Bruce Hayes
May 2001

Department of Linguistics
UCLA

Phonological Acquisition in Optimality Theory: The Early Stages*

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 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. 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 some learning errors that are likely to occur at the earliest stages.

There are two 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.

* For advice that helped to improve this paper I would like to thank Adam Albright, Sun-Ah Jun, René Kager, Patricia Keating, Alan Prince, Charles Reiss, Donca Steriade, Bruce Tesar, Bernard Tranel, Colin Wilson, and audience members at Utrecht, San Diego, Berkeley, and Johns Hopkins. All are absolved from any remaining shortcomings.

Prince and Tesar (1999/this volume) was composed simultaneously with, but independently from, the present paper. The authors present a proposal about phonotactic learning that is quite similar to what is proposed here. In clarifying and focusing the original version of this paper (Hayes 1999c) for the present volume I have benefited considerably from reading Prince and Tesar's work. For a comparison of the ranking algorithms proposed in the two papers, see Appendix A.

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 these experimental and theoretical research lines can be brought closer 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 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 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,[1] 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 literature review, see Gerken (1994, 792-799), and for some recent contributions Levelt (1994), Fikkert (1994), Pater (1996), Bernhardt and Stemberger (1998), Boersma (2000), and various papers in this volume.[2]

[1] With important exceptions; see for instance 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.

[2] Hale and Reiss (1998) likewise propose that the child's output mapping is 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 acquisition from another point of view, focusing on 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. The focus of this paper is the child's conception of the adult language, a research topic which can perhaps be fairly described as neglected by phonologists.

To clarify what is meant here, 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.

I advocate, then, a clear separation between the child's phonological analysis of the ambient language vs. her personal production phonology. This view can be opposed, for example, to that of Smolensky (1996a), who takes the (a priori, rather appealing) view that the child's grammars for production and perception are the same. I will argue that that this cannot be right: children whose production rankings generate very primitive outputs—or none at all—nevertheless can pass the "blick" test. They couldn't do this unless they had also internalized an adult-like constraint ranking, separate from their production grammar.[3]

3. Some Results from the Acquisition Literature

To start, I will present a quick summary of results from the experimental literature on the early stages of phonological acquisition; see Gerken (1994) for more detail on most of the material treated here. 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 a high 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

---

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] Separating the child's internalized conception of the adult grammar from her production grammar also helps clarify various patterns in the child phonology literature. For instance, before age 4;3 Gwendolyn Stemberger produced /niːd+d/ as [niːdəd] 'needed', but /hʌg+d/ as [hʌdd] 'hugged' (Bernhardt and Stemberger 1998, 651). This mapping makes sense if /niːd+d/ → [niːdəd] was part of Gwendolyn's conception of adult phonology, but /hʌg+d/ → [hʌdd] was part of her production mapping from adult outputs to Gwendolyn-outputs. A similar example appears in Dinnsen et al. (in press).

about what interests them in their surroundings when they vary the rate of sucking on an electronically monitored pacifier. Older babies can turn their heads in the direction they choose, and the duration of their head turns can be used to establish their degree of interest in linguistic material presented to them—most crucially, *relative* differences of interest in different linguistic material. Methods have been developed to ensure that the observations (e.g. "How long did this head turn last?") are unbiased 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.

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 et al. 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 phonetic 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 phonetic 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 a number of different ways.

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. 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. Thus, for example, some distributional protocategories may turn out to be only strongly differentiated allophones of the same phoneme. These are only later united into a single category during the next phase of learning, when the child discovers that the protocategories have a predictable phonological distribution. The means by which this might be done are explored below.

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.

I. Studies by Werker and Tees (1984) and Werker and Lalonde (1988) 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 can distinguish words from one another. This effect has been demonstrated by Werker and her colleagues for retroflex/alveolar contrasts in Hindi and for uvular/velar contrasts in Nthlakampx.[4]

II. At more or less the same time, infants start to acquire knowledge of the legal segments and sequences of their language (cf. [blɪk] vs. *[bnɪk], above). This is shown in work by Jusczyk, Friederici, Wessels, Svenkerud, and Jusczyk (1993), Friederici and Wessels (1993), and Jusczyk, Luce, and Charles-Luce (1994). In carefully monitored experimental situations, eight- to ten-month-old infants come to react differently to legal phoneme sequences in their native language than to illegal or near-illegal ones.[5]

Both phenomena suggests that the ages of eight to ten months are the birth of true phonology; the infant at this stage takes the distributional protocategories obtained in earlier infancy and processes them to form a first-pass phonological system. It is worth pondering, I

[4] Best et al. (1988) have shown that English-learning infants 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.

[5] Examples from Jusczyk, Friederici, et al. (1993): Dutch *[rtum], English ?[jiːdʒ]. 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.

think, what might be done to characterize in a formal phonological theory what a ten-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 adult speakers possess. 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.

There are basically three kinds of phonological knowledge. For each, I will review how such knowledge is currently described formally in Optimality Theory (Prince and Smolensky 1993), the approach to phonology adopted here. [6]

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 [**b**u] 'end' vs. [**p**u] '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 contrasts and their distribution 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 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, with Markedness over Faithfulness; 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.[7]

In some cases, the situation is more complex than what was just described: the ranking of constraints is such that a contrast is allowed only in particular contexts. Thus, French generally allows for a voicing distinction in stops, but there is a high-ranking Markedness constraint that requires voicing agreement in obstruent clusters. This constraint outranks Faithfulness for stop voicing, so that the contrast is suspended in certain contexts. For instance, there is no voicing contrast after an initial [s]; there are pairs like [bu] vs. [pu], but no pairs like [s**p**eˈsjal] ('spéciale') vs. *[s**b**eˈsjal].

---

[6] 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 (1999b). Another helpful account, oriented to acquisition issues, is included in Bernhardt and Stemberger (1998).

[7] 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.

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 other restrictions on 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. This point is made by Smolensky (1996b).

1.2 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 suprasegmental phenomena such as syllables, stress, and tone. The case of legal [blɪk] vs. illegal *[bnɪk] noted above is an example. To designate this sort of knowledge, I will use the somewhat archaic term PHONOTACTICS: a speaker who knows the phonotactics of a language knows its legal sequences and structures.

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 mechanisms other than constraint ranking in defining the phonotactics.

1.3 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 *cans* [kæn**z**]; but it is realized as [s] when it follows a voiceless consonant: *caps* [kæp**s**].

The [s] realization is related to the phonotactics in an important way: English does not tolerate final sequences like [pz], in which a voiced obstruent follows a voiceless one. For example, there are monomorphemic words like *lapse* [læ**ps**], but no words like *[læ**pz**].

Optimality Theory treats most alternations as the selection of an output candidate that deviates from the underlying form in order to conform to a phonotactic pattern. In this way, it establishes an especially close relationship between phonotactics and alternation. Thus, for underlying /kæp+z/, the winning candidate is [kæps], in which the underlying value of [voice] for /z/ is altered in order to obey the Markedness constraint that forbids final heterovoiced obstruent clusters. We will return to the connection between phonotactics and alternation below.

1.4 Interpreting the Acquisition Literature

Turning now to the acquisition results reviewed earlier, I adopt the following interpretations of them within Optimality Theory.

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 Optimality Theory establishes what is phonemic.

Phonotactics: the work of Jusczyk and others suggests that by the time infants are eight to ten months hold, 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 infer 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. [wut] and [wiːt] 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* [dɔg**z**] and *cats* [kæt**s**].

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.

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—contrasts, phonotactics, and alternation. For the first two, it appears that acquisition is precocious, with much progress made by the age of 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 phonotactics 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 get started when the child merely possesses a more-or-less random inventory of words and short phrases.[8] 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. It was noticed by Kisseberth (1970) and subsequent work that this alternation-driven approach characteristically missed crucial generalizations about phonologies, generalizations that were 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

[8] Evidence for rather impressive ability among infants to extract and remember words and phrases from the speech stream is presented in Jusczyk and Aslin (1995), Jusczyk and Hohne (1997), and Gomez and Gerken (1999).

conspiracy problem disappears, since the rules that formerly "conspired" are absent, and the target of the conspiracy is itself the core of the analysis.[9]

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 before knowledge of alternations.

More important, I believe that prior knowledge of phonotactics would actually *facilitate* the acquisition of alternation. The reason is that most alternation is directly driven by the need 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* [kæn**z**]) but [əz] after sibilants (*edges* [ɛdʒ**əz**]) and [s] after voiceless sounds other than sibilants: *caps* [kæp**s**]. 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 *[ɛdʒ**z**]), nor can they end in a sequence of the type *voiceless obstruent* + *voiced obstruent* (*[kæ**pz**]). 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].

We can put this slightly more generally if we adopt some terminology from the older rule-based theory. A rule has a STRUCTURAL DESCRIPTION (the configuration that must be present for a rule to apply) and a STRUCTURAL CHANGE (the change the rule carries out). In these terms, a child who has already achieved a good notion of the phonotactics of her language need not, in general, locate 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. More precisely, within Optimality Theory, the learner must locate 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 *[wʌ**gs**] or *[gʌ**tʃs**] (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 children's' greater of

[9] The production phonologies of 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 production phonologies.

reliability in this area results from their having 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 the core issue of this paper: 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. Research in this tradition began with Tesar and Smolensky (1993) and includes later work such as Tesar (1995a, 1995b, 1999), Tesar and Smolensky (1996, 1998, 2000), Boersma (1997), Pulleyblank and Turkel (2000), and Boersma and Hayes (2001).

