> at <sup>h</sup> a], *[t <sup>h</sup> at <sup>h</sup> a], *[tata], *[tada], *[tad <sup>h</sup> a]
/t <sup>h</sup> ada/	[t <sup>h</sup> ada]	*[t <sup>h</sup> ata], *[t <sup>h</sup> ad <sup>h</sup> a], *[tada], *[dada], *[d <sup>h</sup> ada], *[t <sup>h</sup> at <sup>h</sup> a]
/t <sup>h</sup> at <sup>h</sup> a/	[t <sup>h</sup> at <sup>h</sup> a]	*[tat <sup>h</sup> a], *[dat <sup>h</sup> a], *[d <sup>h</sup> at <sup>h</sup> a], *[t <sup>h</sup> ata], *[t <sup>h</sup> ada], *[t <sup>h</sup> ad <sup>h</sup> a], *[tata], *[tada]
/tat/	[tat]	*[t <sup>h</sup> at], *[dat], *[d <sup>h</sup> at], *[tat <sup>h</sup> ], *[tad], *[tad <sup>h</sup> ]
/t <sup>h</sup> at/	[t <sup>h</sup> at]	*[tat], *[dat], *[d <sup>h</sup> at], *[t <sup>h</sup> ad], *[t <sup>h</sup> at <sup>h</sup> ], *[t <sup>h</sup> ad <sup>h</sup> ]

Following the assumption made above in section 7, I consistently made the underlying form the same as the winning candidate. Note that all the forms in the simulation were *legal surface forms* of Pseudo-Korean, such as would be heard in real-life data; thus, the training set provided only positive evidence.

#### 7.4 Application of Constraint Demotion to Pseudo-Korean

I will now show how, and why, Tesar and Smolensky's Constraint Demotion algorithm is not suited to pure phonotactic learning.

I submitted the following material to a software implementation of Constraint Demotion:<sup>18</sup> the full set of Pseudo-Korean inputs, winners, and rivals, the Markedness and Faithfulness constraints given in 7.2.1 and 7.2.2, and a machine-generated set of violations for every constraint and candidate combination. Constraint Demotion performed its work, and output the following strata:

---

<sup>18</sup> The software is available at the Web address given in the preceding footnote. It has been used extensively in teaching and appears from experience to be a fully faithful rendition of Constraint Demotion.

**Stratum #1**

IDENT(ASP)

IDENT(VOICE)

IDENT(ASP) / \_\_\_\_ V

IDENT(VOICE) / \_\_\_\_ V

\*D<sup>h</sup>**Stratum #2**

\*[+V][-V][+V]

\*[-SON/+VOICE]

\*ASPIRATION

This outcome is, in fact, a perfectly good grammar for the data that fed it, in the sense that it generates the correct outcome for every input form. But it is *not* a good grammar for Pseudo-Korean, because it fails to describe Pseudo-Korean phonotactics. Indeed, under this grammar, *any* combination of voicing and aspiration is legal in any position, contrary to the facts of the language.

To illustrate this, I added to the Pseudo-Korean “data” a set of underlying forms that are illegal in Pseudo-Korean. For each one, I provided all reasonable logical possibilities as candidates, and this time simply checked what emerged as the winner.<sup>19</sup> Some representative forms resulting from this procedure were as follows:

Input	Choices for output	Output
/da/	[da], [ta], [t <sup>h</sup> a], [d <sup>h</sup> a]	*[da]
/d <sup>h</sup> a/	[d <sup>h</sup> a], [t <sup>h</sup> a], [ta], [da]	*[d <sup>h</sup> a] or *[t <sup>h</sup> a] or *[da]
/ata/	[ata], [ada], [ad <sup>h</sup> a], [at <sup>h</sup> a]	*[ata]
/ad <sup>h</sup> a/	[ad <sup>h</sup> a], [at <sup>h</sup> a], [ata], [ada]	*[ad <sup>h</sup> a] or *[at <sup>h</sup> a] or [ada]
/ad/	[ad], [at], [ad <sup>h</sup> ], [at <sup>h</sup> ]	*[ad]
/at <sup>h</sup> /	[at <sup>h</sup> ], [at], [ad], [ad <sup>h</sup> ]	*[at <sup>h</sup> ]
/ad <sup>h</sup> /	[ad <sup>h</sup> ], [at], [ad], [at <sup>h</sup> ]	*[ad <sup>h</sup> ] or [at <sup>h</sup> ] or *[ad]

The crucial point is a large number of illegal forms were generated. It also can be noted in passing that there was also a good deal of free variation: it matters how \*D<sup>h</sup> is ranked with respect to the Faithfulness constraints; but the original learning data do not suffice to establish this ranking.

The basis of the bad outcomes is not hard to see: since all the Faithfulness constraints are at the top of the hierarchy, it is always possible to generate an output that is identical (or at least, very similar) to an illegal input. Moreover, this is an *inevitable* result, given the nature of Constraint Demotion as

---

<sup>19</sup> Method used: compute a “constrained factorial typology”, namely the full set of outcomes under all complete rankings that respect the stratal arrangement created by Constraint Demotion.

applied to the learning data of the type considered here. Given that only positive data are considered, and that underlying forms are always identical to surface forms, the Faithfulness constraints of the grammar are never violated in the learning data. Therefore, they are all placed in the topmost stratum, which is too high.

Recall now what we wanted our grammar to do: given a legal input, it should simply reproduce it as an output; and given an illegal input, it should alter it to form a legal output. It is evident that the ranking learned by Constraint Demotion succeeds in the first task, but not the second.

