Formal and Empirical Arguments Concerning Phonological Acquisition*

Mark Hale
Concordia University, Montréal
hale1@alcor.concordia.ca

Charles Reiss
Concordia University, Montréal
reiss@alcor.concordia.ca

November 25, 1997

1 Introduction

1.1 Overview

This paper draws on the generative literature in phonological acquisition, as well as on the work of phoneticians and psycholinguists, in an attempt to propose a unified view of the acquisition of phonological competence. We start with an examination of the Optimality Theoretic (OT) account of the well-known comprehension/production dilemma in child language presented in Smolensky (1996ab, henceforth S). We will argue that S's model encounters two serious difficulties — the first concerning his proposed parsing algorithm and the second concerning the issue of the learnability of underlying forms. We offer alternative parsing algorithms and examine their implications for learnability and the initial ranking of OT constraints. Finally, we propose, based on evidence from a variety of sources, that the resolution of the comprehension/production dilemma lies not in phonological domain (linguistic competence), but rather in the domain of implementation of linguistic knowledge (performance). With a revision of certain aspects of the OT model for children's phonologies and of learnability theory in phonology, the paper attempts to contribute both to research on OT and to the study of phonological acquisition generally. It is not our aim here to argue for or against OT approaches to phonology. We will not hesitate, however, to point out flaws in the *application* of OT in the current literature. We hope that the improvements in the application of OT which we propose, as well as mention of points in which the theory seems to fare worse than alternative approaches, will lead to the kind of progress in

*The authors would like to thank Madelyn Kissock and Ida Toivonen for reading and discussing many drafts of this paper. Lisa Quesnel provided a very useful reference. Morris Halle, Sharon Inkelas, Orhan Orgun, Glyne Piggott and Donca Steriade provided encouragement and helpful discussion. Many other individuals, including an *LI* reviewer, and audiences at *NELS, WECOL, GALA, WCHTSALT, CLRF, FLSOM,* Stanford, Berkeley, Fresno, Toronto and Connecticut, have also been helpful in making us improve the presentation of this material.

the fields of phonology and acquisition that transcends the parochialism of particular frameworks.

1.2 The Phonological Enterprise

We believe that the study of phonology involves a characterization of mental representations and computations involving these representations. This view can be contrasted with with conceptions of phonology centered around the characterization of 'tendencies' and 'trends' sometimes subsumed under the ill-defined notion of 'markedness,' e.g., "The goal of phonology is the construction of a theory in which *cross-linguistically common* [but not necessarily universal-mh&cr] and well-established processes emerge from very simple combinations of the descriptive parameters of the model" (McCarthy 1988) [emphasis added-mh&cr]. We take the goal to be the development of a theory of *possible* human languages (i.e., the restrictions imposed upon the human language faculty by UG), not 'common' (statistically preponderant) human languages. Briefly, the question of what gets counted in determining 'commonness' leads to insurmountable difficulties, given the standard assumption among generative linguists that the object of study is computational systems, not speech communities. 'Common' features are artifacts of the sampling process, phonetic factors grammaticalized through historical change/acquisition (cf. Hale, to appear), etc., some of which are interesting and important domains of inquiry, but, which are strictly speaking, extra-grammatical.

1.3 Acquisition

The study of acquisition includes a characterization of the initial state of the grammar, S_0, and a theory of a learning path from S_0 to a subsequent state S_n. Our point of view is a strongly innatist approach to child grammar, but a performance-based account of many of the peculiarities of children's speech production. There exist opposing hypotheses in the acquisition literature which are directly relevant to these assumptions. These are outlined in (1) and (2) below.

(1) The Nature of Child Phonology

a. The Strong Identity Hypothesis, which holds that child phonology is governed by the same principles as adult phonology

b. The view that child phonology is fundamentally distinct from adult phonology — licensing processes unattested in adult language, dependent on a series of developmental stages, etc.