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:

(1) 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, this definition stipulates 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 seems to be a realistic requirement in the present case. [10]

The rankings obtained by a pure phonotactic learner 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

---

[10] Phonological alternations provide a weak form of negative evidence: the fact that the [-z] suffix of *cans* [kæn**z**] shows up altered to [-s] in *caps* [kæp**s**] is a clue that final *[pz] is not legal in English. It is for this reason that constraint ranking is often an easier problem for alternation data than for purely phonotactic data. Given that alternations can provide negative evidence, it is relevant that the learning of alternations appears to happen relatively late (section 5); and also that in many languages only a fraction of the phonotactic principles are supported by evidence from alternations.

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).[11]

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, 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 assumption is likewise not needed in pure phonotactic learning: our (somewhat idealized) view is that we are dealing with a stage at which the child has identified individual words but not yet parsed them into morphemes. In such a system, there are no alternations, so there is no need for underlying forms to account for them.[12]

With both functions of underlying forms dispensed with in the present context, we can suppose that underlying representations are the same as surface representations; [13] 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.

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, 2000). When applied to conventional data sets (involving alternation), Constraint Demotion arrives quite efficiently 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

---

[11] The 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 is undertaken.

[12] 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 one-word utterances, where the effects of phrasal phonology would be at a minimum. Another possibility is for the child to internalize a supply of short phrases, and learn their phonotactics without necessarily parsing them.

[13] I am grateful to Daniel Albro for suggesting this as a basis for pure phonotactic learning.

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).[14]

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; (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 the correct output for each underlying form.

A term from Prince and Tesar (1999/this volume) that is useful in understanding Constraint Demotion is PREFERS A LOSER: a constraint prefers a loser if an ill-formed rival candidate violates it fewer times than the correct form does. The leading idea of Constraint Demotion is to demote those constraints that prefer losers to a 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 batch version of Constraint Demotion is summarized as follows:

(2) **Constraint Demotion**

I. Find all constraints that don't prefer any losers. Place them in a "stratum," a set of constraints assumed to occur together at the top of the ranking hierarchy.

II. 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. 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.

III. Of the constraints that have not yet been placed in a stratum, find those which prefer no losers in the remaining data. Place them in the next stratum of constraints.

IV. Cull out explained rivals again, as in II.

V. Repeat steps III and IV 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.

---

[14] In work not described here, I have examined the more sophisticated Error Driven Constraint Demotion variant of the algorithm, obtaining essentially the same results as are described below for the batch version.

Sometimes, step III of the algorithm yields no constraints at all; all the remaining constraints prefer losers. In such cases, it turns out, there *is no ranking* of the constraints that will generate only winners. Thus, Constraint Demotion has the ability to detect failed constraint sets.[15] A constraint set also fails if it is insufficiently rich, assigning identical violations to a winner and rival.

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, as I will demonstrate.

The next few sections of the paper are laid out as follows. In sections 7.2 and 7.3, I present a simple data example, “Pseudo-Korean,” to be used as an illustration for the ranking algorithms. Section 7.4 applies Constraint Demotion to Pseudo-Korean, and shows how it is unable to learn its phonotactics. Sections 7.5-7.7 lay out my own algorithm, and 7.8 shows how it learns the Pseudo-Korean pattern.

7.2 “Pseudo-Korean”: Basic Pattern and Constraints

Imagine a language in which stops contrast for aspiration; thus /ptk/ and /pʰtʰkʰ/ form separate phonemic series and are attested in minimal pairs, such as [tal] ‘moon’ vs. [tʰal] ‘mask’. Assume further that, while /pʰtʰkʰ/ show no significant allophonic variation, /ptk/ are voiced to [bdg] when intervocalic: thus [**k**e] ‘dog’ but [i **g**e] ‘this dog’. 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 [tʃi**pʰ**i] ‘straw-nom.’ and [tʃi**b**i] ‘house-nom.’ show the phonemic contrast between /pʰ/ and the [b]-allophone of /p/, this contrast is neutralized to plain [p] in final position, so that unsuffixed [tʃi**p**] 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 (Lombardi 1995, 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 and rankings for analyzing Pseudo-Korean in Optimality Theory is given below.

7.2.1 Markedness Constraints

(3) *[–SONORANT, +VOICE]

---

[15] Or, in some cases, to correct assumptions made earlier about “hidden structure” in the input data; Tesar and Smolensky (2000).

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.

(4) *[+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 these two constraints in hand, we may consider their role in Pseudo-Korean. As will be seen, the Faithfulness constraints for voicing are ranked so low so as to make no difference in Pseudo-Korean; therefore voicing is allophonic. The distribution of voicing is thus determined by the ranking of the markedness constraints. In particular, *[+V][–V][+V] must dominate *[–SON, +VOICE], so that obstruents will be voiced in voiced surroundings:

(5)

/ada/	*[+V][–V][+V]	*[–SON, +VOICE]
☞ [ada]		*
*[ata]	*!	

Under the opposite ranking, obstruents will be voiceless everywhere; Keating et al. (1983) note that this is the pattern found in Hawaiian and various other languages.

(6) *[+SPREAD GLOTTIS] (abbr. *ASPIRATION or *ASP)

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

(7) *[+VOICE, +SPREAD GLOTTIS] (abbr. $*D^h$)

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

In pseudo-Korean, $*D^h$ must be ranked above *[+V][–V][+V]; otherwise, aspirated stops would be voiced intervocalically.

(8)

$/at^ha/$	$*D^h$	*[+V][–V][+V]
☞ $[at^ha]$		*
$*[ad^ha]$	*!	

Hence we have the three-way ranking $*D^h$ >> *[+V][–V][+V] >> *[–SON, +VOICE].

7.2.2 Faithfulness Constraints

The constraint in (9):

(9) IDENT([SPREAD GLOTTIS]) / ___ V (Abbr.: IDENT(ASP) / ___ V)

is a "positional Faithfulness" constraint, of the type explored by Beckman (1998) and others. It is violated when a segment occurring prevocalically in a candidate surface form corresponds to an underlying segment with which it disagrees in the feature [spread glottis]; i.e., aspiration. The constraint is "positional" because it relies on a phonological context. [16]

The rationale for the particular context invoked, prevocalic, has been explicated by Steriade (1997), who offers an explanation for the fact that aspiration and other laryngeal contrasts gravitate cross-linguistically to prevocalic position. In Steriade's view, this has an acoustic explanation: vowels[17] 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.

I also assume a general, context-free Faithfulness constraint for aspiration:

(10) IDENT([SPREAD GLOTTIS]) (Abbr.: IDENT(ASP))

The type of aspiration pattern a language will allow depends on the ranking of *ASPIRATION in the hierarchy: if *ASPIRATION is on top, then aspiration will be missing entirely (as in French); if *ASPIRATION is outranked by IDENT(ASP) (and perhaps, also redundantly by IDENT(ASP) / ___ V), aspiration will be contrastive in all positions, as in Hindi. Pseudo-Korean reflects the ranking IDENT(ASP) / ___ V >> *ASP >> IDENT(ASP), which permits aspiration prevocalically ((11)) but not in other positions ((12), where an illegal input */at^h/ loses out to a legal winner [at]).

(11)

/t^{h}a/	ID (ASP)/_V	*ASP
☞ [t^{h}a]		*
*[ta]	*!	

(12)

*/at^h/	*ASP	ID (ASP)
☞ [at]		*
*[at^h]	*!	

[16] 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 here places contexts primarily in the Faithfulness constraints; I have also tried a parallel simulation using the opposite strategy, and obtained very similar results.

[17] More accurately: sonorants, but for Pseudo-Korean I will stick with vowels for brevity.

A more subtle ranking argument concerns the possibility of *[ada] as the winning candidate for underlying /at^{h}a/. *[ada] satisfies both *D^h and *[+V][–V][+V], at the expense of IDENT(VOICE) / ___ V. In order for [at^{h}a] to win out over [ada], IDENT(ASP) / ___ V must be ranked above *[+V][–V][+V].

(13)

/at^{h}a/	*D^h	ID (ASP) /_V	*ASP
☞ [at^{h}a]			*
*[ada]		*!	
*[ad^{h}a]	*!		*

The following Faithfulness constraints for voicing:

(14) IDENT(VOICE) / ___ V

(15) IDENT(VOICE)

work just like the analogous constraints for aspiration. However, in Pseudo-Korean, where voicing is allophonic, both are at the bottom of the grammar, where they have no influence. The ranking *[+V][–V][+V] >> *[-SON, +VOICE] >> { IDENT(VOICE) / ___ V, IDENT(VOICE) } forces a distribution where unaspirated stops are voiced in intervoiced position, and voiceless elsewhere. The following tableaux illustrate this; they show how underlying forms containing stops with the wrong voicing would lose out to well-formed alternatives:

(16)

/da/	*[+V][–V][+V]	*[–SON, +VOICE]	ID(VOICE)/_V	ID(VOICE)
☞ [ta]			*	*
*[da]		*!		

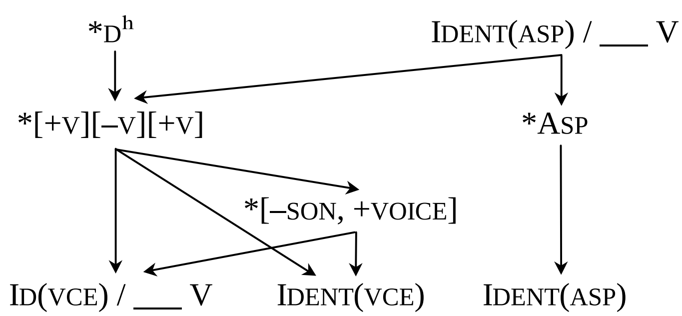
(17)

/ata/	*[+V][–V][+V]	*[–SON, +VOICE]	ID(VOICE)/_V	ID(VOICE)
☞ [ada]		*	*	*
*[ata]	*!			