### 7.5 *Adapting Constraint Demotion to Pure-Phonotactic Learning*

A fundamental idea that has emerged from recent theoretical acquisition work in OT (Gnanadesikan 1995, Smolensky 1996) is that Faithfulness constraints should be assigned a default location at the bottom of the constraint hierarchy. This idea has thus far been applied only to the “production grammars” that children use for their own outputs: the gradual approximation by the child’s own output to adult speech reflects a gradual rise of the Faithfulness constraints upward from the bottom of the hierarchy. When they are at the bottom, children can’t say anything; when they have reached adult level, child speech becomes the same as adult speech.

The same idea, however, can be used to analyze the child’s (actually, the infant’s) passively internalized conception of the adult language. Here, the need to favor rankings with low Faithfulness arises for a different reason: the problem of learning in the absence of negative evidence. To learn that (say) \*[ad] is ill-formed, Pseudo-Korean infants must use a conservative strategy, plausibly along the lines “if you haven’t heard it, or something like it, then it’s not possible.” In the present, Optimality-theoretic context, this has a specific formal translation: we must locate a constraint ranking that places Faithfulness as low as possible.

Note that by using general phonological constraints, we can in principle solve a major problem. We don’t want the finished grammar to admit only those words that it has heard before; rather, we want the grammar to *project* beyond this minimum to allow similar forms. Thus (to take up a familiar example again) English speakers accept *blick* [blik] as well-formed because the real words they learned in childhood (such as *blink* and *kick*) led them to adopt a constraint ranking in which *blick* emerges as a legal form. It is the phonological generality of the constraint inventory that makes this possible.

Turning to the question of an actual algorithm: what we want is an algorithm that will produce a ranking that (a) correctly derives all attested forms; and (b) places the Faithfulness constraints as low as possible, in some sense yet to be defined precisely.

### 7.6 *Low-Faithfulness Constraint Demotion*

My proposed Low-Faithfulness Constraint Demotion algorithm is identical to the batch version of Tesar and Smolensky’s algorithm, with the following crucial exception. Whenever a new stratum is to be created (that is, at either stage 1 or stage 3 of Constraint Demotion, given in 7.1 above), the criteria

that a constraint must pass to be eligible to be installed in the new stratum are made more stringent. There are three areas of increased stringency. As I have determined by trying out the algorithm on various input files, all three are crucial to learning success.

### 7.6.1 *Bias Against Installing Faithfulness Constraints in a New Stratum*

Suppose that the set of constraints that are not crucially violated in the remaining learning data includes *both Faithfulness and Markedness constraints*. In such cases, only the Markedness constraints are allowed to join the new stratum. The Faithfulness constraints must await a later opportunity to be ranked, often the next stratum down.

Here is the rationale: often a rival candidate can be ruled out either because it violates a Markedness constraint, or because it is unfaithful. In such cases, we want the Markedness constraint to do the job, because if we let Faithfulness do it, it is likely to lead to overgeneration in the finished grammar.

### 7.6.2 *Only Specific Members of Specific/General Pairs May be Installed*

Often, two Faithfulness constraints have violation patterns that are in a subset relation. Thus, for example, the violations of IDENT(ASP) / \_\_\_ V form a subset of the violations of IDENT(ASP). In such cases, where both constraints are eligible to join the current stratum (because both are not crucially violated), only the more specific constraint (the one with a subset of the other's violations) is admitted. This will push the general constraint down to a lower stratum, again often resulting in a tighter grammar.

### 7.6.3 *Only Effective Faithfulness Constraints May be Installed*

Suppose now that we have cut down the list of constraints eligible for installation in the current stratum to a set F, consisting of Faithfulness constraints that are not yet ranked, not crucially violated, and not excluded on grounds of occurring in a general/specific relation. Let C be some member of F. It is sensible to require that C actually "do some work" in order to be installed in the newly formed stratum. We can do this by requiring that C exclude at least one ill-formed rival candidate R that is yet unexplained. (By "exclude" it is meant that R violates C more times than the winning candidate.)

In simple cases, this procedure suffices. However, in more complex situations, multiple Faithfulness constraints often rule out the very same rival candidates in parallel. To tease apart such cases, the algorithm should require that constraint C exclude some rival *without "help"* from any other constraint. Should this fail to locate at least one "installable" constraint, the criterion is loosened: C must exclude some rival with the help of just one other constraint, or just two, and so on. Eventually, this procedure usually locates a constraint or set of constraints that can be installed in the new stratum.

By only allowing maximally effective Faithfulness constraints to be installed, we can limit installations to cases that are most likely to be truly necessary, letting the ineffective constraints sink further down, to a point where they will not lead to overgeneration.

#### 7.6.4 Termination

The three conditions above sometimes result in the inability to form any stratum at all. This will occur when all remaining constraints are Faithfulness constraints that do not exclude any rivals. Once this has happened, the remaining constraints are simply relegated to a final stratum, placed below all others.

Other than what has just been said, Low-Faithfulness Constraint Demotion works just like regular Constraint Demotion.

#### 7.7 Pseudo-Korean and Low-Faithfulness Constraint Demotion

To test Low-Faithfulness Constraint Demotion, I implemented it as a computer program and fed it the same input file, containing only well-formed examples, that I had earlier given to regular Constraint Demotion. The new algorithm ranked the constraints as follows:

##### **Stratum #1**

\*d<sup>h</sup>

##### **Stratum #2**

IDENT(ASP) / \_\_\_ V

##### **Stratum #3**

\*[+V][-V][+V]

\*ASPIRATION

##### **Stratum #4**

\*[-SON/+VOICE]

##### **Stratum #5**

IDENT(ASP)

IDENT(VOICE)

IDENT(VOICE) / \_\_\_ V

The strata can be seen to form alternating bands of Markedness and Faithfulness, just as we would expect given the restrictions made on when a constraint may be installed in a stratum. More important, we can inspect the ranking of the Faithfulness constraints and see exactly what is phonemic in Pseudo-Korean stops: aspiration in prevocalic position. This is because IDENT(ASP) / \_\_\_ V is the only Faithfulness constraint that doesn't reside at the bottom of the grammar. Voicing is allophonic, and aspiration in non-prevocalic position is likewise predictable.