(2) The Nature of the Evidence

a. Deviations from target forms — in children's as well as adults' grammars — are to be attributed to performance effects, including non-linguistic cognitive and motor processing

b. Many deviations from target forms are the result of 'child phonology' (i.e., the child's phonological competence) — grammatical effects for which the target language provides no evidence

We will argue that both empirical evidence and learnability considerations favor the (a) hypotheses — that is, we support the Strong Identity Hypothesis and the hypothesis that deviations from targets are largely due to performance effects.

2 Smolensky (1996a)

By contrast, Smolensky (1996a) attempts to account for the peculiarities of children's speech output and the well-known discrepancy between their inaccurate production of adult words and their extremely accurate parsing of adult speech by appealing to the state of their grammars. In particular S rejects the notion that there is a 'dramatically greater performance/competence gap for children' (p.1). S proposes that a single OT grammar can generate both adult-like comprehension and child-like production. In order to model this discrepancy, S assumes that at the initial state of the grammar, S_0, OT Well-formedness (W) constraints are ranked above Faithfulness (F) constraints.[1] S's proposal is represented in (3):

(3) A single OT grammar generates adult-like comprehension and degenerate, unmarked child output if at S_0:
Wellformedness constraints >> Faithfulness constraints

2.1 Smolensky's Parsing Algorithm

S ingeniously proposes a distinction between the nature of production and comprehension in an OT model. This distinction is sketched in (4) and (5):

(4) PRODUCTION: OT-grammar selects the most 'harmonic' *output/surface form* (from the set of candidates which GEN provides) for a given *input/UR*

(5) COMPREHENSION: (same) OT-grammar selects the most harmonic *input/UR* for a given observed *output/surface form*

S's model is intuitively satisfying. The production of phonological forms represents the OT-grammar's selection of the most 'harmonic' output for a given input; the comprehension of phonological forms, on the other hand, represents the (same) OT-grammar's selection of the most harmonic *input* for a given output (again with the candidate set being generated by GEN). As S states, "What differs between 'production' and 'comprehension' is only *which structures compete*: structures that share the same underlying form in the former case, structures that share the same surface form in the latter case" (p.3). Note that this parsing model is meant to be valid for all grammars – those of adults as well as those of children.

[1] F-constraints value correspondence between input forms (underlying representations) and output forms (surface phonetic representations).

The result of making such a distinction is that the two operations will not always lead to the same input-output mapping, for example, at S_0. In (6) we have adapted S's constraint tableaux to show how the distinction works. Compare a child's pronunciation of a stored lexeme /kæt/ to the comprehension of this same lexeme as pronounced by an adult.[2] Since the W-constraints are all ranked high, every possible candidate form except for the most unmarked will violate some W-constraints. Like S, we have not distinguished among candidates on the basis of which specific W-constraints they violate, since this does not affect the argument. Again following S, we assume that the universally least marked output representation is [ta]. Since this candidate violates no W-constraints, it is selected by the grammar at this stage as the optimal surface form. Note that the same candidate will surface no matter what input form is used at this stage of the grammar.

In the bottom half of the tableau we illustrate how, in S's system, the child is able to parse adult [kæt] accurately as /kæt/ using the same grammar. Since the W-constraints represent surface well-formedness conditions, and the output, [kæt], is a given, the mapping from any possible underlying representation to this surface form will violate the same W-constraints. The surface form is known *a priori* to violate constraints against the presence of a coda, of an [æ] and of a dorsal consonant. Therefore, it is left to the F-constraints to select the most harmonic, the optimal, input-output mapping.

(6) The grammar at the initial state following Smolensky 1996a:[3]

- PRODUCTION: /kæt/ pronounced [ta] ('emergence of the unmarked')
- COMPREHENSION: [kæt] parsed as /kæt/, not [skæti], since mapping of /kæt/ to [kæt] is more harmonic than /skæti/ to [kæt] (only F-constraints matter).

[2] Following S, we ignore details of the pronunciation such as the aspiration on the initial voiceless stop.