7.2.3 Target Ranking

The following diagram summarizes all rankings required to derive the Pseudo-Korean pattern.

(18)



An adequate ranking algorithm must learn *at least* these rankings. It may also without harm posit additional ranking relationships, so long as they do not conflict with those of (18).

7.3 Pseudo-Korean: Candidates

To provide a reasonable test for ranking algorithms, I developed a large set of Pseudo-Korean forms, with rival candidates for each. The rivals included all possible combinations of voicing and aspiration in all positions. A representative subset is given in (19):

(19)

Input	**Winning Output**	**Rivals**
/ta/	[ta]	*[t^ha], *[da], *[d^ha]
/t^ha/	[t^ha]	*[ta], *[da], *[d^ha]
/ada/	[ada]	*[ata], *[at^ha], *[ad^ha]
/at^ha/	[at^ha]	*[ata], *[ada], *[ad^ha]
/at/	[at]	*[ad], *[at^h], *[ad^h]
/tada/	[tada]	*[tata], *[tat^ha], *[tad^ha], *[dada], *[t^hada], *[d^hada]
/tat^ha/	[tat^ha]	*[tata], *[tada], *[tad^ha], *[dat^ha], *[d^hat^ha], *[t^hat^ha]
/t^hada/	[t^hada]	*[t^hat^ha], *[t^hata], *[t^had^ha], *[tada], *[dada], *[d^hada]
/t^hat^ha/	[t^hat^ha]	*[t^hata], *[t^hada], *[t^had^ha], *[tat^ha], *[dat^ha], *[d^hat^ha], *[tata], *[tada]
/tat/	[tat]	*[tat^h], *[tad], *[tad^h], *[t^hat], *[dat], *[d^hat]
/t^hat/	[t^hat]	*[t^had], *[t^hat^h], *[t^had^h], *[tat], *[dat], *[d^hat]

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; thus, the training set provided only positive evidence.

7.4 Application of Constraint Demotion to Pseudo-Korean

I will first show how, and why, Tesar and Smolensky's Constraint Demotion algorithm is not suited to pure phonotactic learning. If we apply the Constraint Demotion to the Pseudo-Korean data, it will install in the top stratum the following five constraints, which never prefer losers:

(20) **Stratum #1**

IDENT(ASPIRATION)
IDENT(VOICE)
IDENT(ASPIRATION) / ___ V
IDENT(VOICE) / ___ V
*D^h

*D^h reflects a patently true regularity of Pseudo-Korean, and the four Faithfulness constraints are never violated, as a result of the positive-data-only learning situation.

With all these constraints in place, all of the learning data are explained, and the remaining, loser-preferring constraints end up in the second stratum:

(21) **Stratum #2**

*[+V][–V][+V]
*[–SON/+VOICE]
*ASPIRATION

The result is 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. To illustrate this, we can feed to this grammar a set of underlying forms that are illegal in Pseudo-Korean, and check to see what emerges as the winner. Omitting the tableaux, I will simply list here some representative cases that emerge from this procedure:

(22)

Input	Output	Input	Output
/da/	*[da]	/ad/	*[ad]
/d^{h}a/	*[d^{h}a] or *[t^{h}a] or *[da]	/at^h/	*[at^h]
/ata/	*[ata]	/ad^h/	*[ad^h] or [at^h] or *[ad]
/ad^{h}a/	*[ad^{h}a] or *[at^{h}a] or [ada]		

The crucial point is that a large number of illegal forms is generated. It also can be noted in passing that there is also a good deal of free variation: it matters how *D^h is ranked with respect to the Faithfulness constraints; but Constraint Demotion is unable 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 applied to learning data of the type considered here.

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

We begin by invoking a fundamental idea of recent theoretical acquisition work in Optimality Theory (Gnanadesikan (1995/this volume), Smolensky (1996b), and much other work): Faithfulness constraints should be assigned a default location at the bottom of the constraint hierarchy. From the viewpoint of this article, this proposal is actually two proposals.

First, in the production grammars that children use to generate 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.

Second, and more relevant here, default low-Faithfulness ranking can be applied to the learner's 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." As Smolensky (1996b) showed, this has a direct formal implementation in Optimality Theory: 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* [blɪk] 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 An Approach That Won't Work: Initial Rankings

The simplest approach, which emerges directly from the discussion above, is to let Constraint Demotion start not with a "blank slate," with all constraints considered equally, but rather with an *a priori* initial ranking in which all Markedness constraints outrank all Faithfulness constraints. In this approach, there would be a Stratum 1, containing all of the Markedness constraints, and a Stratum 2, containing all of the Faithfulness constraints.

It appears that this approach is unworkable, as we can see by applying it to Pseudo-Korean. When first encounters the data, Constraint Demotion will find that only *Dʰ prefers no losers and may thus remain in Stratum 1. The remaining three Markedness constraints of Pseudo-Korean prefer losers:

(23) a. *[+V][–V][+V] prefers loser *[ada] to winner [atʰa].
b. *[–SON/+VOICE] prefers loser *[ata] to winner [ada].
c. *ASPIRATION prefers loser *[ta] to winner [tʰa].

Thus, they must be demoted to a lower stratum: either Stratum 2, or a newly formed Stratum 3, depending on the details of the algorithm. If the demoted constraints move only to Stratum 2, then on the next round they will *still* prefer losers, and thus will move on down to Stratum 3. At the same time, all the Faithfulness constraints remain in Stratum 2, because they never prefer a loser. The result is again a grammar with all Faithfulness constraints ranked high, resulting in massive overgeneration.

The upshot is a point made by Prince and Tesar (1999/this volume): a ranking algorithm is unlikely to succeed simply by using a starting point that favors Markedness over Faithfulness; rather, the bias must be enforced throughout the ranking process.

7.7 A New Algorithm: Low Faithfulness Constraint Demotion[18]

As a tentative solution to the problem of pure phonotactic learning, I propose a Low Faithfulness Constraint Demotion algorithm. This algorithm is identical to the batch version of classical Constraint Demotion, with the following crucial exception. Whenever a new stratum is to be created, the criteria that a constraint must pass to be eligible to be installed in the new stratum are made more stringent in various ways. Among other things, this implements the consistent bias for Markedness just described.

In Low Faithfulness Constraint Demotion, the principles for ranking are prioritized; higher-priority principles get first say in which constraints are or are not eligible for installation in the new stratum. The exclusion of constraints from the next stratum is formally reminiscent of the exclusion of candidates in OT derivations. The full set of principles, in priority order, are given below; ranking principles will be designated typographically with boldface italics.

7.7.1 ***Avoid Preference For Losers***

This principle is taken from classical Constraint Demotion. It forbids the admission to the current stratum of constraints that prefer a losing to a winning candidate. It plainly must be an “undominated” ranking principle, since if one admits such a constraint to the current stratum, the finished grammar will itself prefer losers and generate incorrect outputs.

7.7.2 ***Favor Markedness***

Suppose that, after we have culled out the constraints that prefer losers, we are left with both Faithfulness and Markedness constraints. In such cases, Low Faithfulness Constraint Demotion installs only the Markedness constraints in the new stratum. The Faithfulness constraints must await a later opportunity to be ranked, often the next stratum down. Only when the eligible set consists entirely of Faithfulness constraints may Faithfulness constraints be considered for ranking.

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

---

[18] The algorithm given here is the same as the one in the earlier Web-posted version of this article (Hayes 1999c). However, the presentation has been substantially reshaped, taking a lesson from the presentation of similar material in Prince and Tesar (1999/this volume). In addition to increasing clarity, the changes should also facilitate comparison of the two algorithms.

constraint to do the job, because if we let Faithfulness do it, it is likely to lead to overgeneration in the finished grammar.

7.7.3 *Favor Activeness*

We consider next the principles that must be invoked when the pool of rankable constraints contains only Faithfulness constraints, so that ***Favor Markedness*** is unable to decide.

Consider that it is quite common for a constraint not to prefer any losers—but not to prefer any winners either. Such constraints are, in the terminology of Prince and Tesar (1999/this volume), INACTIVE. It plainly does no good to rank inactive constraints high, and may well do harm in terms of overgeneration for data yet unencountered. Therefore, Low Faithfulness Constraint Demotion excludes such constraints from the set allowed to join the next stratum.

An exception is necessary: if the set of eligible candidates contains only inactive Faithfulness constraints, they must all be assigned to the next stratum anyway, to permit the algorithm to terminate. This step usually turns out to be the last stage of ranking: it dumps the inactive Faithfulness constraints into the bottom stratum, where they will do no harm in overgenerating.[19]

7.7.4 *Favor Specificity*

Often, two Faithfulness constraints fall into a specific-general relation; for example, IDENT (ASP) / ___ V is a more specific version of IDENT (ASP). Suppose we have two such constraints, both found in the pool of constraints still eligible to be ranked after we have invoked ***Avoid Preference For Losers***, ***Favor Markedness***, and ***Favor Activeness***. In such cases, the principle ***Favor Specificity*** requires that the general constraint must be excluded from the next stratum; only the specific constraint remains eligible.

The rationale for ***Favor Specificity*** is again conservatism: there is no point in using a general constraint to rule out a rival candidate if a specific one will suffice. Ranking the specific constraint alone often permits the possibility of ranking its general partner deep in the grammar, below Markedness constraints not yet been ranked, so overgeneration is avoided. In contrast, if one admits the general constraint to the current stratum right away, one gives up on this possibility from the start.