The crucial test for this grammar, however, is: does it overgenerate? To test this, I fed the grammar the larger set of inputs which had earlier shown that regular Constraint Demotion overgenerates. From these inputs, the new grammar derived outputs like the following:

Well-Formed Inputs		Ill-Formed Inputs	
input	output	input	output
/ta/	[ta]	/da/	[ta]
/ada/	[ada]	/d <sup>h</sup> a/	[t <sup>h</sup> a]
/t <sup>h</sup> a/	[t <sup>h</sup> a]	/ata/	[ada]
/at <sup>h</sup> a/	[at <sup>h</sup> a]	/ad <sup>h</sup> a/	[at <sup>h</sup> a]
/at/	[at]	/ad/	[at]
/tada/	[tada]	/at <sup>h</sup> /	[at]
/tat <sup>h</sup> a/	[tat <sup>h</sup> a]	/ad <sup>h</sup> /	[at]
/t <sup>h</sup> ada/	[t <sup>h</sup> ada]		
/t <sup>h</sup> at <sup>h</sup> a/	[t <sup>h</sup> at <sup>h</sup> a]		
/tat/	[tat]		
/t <sup>h</sup> at/	[t <sup>h</sup> at]		

Specifically, all well-formed inputs were retained, and all ill-formed inputs were “fixed”; that is, converted by the grammar into a well-formed output.<sup>20</sup> Thus, Low-Faithfulness Constraint Demotion succeeded in learning a ranking that defines Pseudo-Korean phonotactics, based on only positive evidence.

I have tried out Low-Faithfulness Constraint Demotion on a number of data files similar in scope to Pseudo-Korean.<sup>21</sup> So far, I have found that it always succeeds in producing “tight” grammars, which generate only forms that match (or are less marked than)<sup>22</sup> those given to it in the input.

### 7.8 Caveats

To recapitulate: in the approach taken here, infants are able to learn a lot about their phonology (specifically, what structures are ill-formed or well-formed) in the absence of any negative evidence. Moreover, they apparently accomplish their learning with little or no information about morphology and

<sup>20</sup> The reader may have noted that the “fixes” imposed by the grammar conform to the behavior of alternating forms in real Korean. This outcome is accidental. A larger Pseudo-Korean simulation, not reported here, included candidates with deletion and insertion, and indeed uncovered grammars in which illegal forms were repaired by vowel epenthesis and consonant deletion, rather than by alternation of laryngeal feature values.

<sup>21</sup> Specifically: a file with the legal vowel sequences of (the native vocabulary of) Turkish, and a family of files containing schematic “CV” languages of the familiar type, that is, languages banning codas, requiring onsets, banning hiatus, and so on.

<sup>22</sup> Thus, for instance, when given a (rather unrealistic) input set consisting solely of [CV.V], the algorithm arrived at the view that [CV.CV] is also well formed. This is because, given the constraints that were used, there was no ranking available that would permit [CV.V] but rule out [CV.CV]. This fits in with a general prediction made by Optimality Theory, not just with Low-Faithfulness Constraint Demotion: in any language, a hypothetical form that incurs a subset of the Markedness violations of any actual form should be regarded by speakers as well-formed. This is a hard claim to test, since our knowledge of what is a possible Markedness constraint remains sketchy.

alternations, hence in the absence of knowledge of underlying forms. I have developed a ranking algorithm with the goal of demonstrating the feasibility of this kind of acquisition: that in principle it can be done, given an algorithm that suitably downgrades the ranking of Faithfulness constraints whenever possible.

This said, I wish to mention three limitations of Low-Faithfulness Constraint Demotion.

First, the algorithm cannot deal with the fact that judgments of phonotactic well-formedness are gradient (Algeo 1978); for example, a form like  $?[dw\epsilon f]$  seems neither perfectly right nor completely ill-formed. There *is* an algorithm that has proven capable of treating gradient well-formedness, namely the Gradual Learning Algorithm of Boersma (1997, 1998), applied to gradient well-formedness in Boersma and Hayes (in progress). I have not yet succeeded in incorporating a suitable downward bias for Faithfulness into this algorithm.

Second, I find it a source of discontent that Low-Faithfulness Constraint Demotion, like regular Constraint Demotion, relies so heavily on *a priori* knowledge: specifically, a universal inventory of constraints and a universal feature system. It would count as a considerable advance, I think, if it could be shown that these analytical elements are themselves learnable. For discussion along these lines, see Boersma (1998).

Lastly, Low-Faithfulness Constraint Demotion, unlike Constraint Demotion, is not backed by a mathematical proof of its effectiveness. Indeed, there is not even a criterion, other than phonologists' judgments, as to what should be considered effective.<sup>23</sup>

In the face of these caveats, I would take the main point of my study to be a demonstration of feasibility. In particular: the information for learning phonotactics really does seem to be there, negative evidence is not required, and at least in fairly simple cases the necessary ranking can be learned by an algorithm.