[3] S p. 7: "What is given is the surface form, so the competing structures now [i.e., in comprehension—mrh&cr] are all those which are pronounced [kæt]." This is uninterpretable for child language as stated, since, by S's hypothesis, the grammar is responsible initially for maximally 'unmarked' pronunciation. There is no UR at this stage which is *pronounced* [kæt], even though adult [kæt] can be parsed as such. It seems clear from the discussion that S is trying to say that the mappings from every possible UR to surface candidate [kæt] are compared. The UR corresponding to the most harmonic mapping is the winner.

	Candidates	W-constraints (*æ, *Dorsal, *Coda…)	F-constraints (Parse, Fill,…)
PRODUCTION			
UR /kæt/			
☞	[ta]		*
	[kæt]	*!	
	[skæti]	*!	*
	[dajpɚræʃ]	*!	*
	etc.	*!	*
COMPREHENSION			
Surface [kæt]			
	/ta/	*	*!
→	/kæt/	*	
	/skæti/	*	*!
	/dajpɚræʃ/	*	*!
	etc.	*	*!

The winner in comprehension is marked with an arrow, →.

Note that S departs from most work on phonological acquisition in implicitly assuming (correctly, we believe) that children have access to the full set of universal features in constructing URs and that they store URs fully and accurately specified, according to what they hear in the target language.[4] Therefore, under S's own analysis the notion of *richness of the base* (e.g. Prince & Smolensky 1993, 191) becomes irrelevant (at least in the case of non-alternating forms) to the acquisition process. Richness of the base is a claim about the nature of OT grammars which states that there can be great latitude in the form of URs. For example, someone with a grammar of English could have all voiceless surface velar stops stored as clicks. Given the appropriate constraint ranking, *viz.*, with constraints against clicks ranked high, the surface forms could still be pronounced with normal velar stops. But given S's own assumptions about how parsing and the acquisition of URs proceeds, /kæt/, for example, could never be stored with a click. As a result, the notion of richness of the base becomes a computational curiosity of OT grammars which may be quite irrelevant to human language.

Note, in anticipation of the discussion to follow, that (a) given the arbitrariness of the sign, the child must first hear adult [kæt], store /kæt/, and associate this UR with the appropriate meaning before being able to generate the production mapping of /kæt/ to [ta] that S assumes; and (b) that as soon as the child does generate this form, the discrepancy between the child form and the adult form leads to a massive reranking of the constraints. In particular, it will trigger the demotion of most relevant W-constraints below the F-constraints that (even in the adult grammar) assure that the UR and surface forms of this word are in most respects identical. S's presumed initial high ranking

[4]We would qualify this by allowing for errors in parsing, which lead to incorrect representations. This is to be distinguished from merely incomplete representations assumed by researchers who posit that the child does not have access to all the features of the universal feature inventory. See Hale and Reiss 1996a-d for arguments against this position. Since S denies the relevance of performance in the characterization of language acquisition, he cannot appeal to such parsing errors. This idealization does not, however, affect the structure of his argument, which up to this point we accept.

of W-constraints thus leads to immediate promotion of those F-constraints which are relevant to each lexical item acquired.

2.2 Flaws in the Parsing Algorithm

The parsing algorithm that S proposes is meant to characterize comprehension by both children and adults. Unfortunately, this algorithm suffers from a serious flaw. We believe that this flaw precludes using such a model for the purposes of either child or adult language. The flaw lies in the fact that since the algorithm generates the **most** harmonic mapping from a UR to a surface form, it will never be able to account for the well attested and widespread phenomenon of surface ambiguity (merger) in natural language. A simple and well-known example will reveal this, though it is worth pointing out that any example which shows the effects of a phonological merger would do as well.[5]

Our example comes from German, which has two surface forms [rat], one derived from the UR /rat/ and the other from /rad/. We can capture the phenomenon of coda devoicing in German by assuming that a constraint against voiced codas is ranked above constraints demanding faithfulness to underlying voicing values. The relevant aspects of German grammar are sketched in (7).