In most cases, it is straightforward to assess specificity: a constraint that properly includes another in its structural description (Koutsoudas, Sanders, and Noll 1974) is the more specific of the two and will normally have a subset of the other's violations (it will never have a superset, which is most crucial). Thus the additional material "/ ___ V" in IDENT(ASP) / ___ V identifies it as a more specific version of IDENT(ASP). For further discussion of specificity, see Appendix A.

---

[19] When the constraint set is inadequate (cannot derive the winners under any ranking), the dumping of the inactive Faithfulness constraints into a default stratum will be the penultimate phase. In the final phase, the algorithm learns that it has only loser-preferring Markedness constraints to work with, and is unable to form any further strata. It thus terminates without yielding a working grammar (cf. 7.1, 7.7.6).

7.7.5 Favor Autonomy

If the algorithm has gotten this far, it has culled the set of rankable constraints down to a set of active Faithfulness constraints, none of which is a more general version of some more specific constraint. If this set consists of just one member, it can be selected as the sole constraint of the current stratum, and we can move on. But what if more than one Faithfulness constraint is still eligible? It would certainly be rash to install all of the eligible constraints, because it is quite possible that we can ward off overgeneration by installing just some of them.

My proposal is that there should a criterion of AUTONOMY. As an approximation, we can say that for an eligible Faithfulness constraint F_i to be installed in the current stratum, it should exclude at least one rival R autonomously, acting without help from any other constraint C. By "help," I mean that C, working alone, would likewise exclude R (in other words, R violates C more than the winner). This is the core idea, and I will try to justify it below.

Before doing so, I must add some details. First, we must consider the possibility that there might not be any constraints that exclude a rival without help. In such cases, I propose that we should loosen the criterion, seeking constraints that exclude some rival with the help of just one other constraint, or (failing that) just two, and so on. In other words, autonomy is relative, not absolute. Eventually, the criterion of relative autonomy will select a constraint or set of constraints that can be installed in the new stratum.

The other detail concerns what constraints may be considered as eligible "helpers" in assessing autonomy, filling the role of C. We will see below that Markedness constraints must be considered, even if at the stage in question they prefer losers.[20] I also conjecture, for concreteness' sake, that Faithfulness constraints excluded by ***Favor Specificity*** should not be included, though I do not yet know of any evidence to bear on this question.

With these modifications in place, I state ***Favor Autonomy*** formally in (24c), following two preliminary definitions:

(24) a. Defn.: *helper*

Let F be an yet-unranked Faithfulness constraint, and R be some rival candidate such that F prefers the winner to R. Let C be a yet-unranked constraint not excluded by ***Favor Specificity***. If C also prefers the winner to R, then C is a *helper* of F.

b. Defn.: *minimum number of helpers*

Let F be an yet-unranked Faithfulness constraint, and { R_1, R_2, ... R_n } be the set of all rival candidates R_i such that F prefers the winner to R_i. Let { h_1, h_2, ... h_n } be the number

[20] Markedness constraints already assigned to a stratum can never be helpers in any event: the rivals they explain have been culled from the learning set.

of helpers of F for each of { R_1, R_2, ... R_n } respectively. The minimum number of helpers of F is the lowest value in { h_1, h_2, ... h_n }.

c. ***Favor Autonomy***

Of the constraints that satisfy all higher-priority ranking principles (see 7.7.1-7.7.4), install on the current stratum the constraint or constraints whose minimum number of helpers is the lowest.

I now turn to why ***Favor Autonomy*** is a plausible ranking principle. In brief, what is does is *force the algorithm to use data that are maximally informative*. When we are dealing with a rival candidate in which the number of helpers for each constraint is high, it means that there are many possible explanations for why the rival is excluded. Such cases tell us little. In contrast, when a Faithfulness constraint excludes a rival on its own, it is usually a very simple case, with implications that are clear. Intermediate cases, with few but more than zero helpers, have an intermediate evidentiary value, and for this reason Low Faithfulness Constraint Demotion is set up gradiently, to go with the best cases available.

As an illustration, let us compare two cases. When Low Faithfulness Constraint Demotion is run on Pseudo-Korean (full details below in 7.8), it reaches a stage where it must select among the pair { IDENT (ASP) / ___V, IDENT (VOICE) /___V } for which should join the new stratum. If we leave out constraints already ranked and constraints excluded by ***Favor Specificity***, the constraint violations for the winner-rival pair [ada] ~ *[ata] look like this:

(25)

/ada/	ID (ASP) /_V	*[+V][–V][+V]	*ASP	*[–SON/+VCE]	ID (VCE) /_V
☞ [ada]				*	
*[ata]		*			*

From this pattern, it is evident that if we let ID (VCE) / ___V join the next stratum, *[ata] will be ruled out. But ID (VCE) / ___V has a helper, namely *[+V][–V][+V], so perhaps it is really *[+V][–V][+V] (yet unrankable) that is ultimately responsible for ruling out *[ata]. We cannot know at this stage. As it turns out in the end, it would be rash to pick IDENT (VOICE) / ___V, because in the correct grammar ((18) above), it emerges that *[+V][–V][+V] must outrank IDENT (VOICE) / ___V to avoid overgeneration.

By way of comparison, (26) shows the violation pattern for winner [t^ha] vs. rival *[ta], at the same stage of learning:

(26)

/t^ha/	ID (ASP)/_V	*[+V][–V][+V]	*ASP	*[–SON/+VCE]	ID (VCE)/_V
☞ [t^ha]			*		
*[ta]	*				

Here, IDENT (ASP) / ___V rules out *[ta] alone, with no help from any other eligible constraint. Its autonomy constitutes plain evidence that IDENT (ASP) / ___V needs to be ranked high. ***Favor Autonomy*** thus selects it for the stratum under construction, which, as it turns out, permits the rest of the grammar to be constructed straightforwardly (see 7.8).

Summing up: ***Favor Autonomy***, by seeking the examples where a constraint acts with the fewest possible helpers, singles out the cases where a constraint is most clearly shown to require a high ranking. By only allowing maximally autonomous Faithfulness constraints to be installed, we can limit installation to cases that are most likely to be truly necessary, letting the less autonomous constraints sink further down to a point where they will not lead to overgeneration.

Favor Autonomy, being relatively defined, will always place at least one constraint in the current stratum. Thus, when invoked, it is the last criterion used in stratum formation. Once the stratum is created, Low Faithfulness Constraint Demotion continues just like classical Constraint Demotion: rival forms explained by the constraints of the new stratum are culled from the learning set, and the next stratum down is created by invoking the criterion hierarchy again, beginning with ***Avoid Preference For Losers***.

7.7.6 Terminating the Algorithm

Low Faithfulness Constraint Demotion terminates under conditions identical to those of classical Constraint Demotion. Specifically, when the constraint inventory is adequate (can be ranked to prefer all and only winners), both algorithms terminate in success when all constraints have been assigned to a stratum and all rival candidates have been ruled out. Both terminate in failure when ***Avoid Preference For Losers*** is unable to locate constraints that favor no losers.

7.8 Pseudo-Korean and Low Faithfulness Constraint Demotion

To illustrate how Low Faithfulness Constraint Demotion works, I will now apply it to the Pseudo-Korean problem, for which the candidate set was given in (19). For convenience, the full constraint inventory is repeated below:

(27)

a. **Markedness**	b. **Faithfulness**
*D^h	IDENT(ASPIRATION) / ___ V
*[+V][–V][+V]	IDENT(ASPIRATION)
*[–SON/+VOICE]	IDENT(VOICE) / ___ V
*ASPIRATION	IDENT(VOICE)

The strata of the grammar are formed as follows:

Stratum 1: As it would in classical Constraint Demotion, ***Avoid Preference For Losers*** excludes the markedness constraints *[+V][–V][+V], *[–SON/+VOICE], and *ASPIRATION. The losers that these constraints prefer were given in (23).

The remaining five constraints are *D^h plus the four Faithfulness constraints. ***Favor Markedness*** selects *D^h, which forms the top stratum alone. All rival candidates containing [d^h] are pruned from the learning set.

Stratum 2: We first evaluate ***Avoid Preference For Losers***. *[+V][–V][+V], *[–SON/+VOICE], and *ASPIRATION continue to prefer losers and still cannot be ranked. ***Favor Markedness*** cannot be satisfied, so we must consider the Faithfulness constraints. All of them pass ***Favor Activeness***, as shown in (28):

(28)	IDENT(ASP) / ___ V	prefers winner	[t^ha]	over rival	*[ta]
	IDENT(ASP)	prefers winner	[t^ha]	over rival	*[ta]
	IDENT(VOICE) / ___ V	prefers winner	[ada]	over rival	*[ata]
	IDENT(VOICE)	prefers winner	[ada]	over rival	*[ata]

Control thus passes to ***Favor Specificity***, which excludes IDENT(ASPIRATION) and IDENT(VOICE).

With two constraints left in the running (IDENT(ASP) / ___ V and IDENT(VCE) / ___ V), control passes to ***Favor Autonomy***. The assessment of autonomy for these constraints was already illustrated in (25) and (26), which show the fewest-helper cases for each. Since IDENT(ASP) / ___ V is the more autonomous (zero helpers to one), it is selected as the sole member of Stratum 2. The rival candidates that it explains, such as *[ta] for /t^ha/, are excluded from the learning set.