## 8. The Learning of Alternations

What happens in the phonology as the child comes to parse words into their component morphemes and starts to notice alternations? There are various possibilities here. One is that the morphemes are assigned underlying forms, which abstract away from the variety of surface realizations in a way that permits all allomorphs to be derived. More recently, various proposals within OT have been made that call into question the need for positing underlying representations at all. In such theories, the task of guaranteeing that allomorphs should resemble one another is taken over by the constraint system. This is accomplished with "output-to-output" correspondence constraints (see for example

---

<sup>23</sup> One possible formal effectiveness criterion might be: if for any pair of rankings  $R$  and  $R'$ ,  $R$  permits only a subset of the forms permitted by  $R'$ , then an effective algorithm will never settle for  $R'$  in preference to  $R$  (though it might opt for a third, still better ranking). In the end, though, I think the right criterion for any algorithm should be empirical, namely the ability to match human intuition.

Burzio 1996, Kager 1996, Steriade 1996, Benua 1997, Hayes 1997, Kenstowicz 1997), which directly enforce uniformity within paradigms.

It will not be crucial here to decide whether underlying representations exist, but I will be invoking output-to-output constraints in the discussion below. In what follows, I will outline some areas in which further learning must adjust, or perhaps retract, conclusions made by the child during the earliest stages, and some suggestions about how this might be accomplished.

### 8.1 *Knowing the Phonotactics First is Helpful*

To start, it is worth reemphasizing a point made above: because phonology is conspiratorial, knowing the phonotactics in advance of morphology is a powerful tool to use in learning alternation. We have seen three cases above: the accommodation of the English plural suffix /z/ to voiceless and sibilant-final stems (section 6); the appearance of voicing in Korean plain stops in intervocalic position (7.2); and the disappearance of aspiration on Korean stops when they occur in final position (7.2). We have also seen that experimental work in “Wug” testing supports this view.

### 8.2 *A Trim-Back Problem: Grammatically-Conditioned Allophones*

An interesting problem for the study of post-infancy phonological learning is posed by the existence of “grammatically-conditioned allophones”: sounds whose distribution is predictable, but only if one knows the grammatical structure of the words in question.

Such allophones arise in part from what the classical generative literature called “boundary phenomena”: instances where a stem + affix combination receives a different treatment than the same sequence occurring within a morpheme. For instance, in many English idiolects *bonus* [ˈbõʊnəs], with nasalized [õ], fails to form a perfect rhyme with *slowness* [ˈslõʊnəs], with oral [ou]. In traditional terms, nasalization is said to be “blocked across the suffix boundary.” A similar case, worked out in Optimality-theoretic terms in Hayes (in press), is the non-rhyming pair *holy* [holi] vs. *slowly* [slõʊli]: *slowly* avoids the monophthongal [o] characteristic of pre-/l/ position in stems, and thus shows blockage of monophthongization across the suffix boundary.<sup>24</sup> Further cases are cited by Kiparsky (1988, 367).

Another case of grammatically-conditioned allophones is found with the dialectal English forms *writer* [ˈrʌɪtə] and *rider* [ˈraɪdə]. These are well known from the early rule-ordering analysis of Chomsky (1964): in Chomsky’s account, /raɪt+ə/ becomes intermediate [rʌɪt+ə] by raising before voiceless consonants, then [rʌɪtə] by Flapping of pre-atonic intervocalic /t/. The result is a surface minimal pair.

---

<sup>24</sup> The facts just given hold for most American speech. Readers who speak a dialect in which *holy* and *slowly* form a perfect rhyme may be able to locate a similar effect in their own speech if they compare *holy* with *holey* ‘filled with holes’; in non-American dialects these often appear as [ˈhõli] vs. [ˈholi]. In traditional terms this would also be a boundary effect: the relevant dialects deploy [õ] only before pre-boundary /l/.

Infants, who often lack the morphological knowledge needed to identify grammatically-conditioned allophones, are liable to mistake them for cases of outright phonemic contrast.<sup>25</sup> Assuming (correctly, I think) that such sounds do not have phonemic status for adults, we thus have an important question: how can the older child, who has come to know the relevant morphology, do the **backtracking** needed to achieve a full understanding of the system?

I believe there is a straightforward way to do this, based on output-to-output correspondence constraints.

### 8.3 *The Ranking of Output-to-Output Correspondence*

There is evidence that output-to-output correspondence constraints are ranked a priori very high—probably, undominated—by the child. The evidence for this is that children are able to innovate sequences that are illegal in the target language, in the interest of maintaining output-to-output correspondence. This was observed by Kazazis (1969) in the speech of Marina, a four-year-old learning Modern Greek. Marina innovated the sequence \*[xe] (velar consonant before front vowel), which is quite illegal in the target language. She did this in the course of regularizing the verbal paradigm: thus [ˈexete] ‘you-pl. have’ (adult [ˈeçete]), on the model of [ˈexo] ‘I have’.

The example is interesting from the viewpoint of the *a priori* assumptions brought by the child to acquisition. Marina presumably had never heard an adult say [xe], and had every reason to think that the constraint banning it should be ranked at the top of the hierarchy. Yet she ranked an output-to-output correspondence constraint (the one regulating the [x]/[ç] distinction) even higher, to establish a non-alternating paradigm. A reasonable guess, then, is that output-to-output correspondence constraints have a default ranking at the very top of the hierarchy, and that they are demoted only as the child processes the evidence that justifies their demotion.<sup>26</sup>

### 8.4 *Output-to-Output Constraints Facilitate Backtracking*

Let us now return to grammatically-conditioned allophones and the backtracking problem. One important point about grammatically-conditioned allophones is that they seem quite generally to be amenable to analyses making use of output-to-output correspondence. For example, the oral vowel of *slowness* [ˈslounəs] is plausibly attributed to a correspondence effect with its base form *slow* [ˈslou], where orality is phonologically expected. Likewise the diphthongal vowel quality of /ou/ of *slowly*

---

<sup>25</sup> The reader who doubts this might further consider the effects of forms that are not entirely morphologically transparent. For example, not every learner will realize immediately that a *nightie* [ˈnɪɹi] is so called because it is worn at night; hence it will form a near-minimal pair with, e.g. *Heidi* [ˈhaɹi]. A child exposed to the children’s book character “Lowly [ˈloʊli] Worm” will not necessarily be aware that most worms live underground (Lowly doesn’t); and will take *Lowly* to form a near-minimal pair with, e.g. *roly-poly* [ˈroli ˈpoli].