(7) German surface ambiguity

- /rat/ > [rat] 'advice' & /rad/ > [rat] 'wheel'
- *VOICEDCODA >> FAITH[VOICE]

Consider what happens when a surface form [rat] is parsed by a speaker of German, using S's algorithm. Since the surface form is a given in parsing and since the choice of UR is left to the F-constraints, the most harmonic mapping from a UR to [rat] will be from the UR /rat/. The mapping from UR /rad/ to surface [rat] violates the same W-constraints as the mapping from /rat/ to [rat], but the former violates more F-constraints than the latter.

Indeed, in any case of surface merger, only the most 'unmarked' underlying lexeme will be chosen by the parse, since this lexeme provides the most faithful mapping. Note that this flaw in S's comprehension model is independent of the issue of the initial ranking of F-constraints — i.e, the model produces the wrong result both for children (who, according to S — though we do not share this assumption — have low-ranked F-constraints) and adults (who have elevated selected Faithfulness constraints).

S's proposed resolution of the comprehension/production dilemma thus gives rise to an unresolvable empirical problem — in any case of surface phonological merger, only the more 'unmarked' underlying lexeme can be comprehended by the parser. This is contraindicated by a wealth of evidence from virtually every human language. Note that the same difficulty arises in the case of syntactic comprehension: S's algorithm generates only the *most* harmonic parse for a given, potentially ambiguous, overt string.

[5] Hale and Reiss 1995 used an example from Fijian.

Note that any appeal to top-down processing to resolve this dilemma is inconsistent with well-established priming effects: "The general picture of lexical access during speech perception, then, is that it initially can discriminate only on phonological grounds. Only somewhat later in processing, after the syntactic and conceptual processors have gotten access to the list of possible candidates, can the ultimate choice of word be determined" (Jackendoff 1987:103; cf. references therein). In other words, the phonology makes multiple candidates available to further processing.

2.3 S's Learnability Argument

S provides a brief discussion of the mechanism of constraint-demotion in an OT model of the acquisition process (1996a:12). The learning algorithm is given as follows. At the initial stage, the child uses his/her grammar to parse (and produce) overt phonetic forms. This grammar diverges from the target, since all W-constraints outrank all F-constraints. Subsequently,

> The full structural descriptions assigned to the overt data are then used in the Error-Driven version… of the Constraint Demotion ranking algorithm (Tesar and Smolensky 1993): *whenever* the structural description which has just been assigned to the overt data (comprehension) is less harmonic than the current grammar's output (production), relevant constraints are demoted to make the comprehension parse the more harmonic. This yields a new grammar… [S 1996a:12, emphasis added]

In the case discussed by S in which a child produces [ta] for the underlying representation /kæt/, but correctly parses [kæt] as /kæt/, the structural description of the production mapping is more harmonic than that of the comprehension mapping. This is because the production process does not contain violations of highly-ranked W-constraints such as *CODA, *DORSAL, etc. whereas the comprehension process violates these constraints. When the child compares the structural description assigned to the overt data with that of the grammar's output and finds that the former is less harmonic, the necessary W-constraints are reranked such that they are lower than the relevant F-constraints.

2.4 Flaws in the Learnability Argument

The state of knowledge ('grammar') required for S's resolution of the comprehension/production dilemma in child language *cannot exist* given his learning algorithm. The virtual simultaneity of events which share a cause-effect relationship resulting from the application of this algorithm *precludes* a difference in comprehension and production via the mechanism asserted by S. We can illustrate this with the example just cited. The child must first correctly parse [kæt] as corresponding to underlying /kæt/. This is a necessary prerequisite to the acquisition of that lexeme (and the assumption is that the child must be able to do this in spite of having, e.g., *DORSAL ranked high). As soon as the child has done this, he/she will make the relevant comparison between the harmony of the comprehension form and that of his/her production form. The child will ascertain that the comprehension form is less harmonic than the production form and the relevant W-constraints will be demoted below the relevant F-constraints. As a result, there can be no steady-state period during which the child produces [ta] for /kæt/, i.e., no steady state in which production is consistently different (for a given stage in acquisition) than comprehension. Thus the grammar posited by S, which produces [ta] for /kæt/ but correctly parses [kæt] as /kæt/ could never exist. A stage which cannot exist cannot provide an account of stable features of child speech output of this type.