Stratum 3: We first evaluate ***Avoid Preference For Losers***: *[–SON/+VOICE] still prefers *[ata] for underlying /ada/, and so must remain unranked. But two other Markedness constraints now prefer no losers among the remaining data. One such constraint is *[+V][–V][+V]. Earlier it preferred *[ada] for /at^ha/, but that form is now explained by IDENT(ASP) / ___ V. Therefore, ***Avoid Preference For Losers*** permits *[+V][–V][+V] to join the new stratum. Likewise, *ASP no longer prefers losers (*[ta] for /t^ha/ is now ruled out by IDENT(ASP) / ___ V), so it too may join the third stratum. The remaining constraints are Faithfulness constraints, and they are shut out by ***Favor Markedness***.

Once *[+V][–V][+V] is placed in Stratum 3, then *[ata] for /ada/ is explained and excluded from the learning set. This vindicates the earlier decision ((25)) not to place ID(VCE) / ___ V in Stratum 2.

Stratum 4: ***Avoid Preference For Losers*** : *[–SON/+VOICE] no longer prefers a loser, since *[ata] for [ada] is now ruled out by *[+V][–V][+V]. Thus *[–SON/+VOICE] may now be ranked in the fourth stratum. Since *[-SON/+VOICE] is a Markedness constraint, ***Favor Markedness*** continues to shut out the remaining Faithfulness constraints.

Stratum 5: By this stage, all the rival candidates are excluded by some ranked constraint. The remaining three constraints (ID(ASP), ID(VCE) / ___ V, and ID(VCE)) prefer no losers, so are passed first to the jurisdiction of ***Favor Markedness***, then to ***Favor Activeness***. The latter, noticing that none of these constraints is active, invokes its special termination provision and dumps them all into the bottom stratum. The procedure that checks for termination (7.7.6) notices that all constraints are ranked and that all rival candidates are ruled out, so it records success.

Summing up, the ranking obtained by Low Faithfulness Constraint Demotion for Pseudo-Korean is as follows:

(29) **Stratum #1**

*D^h

Stratum #2

IDENT(ASPIRATION) / ___ V

Stratum #3

*[+V][–V][+V]
*ASPIRATION

Stratum #4

*[–SON/+VOICE]

Stratum #5

IDENT(ASPIRATION)
IDENT(VOICE)
IDENT(VOICE) / ___ V

Comparing this with (18), it can be seen that all of the crucial rankings are reflected in the stratal assignments. The distinction of Strata 1 and 2 (imposed by ***Favor Markedness***) does not reflect a crucial ranking, but no harmful consequences result from ranking *D^h over IDENT(ASPIRATION) / ___ V.[21]

By inspection, the grammar of (29) reveals 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.

As a check on the finished grammar, we can test to see if it overgenerates. To do 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 the following outputs:

(30)

Well-Formed Inputs		**Ill-Formed Inputs**	
input	output	input	output
/ta/	[ta]	/da/	[ta]
/ada/	[ada]	/d^{h}a/	[t^{h}a]
/t^{h}a/	[t^{h}a]	/ata/	[ada]
/at^{h}a/	[at^{h}a]	/ad^{h}a/	[at^{h}a]
/at/	[at]	/ad/	[at]
/tada/	[tada]	/at^h/	[at]
/tat^{h}a/	[tat^{h}a]	/ad^h/	[at]
/t^{h}ada/	[t^{h}ada]		
/t^{h}at^{h}a/	[t^{h}at^{h}a]		
/tat/	[tat]		
/t^{h}at/	[t^{h}at]		

Specifically, all well-formed inputs are retained, and all ill-formed inputs are "fixed"; that is, converted by the grammar into a well-formed output.[22] Thus, Low Faithfulness Constraint

---

[21] This is because the completed grammar maps underlying /d^{h}a/ to [t^{h}a], so that both *D^h and IDENT(ASPIRATION) / ___ V are satisfied.

[22] 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

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.[23] In these data files, it succeeds in producing “tight” grammars, which generate only forms that match (or are less marked than)[24] those given to it in the input.

7.9 Caveats

To recapitulate: infants are apparently able to learn a great deal about the phonology of the ambient language—specifically, the phonotactics—in the absence of negative evidence. Moreover, they most likely accomplish this learning with little or no information about morphology and alternations, hence without knowledge of underlying forms. I have developed a ranking algorithm, Low Faithfulness Constraint Demotion, with the goal of modeling this stage of phonological acquisition, and have found that, at least in a number of representative cases, the algorithm accomplishes its intended purpose.

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

7.9.1 Gradient Well-Formedness

The algorithm cannot deal with the fact that judgments of phonotactic well formedness are gradient (Algeo 1978); for example, a form like ?[dwɛf] seems neither perfectly right nor completely ill formed. Moreover, such patterns appear to be learned by infants. An example is the unusual status of final stress in English polysyllabic words, e.g. *ballóon*. Jusczyk, Cutler, and Redanz (1993) give evidence that this pattern is learned by infants at the same time they are learning the exceptionless patterns.

There *is* an algorithm that has proven capable of treating gradient well formedness, namely the Gradual Learning Algorithm of Boersma (1997), applied to gradient well formedness in Boersma and Hayes (2001). I have not yet succeeded in incorporating a suitable downward bias for Faithfulness into this algorithm.

---

vowel epenthesis and consonant deletion, rather than by alteration of laryngeal feature values. For further discussion, see section 8.1.

[23] Specifically: a file with the legal vowel sequences of (the native vocabulary of) Turkish, a file embodying the *azba* problem of Prince and Tesar (1999/this volume), and a family of files containing schematic “CV” languages of the familiar type, banning codas, requiring onsets, banning hiatus, and so on. See Appendix A for further discussion.

[24] 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 Low Faithfulness Constraint Demotion: in any language, a hypothetical form that incurs a subset of the Markedness violations of any actual form should be well-formed.

7.9.2 A Priori Knowledge

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 some discussion along these lines, see Boersma (1998, 2000).

The close reliance of Low Faithfulness Constraint Demotion on an adequate constraint inventory is highlighted by the following fact: if "irrational" Markedness constraints are included in the constraint set, the algorithm can fail. I have found that, for Pseudo-Korean, the algorithm fails to learn the correct ranking if one adds in a constraint (both typologically and phonetically unmotivated) that bans *un*aspirated stops. The reason is that *NOT ASPIRATED counts as a helper for IDENT(ASP) / ___ V in ruling out /t^ha/ → *[ta], which in turn robs the algorithm of its ability to recognize the special effectiveness of IDENT(ASP) / ___ V.

It follows that, whatever system learns the constraints must learn only sensible ones; it will not do to throw a great pile of arbitrary constraints into the algorithm's lap and hope that it will weed out the foolish ones by ranking them low. For a proposed method for discovering only sensible constraints, see Hayes (1999a).

7.9.3 Effectiveness

Low Faithfulness Constraint Demotion, unlike classical Constraint Demotion, is not backed by a mathematical proof establishing the conditions under which it will find an effective grammar. [25]

In the face of these caveats, I would say that the main point of this section is simply to establish the plausibility of an Optimality-theoretic approach to pure phonotactic learning, making no use of negative evidence.

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? Learning patterns of phonological alternation (i.e., morphophonemics) is probably a harder problem than phonotactics; some efforts in this direction are given by Tesar and Smolensky (2000), Zuraw (2000), and Albright and Hayes (1998).

One major issue in morphophonemics concerns the kinds of representations that are invoked to explain alternation. A classical approach is to assign every morpheme a unique UNDERLYING FORM, which abstracts away from the variety of surface realizations in a way that permits every

---

[25] Moreover, for purposes of such a proof, one would have to state precisely what is meant by "effective" in the context of pure phonotactic learning. One possible definition is the subset definition adopted by Prince and Tesar (1999/this volume): a ranking R is the most effective if there is no other ranking R′ that covers the input data and permits only a subset of the forms permitted by R. However, in the long run I think our main interest should lie in an empirical criterion: a phonotactic learning algorithm should make it possible to mimic precisely the well-formedness intuitions of human speakers.

surface allomorph to be derived by the phonology. The underlying form unifies the paradigm, because all the inflected forms are derived from it. Under this approach, the way to learn morphophonemics is to locate an underlying form for each morpheme, and adjust (perhaps by reapplying the same algorithm) the rankings of Faithfulness that were learned for phonotactics so that they cover alternations as well.

More recently, research has suggested (see for example Benua 1997, Kenstowicz 1998, Burzio 1998, Hayes 1999b, Kager 1999a, Steriade 2000) that underlying representations alone do not suffice to unify the paradigm: there is reason to think that the mutual resemblance of the surface realizations of a morpheme are due (either in addition, or perhaps entirely) to Optimality-theoretic constraints. The new constraint type posited is "output-to-output correspondence" constraints. These formally resemble Faithfulness constraints, but they require that a surface form not deviate in some phonologically defined property from the surface form of its morphological base, rather than from its underlying representation. Plainly, under this view a major task in the learning of phonological alternations is the ranking of the output-to-output correspondence constraints.

With this background, I turn to two particular problems in morphophonemic learning.

8.1 The Multiple-Repair Problem

To start, it is worth reemphasizing a point made above: because phonology is conspiratorial, knowing the phonotactics in advance is a powerful tool to use in learning morphophonemics. For example, Albright and Hayes's (1998) rule-based morphophonemic learning algorithm relies crucially on pre-existing knowledge of the phonotactics. Thus, in English past tenses, the algorithm experiments by taking the [-d] suffix it learned from verbs like *planned* or *hugged*, and attaching it to stems like *sip* or *need*, yielding *[sɪ**pd**] and *[ni**dd**]. Knowledge of phonotactics (the algorithm is told in advance that no words can end in *[pd] or *[dd]) permits the algorithm to discover phonological rules of voicing assimilation (/sɪp+d/ → [sɪpt]) and schwa epenthesis (/nid+d/ → [nidəd]). This in turn permits the unification of the structural change for past tense formation, as simply /X/ → /Xd/.