<sup>26</sup> Note that innovation of grammatically-conditioned allophones probably arises historically from the same effects seen synchronically in Marina. Had Marina been able to transmit her innovation to the speech community as a whole, then Modern Greek would have come to have [x] and [ç] as grammatically-conditioned allophones.

[slouli] can be treated as a correspondence effect from the same base. The pair *writer* [ˈrʌɪrə] vs. *rider* [ˈraɪrə], though treated very differently in traditional phonology, likewise emerges as a correspondence effect: *writer* inherits its raised diphthong from the base form *write* [ˈraɪt], where it is justified by a phonetically-grounded Markedness constraint that forces raising.

Output-to-output correspondence provides a plausible strategy by which the child could backtrack, undoing earlier errors on grammatically-conditioned allophones. The two elements of the strategy are as follows. First, as just proposed, Output-to-Output correspondence must be ranked a priori high. Second, the Faithfulness constraints must be forced to continue to “justify themselves” throughout later childhood, by continuing to rule out ill-formed rival candidates. Otherwise, they are allowed to sink back down in the ranking.

Here is how the scheme would work. As soon as the child learns that (say) *lowly* is derived from *low* [ˈloʊ], she will expect its pronunciation to be [ˈlouli] *irrespective of the ranking of the relevant Faithfulness constraints*. This is because the output-to-output correspondence constraint governing diphthongal [ou] quality is a priori undominated. At this point, *lowly* can no longer serve as an input datum to justify a high ranking for the Faithfulness constraints that support the putative [o]/[ou] distinction. After the other relevant forms are also morphologically analyzed, then the entire burden of accounting for the [o]/[ou] distinction is assumed by output-to-output correspondence, and the erstwhile dominant Faithfulness constraints may safely sink to the bottom of the grammar. The end result is that [o]/[ou] ceases to be a phonemic distinction.

Naturally, where there *is* phonological alternation, the learning process must demote the output-to-output correspondence constraints that would block it. Thus, for example, when the child comes to know that *hitting* [ˈhɪtɪŋ] is derived from *hit* [hɪt], she must demote the constraints IDENT-OO(VOICE) and IDENT-OO(SONORANT), which preserve the distinction of [t] vs. [ɾ], from their originally undominated position.

### 8.5 A Stage of Vulnerability

If the view taken here is correct, then children often go through a stage of innocent delusion: they wrongly believe that certain phones which are lawfully distributed according to a grammatical environment are separate phonemes. The effects of this errorful stage can be seen, I think, in cases where the erroneous belief is accidentally cemented in place by the effects of dialect borrowing.

Consider the varieties of American English noted above in which *writer* [ˈrʌɪrə] and *rider* [ˈraɪrə] form a minimal pair. As just mentioned, they can be analyzed in OT with Markedness constraints that require the appearance of the raised diphthong [ɪɪ] before voiceless consonants (accounting for [ˈrʌɪt]), along with an undominated output-to-output correspondence constraint requiring the vowel quality of bases to be carried over to their morphological derivatives. But to the infant who does not yet understand the morphology, [ˈrʌɪrə] vs. [ˈraɪrə] looks just like a minimal pair.

Further light on the *writer/reader* phenomenon was shed by Vance (1982), who made a careful study of the idiolects of three native speakers. Vance elicited hundreds of relevant words from his consultants, and made a striking discovery. For these speakers, [ʌɪ] and [aɪ] are *phonemes*, with a fair number of straightforward, monomorphemic minimal and near-minimal pairs. There was much variation among the three consultants, but at least one of Vance's speakers provided each of the following cases:

<i>idle</i>	[ <sup>1</sup> ʌɪrə]	<i>idol</i>	[ <sup>1</sup> aɪrə]
<i>tire</i>	[ <sup>1</sup> tʌɪ]	<i>dire</i>	[ <sup>1</sup> daɪ]
<i>bicycle</i>	[ <sup>1</sup> bʌɪsəkə]	<i>bison</i>	[ <sup>1</sup> baisən]
<i>miter</i>	[ <sup>1</sup> mʌɪrə]	<i>colitis</i>	[kə <sup>1</sup> laɪrəs]

It is plausible to imagine that the newly phonemic status of [ʌɪ] and [aɪ] for these speakers had its origin in the failure to do the crucial backtracking. For backtracking to be successful, [ʌɪ] must be discovered to be a grammatically-conditioned allophone. Instead, it was kept as a phoneme.