3 An Alternative Parsing Model

If we are to account for surface ambiguity, S's parsing algorithm must be replaced with one which generates a set of parses, not a single parse. We propose two such algorithms. In (8) we sketch an algorithm which is in the non-procedural spirit of OT. Under the assumption that massive computational complexity will ultimately be amenable to effective modeling, the algorithm culls the set of all possible URs to select those which can serve as a parse for a given surface form.

(8) 'Shrinking' algorithm in the 'spirit of OT':

> To select a set of possible parses for a surface form Φ (a) GEN generates all possible URs; $\Psi_i, i = 1,\ldots$ (b) for each UR Ψ_i GEN generates all possible surface candidates; (c) for each UR Ψ_i whose optimal output is Φ, Ψ_i is a parse for Φ.

In (9) we sketch a more procedural algorithm which starts with a set of parses containing only the one form which is identical to the surface form. The algorithm expands the hypothesis space of the parse by 'undoing' the effects of W-constraints.

(9) 'Expanding' algorithm

- Let the set Ψ of possible parses for Φ be equal to Φ; $\Psi = \{\Phi\}$. (This means that the first member of the set of candidate parses is identical to the surface form.)
- Start at the highest ranked constraint and proceed through the ranked constraint hierarchy;
- When a F-constraint which refers to a feature G is encountered, 'fix' the candidate set with respect to G. That is, all subsequent candidates must be identical to some Ψ_i with respect to the feature G.
- When a W-constraint is encountered, expand candidate set Ψ along precisely the dimension specified by the W-constraint. I.e., add candidates Ψ_i to the hypothesis space Ψ which differ from some preexisting candidate only in violating the current W-constraint.
- The algorithm ends when there is no remaining W-constraint which dominates a F-constraint. The parse candidate set thus produced $\Psi = \{\Psi_i, i = 1, \ldots, k\}$ represents the set of URs which will be neutralized to Φ by the grammar.

We can illustrate the operation of the algorithm in (9) by contrasting the parsing of English [rat] vs. [rad] with that of the ambiguous German [rat], assuming the URs in (10). (We have chosen the algorithm in (9) for purely expository purposes. The same result will be obtained using the algorithm in (8).)

(10) Contrastive parsing

- English: /rat/ 'rot' & /rad/ 'rod'
- German: /rat/ 'advice' & /rad/ 'wheel'

Since English does not have coda devoicing, we can assume that the ranking of *VOICEDCODA relative to FAITH[VOICE] in English is the opposite of that assumed for German, above. The operation of the parsing algorithm is sketched in (11) where a single UR is associated with surface [rat].

(11) English parse of [rat]: FAITH[VOICE] >> *VOICEDCODA

- The candidate set consists of /rat/
- The voicing specification of all segments in /rat/ is fixed by FAITH[VOICE]
- The candidate set is not increased by *VOICEDCODA, since [voice] has been 'fixed' in previous step
- The overt form is associated to a *single* UR, /rat/.

In German, on the other hand, the algorithm leads to an ambiguous parse, as desired, shown in (12).

(12) German parse of [rat]: *VOICEDCODA >> FAITH[VOICE]

- The candidate set consists of /rat/
- The candidate set is expanded to /rat/ **and** /rad/ by 'undoing' the W-constraint *VOICEDCODA
- The voicing specification of all segments in both /rat/ **and** /rad/ is fixed by FAITH[VOICE]
- The overt form is ambiguous — derivable from both /rat/ and /rad/.