Within Optimality Theory, a particular difficulty for the transition from phonotactic to morphophonemic learning is that a phonotactically-illegal form can be "fixed" in various different ways. For example, a final /t^h/ in Pseudo-Korean could be fixed by deaspirating it, by deleting it, or by appending a final vowel. Low Faithfulness Constraint Demotion, however, seizes on just one way (fn. 22). Whatever turns out to be the right algorithm for morphophonemic learning in OT must be able to rearrange the ranking of the Faithfulness constraints during the transition from the phonotactics-only grammar to the more mature grammar. In particular, the Faithfulness constraints that are violated in accommodating underlying forms to the phonotactics must be ranked lower than those that are not—even if this was not the ranking that pure-phonotactic learning had earlier settled on.

8.2 A Trim-Back Problem: Grammatically-Conditioned Allophones

Another interesting problem in the study of morphophonemic 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 classical generative phonology called "boundary phenomena": instances where a stem + affix combination receives a different treatment than the same sequence occurring within a morpheme. For instance, in some English dialects *bonus* [ˈbõʊ̃nəs], with nasalized [õʊ̃], fails to form a perfect rhyme with *slowness* [ˈsloʊnəs], with oral [oʊ]. In traditional terms, nasalization is said to be "blocked across a suffix boundary." A similar case, worked out in Optimality-theoretic terms in Hayes (2000), is the non-rhyming pair some English speakers have for *holy* [ˈholi] vs. *slowly* [ˈsloʊli]: *slowly* avoids the monophthongal [o] characteristic of pre-/l/ position in stems, and thus shows blockage of monophthongization across the suffix boundary. Further cases are cited by Kiparsky (1988, 367).

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

Infants, who often lack the morphological knowledge needed to identify grammatically conditioned allophones, are liable to mistake them for cases of outright phonemic contrast.[26] 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.

8.2.1 The Ranking of Output-to-Output Correspondence

I will borrow from McCarthy (1998) the idea that output-to-output correspondence constraints are ranked high *a priori* by children acquiring language. McCarthy bases his claim on considerations of learnability: high ranking of output-to-output correspondence is needed to learn the absence of certain patterns of phonological alternation. Another source of evidence on this point comes from observations of children during the course of acquisition: children are able to innovate sequences that are illegal in the target language, in the interest of maintaining output-

[26] The reader who doubts this might further consider the effects of forms that are not morphologically transparent. 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]. Lack of morphological knowledge is particularly likely for infants, who probably learn phonology in part on the basis of stored phonological strings whose meaning is unknown to them (for evidence, see Appendix C).

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 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 top of the hierarchy, and that they are demoted only as the child processes the evidence that justifies their demotion.[27]

An example similar to the Marina case, involving American English Flapping (*[ˈsɪtɪŋ] for [ˈsɪɾɪŋ], on the model of [sɪt]), is given by Bernhardt and Stemberger (1998, 641).

8.2.2 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* [ˈsloʊnəs] is plausibly attributed to a correspondence effect with its base form *slow* [ˈsloʊ], where orality is phonologically expected. Likewise the diphthongal vowel quality of /oʊ/ of *slowly* [sloʊli] can be treated as a correspondence effect from the same base. The pair *writer* [ˈrʌɪɾɚ] vs. *rider* [ˈraɪɾɚ], though treated very differently in traditional phonology, likewise emerges as a correspondence effect: *writer* inherits its raised diphthong from the base form *write* [ˈrʌɪt], where is it 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 [ˈloʊli] *irrespective of the ranking of the relevant Faithfulness constraints*. This is because the output-to-output correspondence constraint governing diphthongal [oʊ] 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]/[oʊ] distinction. After the other relevant forms are also morphologically analyzed,

[27] 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.

then the entire burden of accounting for the [o]/[oʊ] distinction is assumed by output-to-output correspondence, and the erstwhile dominant Faithfulness constraints may safely sink to the bottom of the grammar. The result is that [o]/[oʊ] 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 *sitting* [ˈsɪɾɪŋ] is the correct present participle for *sit* [sɪt], 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.[28]

8.2.3 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* [ˈr**ʌ**ɪɾɚ] and *rider* [ˈr**a**ɪɾɚ] form a minimal pair. As just mentioned, they can be analyzed in OT with Markedness constraints that require the appearance of the raised diphthong [ʌɪ] before voiceless consonants (accounting for [ˈr**ʌ**ɪt]), along with a higher-ranked 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, [ˈr**ʌ**ɪɾɚ] vs. [ˈr**a**ɪɾɚ] looks just like a minimal pair.

Further light on the *writer*/*rider* 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:

(31)	*idle*	[ˈ**ʌɪ**ɾəl]	*idol*	[ˈ**aɪ**ɾəl]
	tire	[ˈt**ʌɪɹ**]	*dire*	[ˈd**aɪɹ**]
	bicycle	[ˈb**ʌɪ**səkəl]	*bison*	[ˈb**aɪ**sən]
	miter	[ˈm**ʌɪ**ɾɚ]	*colitis*	[kəˈl**aɪ**ɾəs]

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]

[28] An issue not addressed in the text is which of the child's two emerging grammars contains output-to-output correspondence constraints—is it her own production phonology, or her conception of the adult system? If the discussion above is right, they must occur at least in the internalized adult system, though perhaps they occur in the production system as well. The crucial empirical issue, not yet investigated to my knowledge, is this: do children like Marina and Gwendolyn recognize that forms like *[ˈexete] or *[ˈsɪtɪŋ] would be aberrant coming from adults?

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: [ˈrʌɪɾɚ]/[ˈraɪɾɚ] speakers are constantly migrating to [ˈraɪɾɚ]/[ˈraɪɾɚ] dialect regions, and vice versa. The [ˈraɪɾɚ]/[ˈraɪɾɚ] speakers of course have no [ʌɪ], and say *bison* [ˈb**aɪ**sən], *colitis* [kəˈl**aɪ**ɾəs], and so on. If a young learner of the [ˈrʌɪɾɚ]/[ˈraɪɾɚ] 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 [ˈraɪɾɚ]/[ˈraɪɾɚ] dialect are *not* analogously susceptible when they are exposed to migratory [ˈrʌɪɾɚ]/[ˈraɪɾɚ] 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 [ˈrʌɪɾɚ] and [ˈraɪɾɚ], 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 [ˈrʌɪɾɚ]/[ˈraɪɾɚ] speakers, adult [ˈraɪɾɚ]/[ˈraɪɾɚ] 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 morphophonemics 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 protocategories**. By the age of six months, infants have used knowledge of the statistical distribution of tokens to establish language-specific distributional protocategories, 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 the view presented here, learning at this phase takes place through the ranking of Faithfulness constraints against Markedness constraints, using 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 7.7) 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, following Smolensky (1996b).

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 production phonology that maps adult surface forms onto their own, simpler, output 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 the 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 8.2.2). 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.

Appendix A: Discussion of Prince and Tesar's paper [29]

Shortly after completing the first version of this paper (Hayes 1999c) I was pleased to learn that Alan Prince and Bruce Tesar had been working simultaneously (Prince and Tesar 1999) on lines very similar to ideas in my own paper, in particular those in sections 6 and 7 above. They likewise lay out pure phonotactic learning as a useful goal for learnability theory, and propose their own ranking algorithm, which they call Biased Constraint Demotion. Just as I suggest here, their algorithm ranks constraints based on a identity mapping in which each surface form is taken to be its own underlying form. Their algorithm also makes use of the principles that I refer to here as ***Avoid Preference For Losers***, ***Favor Markedness***, and ***Favor Activeness***. They eschew—deliberately (Prince and Tesar 1999, 19)—the principle of ***Favor Specificity***, and instead of ***Favor Autonomy*** they adopt principles that have the goal of "freeing up" Markedness constraints; i.e. Faithfulness constraints are selected for ranking with the goal of allowing Markedness constraints to be placed in the immediately following strata.

Since the two algorithms have the same purpose, I became curious to compare them. To this end, I tested both on three data simulations:[30] (1) the Pseudo-Korean problem laid out above; (2) the *azba* case invented by Prince and Tesar (1999, 20-22) and presented by them as a problem for Biased Constraint Demotion; (3) a simulation of the legal vowel sequences of Turkish: rounding may occur on noninitial vowels only when they are high and preceded by a rounded vowel in the previous syllable. Since detailed discussion of these simulations would exceed the space available here, I refer the interested reader instead to a Web page from which full descriptions of the simulations can be downloaded (http://www.linguistics.ucla.edu/people/hayes/acquisition). The result of all three simulations was that Low Faithfulness Constraint Demotion learned the correct (i.e., non-overgenerating) grammar, and Biased Constraint Demotion failed to learn this grammar; i.e. learned a less restrictive ranking.

This difference is less than meets the eye, however, because Low Faithfulness Constraint Demotion makes use of a ranking principle that Prince and Tesar deliberately avoid, namely ***Favor Specificity*** (7.7.4). ***Favor Specificity*** could easily be incorporated into Biased Constraint Demotion. In fact, I have experimented with doing this, and I find that when Biased Constraint Demotion is thus modified, it learns all three simulations correctly.