Why did this happen? A reasonable guess can be based on the extreme geographic mobility of American English speakers: [<sup>1</sup>rʌɪrə]/[<sup>1</sup>raɪrə] speakers are constantly migrating to [<sup>1</sup>raɪrə]/[<sup>1</sup>raɪrə] dialect regions, and vice versa. The [<sup>1</sup>raɪrə]/[<sup>1</sup>raɪrə] speakers of course have no [ʌɪ], and say *bison* [<sup>1</sup>baisən], *colitis* [kə<sup>1</sup>laɪrəs], and so on. If a young learner of the [<sup>1</sup>rʌɪrə]/[<sup>1</sup>raɪrə] dialect encountered such speakers during the crucial period of vulnerability, it might indeed prove fatal to the delicate restructuring process described above, whereby what the child thought were phonemes are restructured as grammatically conditioned allophones. Note in particular that the crucial “contaminating” words would likely be encountered from different speakers more or less at random. This fits in well with the rather chaotic situation of lexical and interspeaker variation that Vance found.

It can be added that children whose primary learning source comes from the [<sup>1</sup>raɪrə]/[<sup>1</sup>raɪrə] dialect are *not* analogously susceptible when they are exposed to migratory [<sup>1</sup>rʌɪrə]/[<sup>1</sup>raɪrə] speakers. For these children, [ʌɪ] and [aɪ] are never distinct phonological categories—indeed, they probably never even make it to the status of distributional protocategories (section 3.2). When such children hear outsiders say [<sup>1</sup>rʌɪrə] and [<sup>1</sup>raɪrə], they will mostly likely simply fail to register the difference, which is of course the normal way that listeners hear phonetically similar sounds that are not phonemic for them. Indeed, my impression is that, unlike [<sup>1</sup>rʌɪrə]/[<sup>1</sup>raɪrə] speakers, adult [<sup>1</sup>raɪrə]/[<sup>1</sup>raɪrə] speakers find the [ʌɪ]/[aɪ] distinction to be rather difficult to hear.

Summing up: the overall view taken here that the acquisition of contrast and phonotactics precedes the acquisition of alternations is supported by the vulnerability of young children to dialect contamination. Since the order of acquisition forces them to assume that what ought to be grammatically-conditioned allophones are simply phonemes, exposure to forms from other dialects readily upsets the former system, turning the former allophones into phonemes in the restructured system.

## 9. Synoptic View of Phonological Acquisition

To conclude, we can now assemble the discussion above into a view of phonological acquisition as a whole, which uses Optimality Theory to model the learning process. It is worth pointing out that this scheme involves three types of default ranking.

1) **Starting point.** Phonological learning is facilitated by good language design: through processes that are not well understood, languages come to place their phoneme boundaries at locations that render distinct phonemes readily discriminable, by matching phoneme boundaries with inherent auditory boundaries (Eimas et al. 1971 and subsequent work).

2) **Distributional protcategories.** By the age of six months, infants have used knowledge of the statistical distribution of tokens to establish language-specific distributional protcategories, which form the currency of computation for later phonological acquisition (Kuhl 1995; Guenther and Gjaja 1996).

3) **Acquisition of “pure phonotactics”.** At eight to ten months, infants make very rapid progress in learning the pattern of contrast and phonotactics in their language. They do this largely in ignorance of morphology, and thus (following current OT assumptions) in a model in which underlying and surface representations are the same.

In view presented here, learning at this phase takes place through the ranking of Faithfulness constraints against Markedness constraints, on the basis of positive evidence only. It is assumed (given how effectively they perform the task) that infants must be using some very efficient algorithm, for which Low Faithfulness Constraint Demotion (section 7.6) is intended as a first approximation. What is crucial about this algorithm is that it is designed to place the Faithfulness constraints as low as possible. The prejudice in favor of low Faithfulness implements the common-sense idea “if you haven’t heard it, or something like it, then it’s ill-formed.”

4) **Learning production.** Shortly thereafter, children start to try to say words. Since their articulatory capacities at this stage are limited, they use a powerful existing cognitive capacity—phonology—to make at least some output possible. Specifically, they form a kind of personal phonology that maps adult surface forms onto their own, simpler, surface representations. Through the first years of childhood, this personal phonology gradually recedes to vacuity, as children acquire the physical ability to render accurate surface forms.

Faithfulness also starts out low in this production grammar. This low ranking corresponds to the initial state of an infant, namely an inability to say anything at all.

5) **Morphology and Alternation.** At the same time (roughly one to five years), the child comes to be able to factor words into morphemes, to understand the principles of the ambient language’s morphology, to apprehend phonological alternations, and to develop an internalized grammar to predict them (say, in deriving novel forms). Here, the mechanisms used are not at all clear. But there are two plainly useful tools that child brings to the task. First, her relatively full knowledge of phonotactics is surely useful, since so much phonological alternation exists simply to bring concatenated sequences of

morphemes into conformity with phonotactic principles (that is: phonology is conspiratorial). Second, it appears that output-to-output correspondence constraints are given an a priori high ranking. This ranking gives the child a straightforward means of identifying grammatically-conditioned allophones (section 8.4). Once these are identified and suitably attributed to high-ranking output-to-output correspondence constraints, the Faithfulness constraints that were wrongly promoted too high in infancy are allowed to recede back downward toward their preferred low positions. This yields the final, correct phonemic system.