Whichever algorithm turns out to be more useful, both of our proposed parsing algorithms are superior to S's, since they generate a *set* of candidate URs for a given surface form. Note that the argument developed here for phonology applies to the parsing of syntax, as well, whereas S's model will not generate differing underlying structures for sentences which are ambiguous on the surface. Theories of phonological and syntactic comprehension must account for such ambiguity. Any model which targets the most harmonic parse (i.e. a *single* candidate) instead of a set of acceptable parses fails to capture a critical aspect of natural language.

4 An Alternative Learning Path

We now turn to a consideration of the implications of these parsing algorithms for the study of the learnability of OT grammars. There is, first of all, an intuitive argument to be made against the position held by S and virtually every other scholar writing about the learning of OT grammars. Since surface forms and underlying forms tend to be 'fairly close' in adult grammars, it is clear that most F-constraints must ultimately be ranked higher than W-constraints. A theory which assumes that the F-constraints start out ranked high seems preferable *a priori* to one which posits massive reranking.

This intuitive argument can be supported by a demonstration that a parsing algorithm that actually works **requires** that F-constraints be initially ranked high in UG so that learners can converge on a lexicon. In contrast to S, then, we propose that the initial state of the grammar must be that shown in (13).

(13) At S_0: Faithfulness constraints $>>$ Wellformedness constraints

With the initial ranking proposed in (13) there is a single outcome to each parse at S_0. With the opposite initial ranking proposed by S in (1) a parsing algorithm like (8), which eliminates candidates from an initially infinite set, will generate the empty set; and an algorithm like (9), which adds candidates to an initially unary set, will explode the candidate set to include all possible URs. A lexicon is **unacquirable** under either scenario.

The table in (14) illustrates the acquisition of English /rat/ and /rad/ (forms AB) as opposed to German /rat/ and /rad/ (forms C-F), based on exposure to relevant surface forms. The German forms ending in [-əs] are genitive singular forms of the relevant nouns; because the stem-final stops occur between vowels, i.e., in onset position, in these forms, coda devoicing is not relevant. In the top half of the table we sketch the learning path under our assumption that all F-constraints are ranked high. Using either parsing algorithm, (8) or (9), the learner will be able to converge on a single UR for each surface form. Using (8), the high ranking of all F-constraints ensures that the optimal candidate is identical to the input form. Using (9), the high ranking of all F-constraints 'fixes' the value of all features of the surface form before the W-constraints can expand the set of candidate parses, again producing a single, fully-faithful parse at the initial state.

The parse chosen is the correct one with respect to the relevant adult grammar in each case except for form E. Ultimately, when the grammar generates the alternations due to coda devoicing, forms E and F will have to be collapsed. This process is obviously intimately related to the process of constraint reranking, whereby *VOICEDCODA is raised above FAITH[VOICE] to obtain the grammar of German.

(14) Comparing HIFAITH and LOFAITH at S_0 using a parser that works

With F constraints ranked HIGH		
Surface Form	Initial Hypothesis for UR	Path to adult UR
A. rat	rat	Unique, correct UR is selected initially.
B. rad	rad	Unique, correct UR is selected initially.
C. rat	rat	Unique, correct UR is selected initially.
D. ratəs	rat	Unique, correct UR is selected initially.
E. rat	rat	E & F stored differently, later collapsed by
F. radəs	rad	storing /rad/ and raising *VOICEDCODA.

With F constraints ranked LOW		
Surface Form	Initial Hypothesis for UR	Path to adult UR
A. rat	Ø *or* rat, rat⊙ʂ, bəbə...	There can be no learning path: each
B. rad	Ø *or* rat, rat⊙ʂ, bəbə...	production yields the maximally
C. rat	Ø *or* rat, rat⊙ʂ, bəbə...	unmarked utterance, say *ta*, as S
D. ratəs	Ø *or* rat, rat⊙ʂ, bəbə...	desired, but each parse yields Ø by (8)
E. rat	Ø *or* rat, rat⊙ʂ, bəbə...	or else everything generated by the
F. radəs	Ø *or* rat, rat⊙ʂ, bəbə...	UG-given W-constraints by (9).