My simulations suggest a diagnosis of the problem that is faced by Biased Constraint Demotion when ***Favor Specificity*** is not included: general Faithfulness constraints, because they are general, tend to free up more Markedness constraints down the line. This problem was also noticed by Prince and Tesar (1999, fn. 9).

---

[29] This appendix has benefited greatly from advice I have received from Colin Wilson; he is absolved, however, of responsibility for errors. For Wilson's own new proposals in this area, see Wilson (2001).

[30] Testing was greatly facilitated by software code for Biased Constraint Demotion provided by Bruce Tesar as a contribution to the "OTSoft" constraint ranking software package (http://www.linguistics.ucla.edu/people/hayes/otsoft).

Given that both algorithms benefit from ***Favor Specificity***, it is worth considering Prince and Tesar's objections to it. The crucial point (see their discussion of the "ˈ*pa*:*to*" language; 1999, 19-20) is that we cannot always read special-to-general relations off of the structural descriptions of the constraints themselves (i.e. a superset structural description singles out a subset of forms). In some cases, the special-general relation is established only when we consider the action of higher-ranked constraints. Thus, in the "*pá*:*to*" language, where all initial syllables, and a number of other syllables as well, are stressed, the environment "in initial syllables" is special relative to the general environment "in stressed syllables." ***Favor Specificity*** cannot be applied here, given that the contingent special-general pattern is not yet known.

The answer to this problem, I suggest, may lie in having the learning algorithm work harder. Contingent subset relations among Faithfulness constraints are, I believe, learnable from the input data even where the constraint ranking is yet unknown, using computations whose scope would be extensive but hardly astronomical. In particular, if an algorithm kept track of the comparative applicability of the Faithfulness constraint environments to the available learning data, it would emerge clearly after only a few thousand words that the structural description "initial syllable" in the "*pá*:*to*" language never targets material not also targeted by the structural description "stressed syllable." [31]

I have also noticed what may be a positive attribute of Biased Constraint Demotion as amplified with ***Favor Specificity***. I noted earlier (7.9.2) that Low Faithfulness Constraint Demotion can be fooled by including an "irrational" constraint *NOT ASPIRATED. This is not so for the augmented version of Biased Constraint Demotion. Thus, whereas Low Faithfulness Constraint Demotion is dependent on the pre-existence of a "rational" set of Markedness constraints, specificity-amplified Biased Constraint Demotion perhaps may not be. Simulation histories for this comparison are given at the same Web page noted above.

In general, I would judge that any evaluation of algorithms for pure phonotactic learning must currently be considered tentative: given the history of linguistic theory in general, we can safely predict that no algorithm is likely to survive in the long term without modification. What is likely to produce the fastest progress, I think, is the accumulation of empirically realistic examples that test the capacities of ranking algorithms in a serious way.

Appendix B: Where do rival candidates come from?

An Optimality-theoretic grammar is an abstract formal object that selects for each input a winning candidate provided by the GEN function. The GEN function provides for each input a

---

[31] Method: form an *n* by *n* chart, where *n* is the number of contexts employed in Faithfulness constraints, and enter a mark in row *i* and column *j* whenever structural material in some word is included in context *i* but not context *j*. After sufficient data have been processed, any blank cell in row *p*, column *q*, together with a marked cell in row *q*, column *p*, indicates that context *p* is in a special-to-general relationship with context *q*.

It can be further noted that paired blank cells in row *p*, column *q* and row *q*, column *p* reveal contexts that—in the target language—turn out to be *identical*. This information would also be very helpful to Low Faithfulness Constraint Demotion. Accidentally duplicate constraints would be likely to distort the crucial measure "number of helpers" on which ***Favor Autonomy*** depends. An environment-tabulation procedure of the type just described could be used to ward off this problem.

potentially infinite[32] number of output candidates, each one lined up with the elements of the input form in all possible ways. The freedom provided to GEN implies that the explanatory force of any OT analysis derives entirely from the constraint set.

At a level closer to implementation (for example, as part of a processing model), this abstract conception is supplemented with additional mechanisms that guarantee that the winning candidate will be located correctly and efficiently. In one view, developed by Ellison (1994), Eisner (1997), and Albro (1997), GEN is modeled by a finite-state machine, and all operations using GEN are recast as operations on this grammar. The finite-state machine permits an infinite set to be modeled with finite means. Other approaches that also circumvent the infinite include work of Tesar (1995a, 1995b, 1999) and Walther (1996).

A more specific question for the present context is how to find the rival candidates that will lead to effective phonotactic learning. To solve this problem, Tesar and Smolensky (1996, 1998, 2000) and Prince and Tesar (1999/this volume) propose that learning is ERROR-DRIVEN: rival candidates are provided by running the grammar as learned so far and examining its errors. Another approach, however, would be to inspect fully the nearby regions of "phonotactic space," examining all candidates which are phonologically similar to the winner (differing from it by just one or two feature values or segment insertions/deletions).

In the simulations reported here, I used the latter approach, trying to make sure that my search went far enough to be sure that all the necessary candidates were found. As a check, however, I reran all the simulations with smaller learning data sets, in which every rival candidate could arise as an error from a partial grammar. Every simulation came out identically.

It remains an issue for the long term whether the use of rival candidates that aren't in the error-driven set aids—or perhaps hinders—phonotactic learning.

Appendix C: Batch vs. Serial Processing

Ranking algorithms can process their data either SERIALLY (examine one datum at a time, and modify the grammar on the basis of this examination) or in BATCH MODE (act at all times on the basis of knowledge of all the data). Tesar and Smolensky's original presentation of Constraint Demotion (1993) gives both batch and serial versions of the algorithm.

Low Faithfulness Constraint Demotion, in contrast, is necessarily a batch-mode algorithm. This follows from the definition of "minimum number of helpers" in (24b), which requires that all the rival candidates be inspected at once. Therefore, it presupposes that the child learning the phonotactics of her language can assemble a representative sample of data and retain the sample in memory for processing. This data sample might be the child's lexicon, or perhaps is simply a set of stored utterances that the learner (who is, perhaps, only eight months old) is not yet able to fully interpret; cf. Jusczyk and Aslin (1995, 19-20).

---

[32] Infinite candidate sets arise as the result of epenthesis; thus for underlying /sta/, GEN could in principle provide [əsta], [əəsta], [əəəsta], etc.

The batch processing mechanism of Low Faithfulness Constraint Demotion therefore presupposes that infants and children have capacious memory for phonological forms, an assumption that is plausible, I think, in light of the research literature.[33]

References

Aksu-Koç, Ayhan A. & Dan I. Slobin (1985). The acquisition of Turkish. In Dan Slobin (ed.) *The crosslinguistic study of language acquisition, vol. 1: the data*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Albright, Adam & Bruce Hayes (1998). An automated learner for phonology and morphology. ms, Department of Linguistics, UCLA. [http://www.linguistics.ucla.edu/people/hayes/learning]

Albro, Daniel (1997). Evaluation, implementation, and extension of Primitive Optimality Theory. MA thesis, UCLA. [http://www.linguistics.ucla.edu/people/grads/albro/ papers.html]

Algeo, John (1978). What consonant clusters are possible? *Word* **29**: 206-224.

Baker, William J. & Bruce L. Derwing (1982). Response coincidence analysis as evidence for language acquisition strategies. *Applied Psycholinguistics* **3**: 193-221.

Beckman, Jill (1998). *Positional Faithfulness*, PhD dissertation, University of Massachusetts. [Rutgers Optimality Archive 234, http://ruccs.rutgers.edu/roa.html]

Benua, Laura (1997). *Transderivational identity: phonological relations between words*. PhD dissertation, University of Massachusetts. [Rutgers Optimality Archive 259, http://ruccs.rutgers.edu/roa.html]

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*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bernhardt, Barbara H. & Joseph P. Stemberger (1998). *Handbook of phonological development from the perspective of constraint-based nonlinear phonology*. San Diego: Academic Press.

Best, C. T., G. W. McRoberts & 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.

Boersma, Paul (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam* **21**: 43–58.

Boersma, Paul (1998). *Functional phonology*. The Hague: Holland Academic Graphics.

[33] Jusczyk and Hohne (1997) show that 8-month-old infants recall words they heard in tape-recorded stories when tested in a laboratory two weeks later. Since only a small sampling of the words in the stories was tested, the number memorized by the infants must be far larger. Gomez and Gerken (1998) supplement this result by demonstrating the great speed and facility with which infants can internalize potential words: 12 month olds, on very brief exposure, can rapidly assimilate and remember a set of novel nonsense words, well enough to use them in determining the grammatical principles governing the legal strings of an invented language.

Boersma, Paul (2000). Learning a grammar in Functional Phonology. In Dekkers et al. (2000). 465-523.

Boersma, Paul & Bruce Hayes (2001). Empirical tests of the Gradual Learning Algorithm. *LI* **32**: 45-86.

Burzio, Luigi (1998). Multiple correspondence. *Lingua* **103**: 79-109.

Chomsky, Noam (1964). Current issues in linguistic theory. In Jerry A. Fodor & Jerrold J. Katz (eds.) *The structure of language*. Englewood Cliffs, NJ: Prentice-Hall.

Chomsky, Noam & Morris Halle (1965). Some controversial questions in phonological theory. *JL* **1**: 97-138.

Dekkers, Joost, Frank van der Leeuw & Jeroen van de Weijer (2000) (eds.) *Optimality Theory: phonology, syntax, and acquisition*. Oxford: Oxford University Press

Derwing, Bruce L. & William J. Baker (1986). Assessing morphological development. In Paul Fletcher & Michael Garman (eds.) *Language Acquisition: studies in first language development*. Cambridge: Cambridge University Press.