## References

- Albro, Daniel (1997) "Evaluation, implementation, and extension of Primitive Optimality Theory," Master's thesis, Department of Linguistics, UCLA, Los Angeles, CA.  
[<http://www.humnet.ucla.edu/humnet/linguistics/people/grads/albro/papers.html>]
- Algeo, John (1978) "What consonant clusters are possible?" *Word* 29: 206-224.
- Aksu-Koç, Ayhan A. and Dan I. Slobin (1985) "The acquisition of Turkish," in Dan Slobin, ed., *The Crosslinguistic Study of Language Acquisition, Vol. 1: The Data*, Lawrence Erlbaum Associates, Hillsdale, NJ
- Baker, William J. and Bruce L. Derwing (1982) "Response coincidence analysis as evidence for language acquisition strategies," *Applied Psycholinguistics* 3, 193-221.
- Benua, Laura (1997) *Transderivational Identity: Phonological Relations Between Words*, Ph.D. dissertation, University of Massachusetts, Amherst. [Rutgers Optimality Archive 259, <http://ruccs.rutgers.edu/ROA/search.html>]
- Best, C. T., G. W. McRoberts, and N.M. Sithole (1988) "Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants," *Journal of Experimental Psychology: Human Perception and Performance* 14, 345-360.
- Berko, Jean (1958) "The child's learning of English morphology," *Word* 14, 150-177.
- Berman, Ruth A. (1985) "The acquisition of Hebrew," in Dan Slobin, ed., *The Crosslinguistic Study of Language Acquisition, Vol. 1: The Data*, Lawrence Erlbaum Associates.
- Boersma, Paul (to appear): "Learning a grammar in Functional Phonology," to appear in Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, eds., *Optimality Theory: Phonology, syntax, and acquisition*, Oxford University Press, Oxford.
- Boersma, Paul (1998) *Functional Phonology*, Holland Academic Graphics, The Hague.
- Boersma, Paul and Bruce Hayes (in progress) "Empirical tests of the Gradual Learning Algorithm," ms., Department of Linguistics, University of Amsterdam, and Department of Linguistics, UCLA.
- Burzio, Luigi (1996) "Multiple correspondence," ms., Johns Hopkins University, Baltimore, MD.
- Chomsky, Noam (1964) "Current issues in linguistic theory," in Jerry A. Fodor and Jerrold J. Katz, eds., *The Structure of Language*, Prentice-Hall, Englewood Cliffs, NJ
- Chomsky, Noam and Morris Halle (1965) "Some controversial questions in phonological theory," *Journal of Linguistics* 1, 97-138.
- Derwing, Bruce L. and William J. Baker (1986) "Assessing morphological development," in Paul Fletcher and Michael Garman, eds., *Language Acquisition: Studies in First Language Development*, Cambridge University Press, Cambridge.
- Eimas, Peter D., Einar R. Siqueland, Peter Jusczyk, James Vigorito (1971) "Speech perception in infants," *Science* 171, 303-306.
- Eisner, Jason (1997) "Efficient generation in primitive Optimality Theory," in *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, Madrid. [Rutgers Optimality Archive 259, <http://ruccs.rutgers.edu/ROA/search.html>]
- Ellison, Mark (1994) "Phonological derivation in Optimality Theory," in *COLING*, Vol. 2, 1007-1013, Kyoto. [Rutgers Optimality Archive 75, <http://ruccs.rutgers.edu/ROA/search.html>]

- Friederici, Angela D. and Jeanine E. Wessels (1993) "Phonotactic knowledge of word boundaries and its use in infant speech perception," *Perception and Psychophysics* 54, 287-295.
- Fikkert, Paula (1994) *On the Acquisition of Prosodic Structure*, Holland Institute of Generative Linguistics, Leiden and Amsterdam.
- Fodor, Jerry A., Thomas G. Bever, and Merrill F. Garrett (1974) *The Psychology of Language*, McGraw Hill, New York.
- Gnanadesikan, Amalia (1995) "Markedness and Faithfulness constraints in child phonology," Rutgers Optimality Archive 75, <http://rucss.rutgers.edu/ROA/search.html>.
- Guenther, Frank H. and Marin N. Gjaja (1996) "The perceptual magnet effect as an emergent property of neural map formation," *Journal of the Acoustical Society of America* 100, 1111-1121.
- Hale, Mark and Charles Reiss (1998) "Formal and empirical arguments concerning phonological acquisition," *Linguistic Inquiry* 29, 656-683.
- Hayes, Bruce (1986) "Assimilation as spreading in Toba Batak," *Linguistic Inquiry* 17, 467-499.
- Hayes, Bruce (1999) "Phonetically-driven phonology: the role of Optimality Theory and Inductive Grounding," in Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer, and Kathleen Wheatly, eds., *Functionalism and Formalism in Linguistics, Volume I: General Papers*, John Benjamins, Amsterdam, pp. 243-285.
- Hayes, Bruce (to appear) "Anticorrespondence in Yidij," to appear in Ben Hermans and Marc van Oostendorp, eds., *The Derivational Residue*. [<http://www.humnet.ucla.edu/humnet/linguistics/people/hayes/hayes.htm>]
- Jun, Sun-Ah (1996) *The Phonetics and Phonology of Korean Prosody: Intonational Phonology and Prosodic Structure*, Garland, New York.
- Jusczyk, Peter W., Angela D. Friederici, Jeanine M.I. Wessels, Vigdis Y. Svenkerud, and Ann Marie Jusczyk (1993) "Infants' sensitivity to the sound patterns of native language words," *Journal of Memory and Language* 32, 402-420.
- Jusczyk, Peter W., Paul A. Luce, and Jan Charles-Luce (1994) "Infants' sensitivity to phonotactic patterns in the native language," *Journal of Memory and Language* 33, 630-645.
- Kager, René (1996) "Surface opacity of metrical structure in Optimality Theory," to appear in B. Hermans and M. van Oostendorp, eds., *The Derivational Residue in Phonology*, John Benjamins, Amsterdam.
- Kager, René (forthcoming) *Optimality Theory*, Cambridge University Press, Cambridge.
- Kazazis, Kostas (1969) "Possible evidence for (near-)underlying forms in the speech of a child," *Proceedings of the Chicago Linguistic Society* 5, 382-388.
- Keating, Patricia A. (1984) "Phonetic and phonological representation of stop consonant voicing," *Language* 60, 286-319.
- Keating, Patricia, Wendy Linker, and Marie Huffman (1983) "Patterns in allophone distribution for voiced and voiceless stops," *Journal of Phonetics* 11, 277-290.
- Kenstowicz, Michael (1997) "Uniform Exponence: exemplification and extension," Rutgers Optimality Archive 218, <http://rucss.rutgers.edu/ROA/search.html>.
- Kiparsky, Paul (1988) "Phonological change," in Frederick J. Newmeyer, ed., *Linguistics: The Cambridge Survey, Volume 1, Linguistic Theory: Foundations*, pp. 363-415.
- Kiparsky, Paul and Lisa Menn (1977) "On the acquisition of phonology," in J. Macnamara, ed., *Language Learning and Thought*, Academic Press, New York.
- Kirchner, Robert (1997) "Contrastiveness and Faithfulness," *Phonology* 14, 83-113.
- Kisseberth, Charles (1970) "On the functional unity of phonological rules," *Linguistic Inquiry* 291-306.
- Kuhl, Patricia K. (1991) "Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not," *Perception and Psychophysics* 50, 90-107.
- Kuhl, Patricia K. (1995) "Mechanisms of developmental change in speech and language," *Proceedings of the International Congress of Phonetic Sciences, Stockholm*, Vol. 2, 132-139.
- Kuhl, Patricia K. and Denise M. Padden (1982) "Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques," *Perception and Psychophysics* 32, 542-550.
- Kuhl, Patricia K. and Denise M. Padden (1983) "Enhanced discriminability at the phonetic boundaries for the place feature in macaques," *Journal of the Acoustical Society of America* 73, 1003-1010.
- Kuhl, Patricia K., and Joanne D. Miller (1975) "Speech perception by the chinchilla: Voiced-voiceless distinction in alveolar plosive consonants," *Science* 190, 69-72.

- Kuhl, Patricia K., and Joanne D. Miller (1978) "Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli," *Journal of the Acoustical Society of America* 63, 905-917.
- Levelt, Clara C. (1994) *On the Acquisition of Place*, Holland Institute of Generative Linguistics, Leiden and Amsterdam.
- Macken, Marlys (1980) "The child's lexical representation: the *puzzle-puddle-pickle* evidence," *Journal of Linguistics* 16, 1-17.
- McHugh, Brian (1986) "Cyclicality in phrasal phonology: Evidence from Chaga," in Mary Dalrymple, ed., *Proceedings of the Fifth West Coast Conference on Formal Linguistics*, 154-164. Stanford Linguistics Association, Stanford, CA.
- Menn, Lise (1983) "Development of articulatory, phonetics, and phonological capabilities," in Brian Butterworth, ed., *Language Production, Vol. 1*, Academic Press, London, pp. 3-47.
- Pater, Joe (1996) "Minimal violation and phonological development", ms., Department of Linguistics, University of Alberta, Edmonton.
- Pater, Joe (this volume) "From phonological typology to development of receptive and productive phonological competence: applications of minimal violation."
- Prince, Alan and Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Ms., available from Rutgers University Center for Research in Cognitive Science.
- Pulleyblank, Douglas & William J. Turkel (to appear): "Learning phonology: Genetic algorithms and Yoruba tongue root harmony," in *Optimality Theory: Phonology, syntax, and acquisition*, ed. by Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, Oxford University Press, Oxford.
- Smith, Neilson (1973) *The Acquisition of Phonology*, Cambridge University Press, Cambridge.
- Smolensky, Paul (1993) "Harmony, markedness, and phonological activity," Rutgers Optimality Archive 87, <http://ruccs.rutgers.edu/ROA/search.html>.
- Smolensky, Paul (1996) "On the comprehension/production dilemma in child language," *Linguistic Inquiry* 27, 720-31.
- Steriade, Donca (1997) "Phonetics in phonology: the case of laryngeal neutralization," ms., Department of Linguistics, UCLA.
- Steriade, Donca (1996) "Paradigm Uniformity and the phonetics-phonology interface," ms. To appear in *Papers in Laboratory Phonology V*, Cambridge University Press, Cambridge.
- Tesar, Bruce & Paul Smolensky (1993) "The learnability of Optimality Theory: an algorithm and some basic complexity results," Rutgers Optimality Archive 2, <http://ruccs.rutgers.edu/ROA/search.html>.
- Tesar, Bruce & Paul Smolensky (1998): "Learnability in Optimality Theory," Rutgers Optimality Archive 156, <http://ruccs.rutgers.edu/ROA/search.html>.
- Tesar, Bruce & Paul Smolensky (1998): "Learnability in Optimality Theory," *Linguistic Inquiry* 29, 229-268.
- Vance, Timothy (1987) "'Canadian Raising' in some dialects of the Northern United States," *American Speech* 62, 195-210.
- Vihman, Marilyn (1996) *Phonological Development: The Origins of Language in the Child*, Blackwell, Oxford.
- Werker, Janet F. and Richard C. Tees (1984) "Cross-Language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life," *Infant Behavior and Development* 7, 49-63.
- Westbury, John R. and Patricia Keating (1986) "On the naturalness of stop consonant voicing," *Journal of Linguistics* 22, 145-166.
- Zoll, Cheryl (1998) "Positional asymmetries and licensing," Rutgers Optimality Archive 282, <http://ruccs.rutgers.edu/ROA/search.html>.

Department of Linguistics

UCLA

Los Angeles, CA 90025-1543

bhayes@humnet.ucla.edu