Dinnsen, Daniel A., Laura W. McGarrity, Kathleen O'Connor & Kim Swanson (in press). On the role of sympathy in acquisition. To appear in *Language Acquisition*.

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.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.html]

Fikkert, Paula (1994). *On the Acquisition of prosodic structure*. Leiden and Amsterdam: Holland Institute of Generative Linguistics.

Fodor, Jerry A., Thomas G. Bever & Merrill F. Garrett (1974). *The psychology of language*. New York: McGraw Hill.

Friederici, Angela D. & Jeanine E. Wessels (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics* **54**: 287-295.

Gerken, LouAnn (1994). Child phonology: past research, present questions, future directions. In Morton A. Gernsbacher, *Handbook of psycholinguistics*. San Diego: Academic Press.

Gnanadesikan, Amalia (1995). Markedness and Faithfulness constraints in child phonology. Rutgers Optimality Archive 75, http://ruccs.rutgers.edu/roa.html.

Gomez, Rebecca L. & LouAnn Gerken (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition* **70**: 109–135.

Guenther, Frank H. & Marin N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *JASA* **100**: 1111-1121.

Hale, Mark & Charles Reiss (1998). Formal and empirical arguments concerning phonological acquisition. *LI* **29**: 656-683.

Hayes, Bruce (1986). Assimilation as spreading in Toba Batak. *LI* **17**: 467-499.

Hayes, Bruce (1999a). Phonetically-driven phonology: the role of Optimality Theory and Inductive Grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds.) *Functionalism and formalism in linguistics, volume i: general papers*. Amsterdam: John Benjamins. 243-285.

Hayes, Bruce (1999b). Phonological restructuring in Yidiɲ and its theoretical consequences. In Ben Hermans & Marc Oostendorp, eds. (1999). *The derivational residue in phonological optimality theory*. Amsterdam: John Benjamins. 175-205.
Hayes, Bruce (1999c). Phonological acquisition in Optimality Theory: the early stages. Rutgers Optimality Archive 327, http://ruccs.rutgers.edu/roa.html.
Hayes, Bruce (2000). Gradient well-formedness in Optimality Theory. In Dekkers et al. (2000), 88-120.
Jun, Sun-Ah (1996). *The phonetics and phonology of Korean prosody: intonational phonology and prosodic structure*. New York: Garland.
Jusczyk, Peter W. & Richard N. Aslin (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology* **29**:1-23.
Jusczyk, Peter. W., Anne Cutler & Nancy J. Redanz (1993). Infants' preference for the predominant stress patterns of English words. *Child Development* **64**: 675-687.
Jusczyk, Peter W., Angela D. Friederici, Jeanine M.I. Wessels, Vigdis Y. Svenkerud & 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. & Hohne, Elizabeth A. (1997). Infants' memory for spoken words. *Science* **277**: 1984-1986.
Jusczyk, Peter W., Paul A. Luce & Jan Charles-Luce (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language* **33**: 630-645.
Kager, René (1999a). Surface opacity of metrical structure in Optimality Theory. In Ben Hermans & Marc Oostendorp (eds.) *The derivational residue in phonological optimality theory*. Amsterdam: John Benjamins. 207-245.
Kager, René (1999b). *Optimality Theory*. Cambridge: Cambridge University Press.
Kazazis, Kostas (1969). Possible evidence for (near-)underlying forms in the speech of a child. *CLS* **5**: 382-388.
Keating, Patricia A. (1984). Phonetic and phonological representation of stop consonant voicing. *Lg* **60**: 286-319
Keating, Patricia, Wendy Linker & Marie Huffman (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* **11**: 277-290.
Kenstowicz, Michael (1998). Uniform Exponence: exemplification and extension. Rutgers Optimality Archive 218, http://ruccs.rutgers.edu/roa.html.
Kiparsky, Paul (1988). Phonological change. In Frederick J. Newmeyer (ed.) *Linguistics: the Cambridge survey, volume 1, linguistic theory: foundations*. Cambridge: Cambridge University Press. 363-415.
Kiparsky, Paul & Lise Menn (1977). On the acquisition of phonology. In John Macnamara (ed.) *Language learning and thought*. New York: Academic Press.
Kirchner, Robert (1997). Contrastiveness and Faithfulness. *Phonology* **14**: 83-113.
Kisseberth, Charles (1970). On the functional unity of phonological rules. *LI* **1**: 291-306.
Koutsoudas, Andreas, Gerald Sanders & Craig Noll (1974). The application of phonological rules. *Lg* **50**: 1-28.
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. & Joanne D. Miller (1975). Speech perception by the chinchilla: voiced-voiceless distinction in alveolar plosive consonants. *Science* **190**: 69-72.

Kuhl, Patricia K. & Joanne D. Miller (1978). Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *JASA* **63**: 905-917.

Kuhl, Patricia K. & 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. & Denise M. Padden (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *JASA* **73**: 1003-1010.

Levelt, Clara C. (1994). *On the acquisition of place*. Leiden and Amsterdam: Holland Institute of Generative Linguistics.

Lombardi, Linda (1995). Laryngeal neutralization and syllable well-formedness. *NLLT* **13**: 39-74.

Macken, Marlys (1980). The child's lexical representation: the *puzzle-puddle-pickle* evidence. *JL* **16**: 1-17.

McCarthy, John (1998). Morpheme structure constraints and paradigm occultation. *CLS* **34**, *Vol. II: The Panels*.

McHugh, Brian (1986). Cyclicity in phrasal phonology: evidence from Chaga. *WCCFL* **5**: 154-164.

Menn, Lise (1983). Development of articulatory, phonetics, and phonological capabilities. In Brian Butterworth (ed.) *Language production*, *vol. 1*. London: Academic Press. 3-47.

Ohala, John J. (1983). The origin of sound patterns in vocal tract constraints. In MacNeilage, Peter F. (ed.) *The production of speech*. New York: Springer. 189-216.

Pater, Joe (1997). Minimal violation and phonological development. *Language Acquisition* **6**:3. 201-253.

Pater, Joe (this volume). From phonological typology to development of receptive and productive phonological competence: applications of Minimal Violation.

Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar.* Ms, available from http://ruccs.rutgers.edu/publicationsreports.html.

Prince, Alan & Bruce Tesar (1999). Learning phonotactic distributions. Rutgers Optimality Archive 353, http://ruccs.rutgers.edu/roa.html.

Pulleyblank, Douglas & William J. Turkel (2000). Learning phonology: genetic algorithms and Yoruba tongue root harmony. In Dekkers et al. (2000). 554-591.

Smith, Neilson (1973). *The acquisition of phonology*. Cambridge: Cambridge University Press.

Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Rutgers Optimality Archive 87, http://ruccs.rutgers.edu/roa.html.

Smolensky, Paul (1996a). On the comprehension/production dilemma in child language. *LI* **27**: 720-731.

Smolensky, Paul (1996b). The initial state and 'Richness of the Base' in Optimality Theory. Rutgers Optimality Archive 154, http://ruccs.rutgers.edu/roa.html.

Steriade, Donca (1997). Phonetics in phonology: the case of laryngeal neutralization. http://www.linguistics.ucla.edu/people/steriade/steriade.htm.

Steriade, Donca (2000). Paradigm Uniformity and the phonetics-phonology boundary. In Janet Pierrehumbert & Michael Broe (eds.) *Acquisition and the lexicon: papers in laboratory phonology V*. Cambridge: Cambridge University Press. 313-334.

Tesar, Bruce (1995a). *Computational Optimality Theory*. PhD dissertation, University of Colorado. [Rutgers Optimality Archive 90, http://ruccs.rutgers.edu/roa.html.]

Tesar, Bruce (1995b). Computing optimal forms in Optimality Theory: basic syllabification. Rutgers Optimality Archive 52, http://ruccs.rutgers.edu/roa.html.

Tesar, Bruce (1999). Robust interpretive parsing in metrical stress theory. *WCCFL* 7: 625-639. [Rutgers Optimality Archive 262, http://ruccs.rutgers.edu/roa.html.]

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.html.

Tesar, Bruce & Paul Smolensky (1996). Learnability in Optimality Theory. Rutgers Optimality Archive 156, http://ruccs.rutgers.edu/roa.html.

Tesar, Bruce & Paul Smolensky (1998). Learnability in Optimality Theory. *LI* **29**: 229-268.

Tesar, Bruce & Paul Smolensky (2000). *Learnability in Optimality Theory*, MIT Press, Cambridge, MA.

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*, Oxford: Blackwell.

Walther, Markus (1996). OT SIMPLE—A construction kit approach to Optimality Theory implementation. Rutgers Optimality Archive 152, http://ruccs.rutgers.edu/roa.html.

Werker, Janet F. & Chris E. Lalonde (1988). Cross-language speech perception: initial capabilities and development change. *Developmental Psychology* **24**: 672-683.

Werker, Janet F. & 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. & Patricia Keating (1986). On the naturalness of stop consonant voicing. *JL* **22**: 145-166.

Wilson, Colin (2001). Learning phonotactic distributions with targeted constraints. Paper presented at the 6th Southwest Workshop on Optimality Theory, Department of Linguistics, University of Southern California.

Zoll, Cheryl (1998). Positional asymmetries and licensing. Rutgers Optimality Archive 282, http://ruccs.rutgers.edu/roa.html.

Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, UCLA. [http://www-rcf.usc.edu/~zuraw/papers.html]

Bruce Hayes
Department of Linguistics
UCLA
Los Angeles, CA 90095-1543
(BHAYES@HUMNET.UCLA.EDU)