The bottom half of the table illustrates the problem with assuming that F-constraints are initially ranked low. As shown by S, the production mapping will generate the maximally unmarked [ta] at the initial state. However, (8) will generate no parses — there is no UR which will surface as [rat] at this stage, since every UR will surface as [ta]. Algorithm (9) will generate an infinite set of candidate parses, since no features of the surface form Φ will be 'fixed' before the W-constraints expand the parse set to include forms with every possible W-constraint violation.

To appreciate that constraint reranking and choice of UR are part of the same task, it may be helpful to think about the most basic lesson in rule-based phonology: the interdependence of the choice of UR and the establishment of phonological rules. Unfortunately the obvious fact that the reranking of constraints and the collapsing of predictable allomorphs to a single form are two aspects of a single process has not been consistently recognized in the literature:

(15) Tesar and Smolensky 1993,1

> Under the assumption of innate knowledge of the universal constraints, the primary task of the learner is the **determination of the dominance ranking** of these constraints which is particular to the target language. We will present a simple and efficient algorithm for solving this problem, **assuming a given set of hypothesized underlying forms**. (Concerning the problem of acquiring underlying forms, see the discussion of 'optimality in the lexicon' in P & S 1993:9).[emphasis added—mrh&cr].

Turning to Prince & Smolensky 1993:9 we find

(16) Prince and Smolensky 1993, 192

> Lexicon Optimization. Suppose that several different inputs I_1, I_2,... I_n, when parsed by **a grammar G** [i.e. ranked constraint hierarchy—mrh&cr] lead to corresponding outputs O_1, O_2,..., O_n, all of which are realized as the same phonetic form Φ — these inputs are all *phonetically* identical with respect to G. Now one of these outputs must be the most harmonic, by virtue of incurring the least significant violation marks: suppose this optimal one is labelled O_k. Then the learner should **choose for the underlying form** for Φ the input I_k.

We might refer to this approach as the *Teufelkreis* or 'vicious circle' theory of language acquisition: the child needs a ranking to get URs and needs URs to get a ranking.[6] In a more recent paper Tesar and Smolensky (to appear) acknowledge the flaw in their approach, but consign a solution to the status of "one of the next steps in [their] research program" (sec. 9).[7]. They also propose that the child must make an "initial guess" as to the correct UR for each surface form. In our view, the only coherent interpretation of 'guess' is the initial hypothesis provided by the learning algorithm–this is what 'guess' means in a formal learning theory. Tesar (1997), states explicitly that he will "avoid the challenging problem of identifying and learning underlying forms, and assume that for a given overt form, the underlying form is apparent" (sec. 2). The only coherent interpretation of this is that URs are initially identical to observed (and produced—see below) surface forms. This follows from our initial assumptions, since it is equivalent to assuming that Faithfulness constraints are initially ranked higher than all Wellformedness constraints. Prince and Smolensky's (1993) process of Lexicon Optimization has a second drawback in addition to the *Teufelkreis* issue: as pointed out by Inkelas (1994), it only works for non-alternating morphemes, those that have a single surface realization.

So, constraint reranking and choice of a lexicon are part of the same task. Using the algorithm we have proposed, the learner can converge on a lexicon **only** if F-constraints are initially ranked **above** W-constraints. Again, referring to rule-based phonology, this is the equivalent of saying that the child has no rules at S_0, i.e., that adult phonetic forms are stored as perceived by learners. As we indicated above, this assumption actually leads to the simplest view of the learning path — one that does not require massive reranking with each newly acquired lexeme.

We now summarize the major points in our argument thus far. First, S's parsing algorithm selects only the most 'harmonic' UR, so it fails to account for surface ambiguity in any human language. Second, an algorithm which associates a perceived form with a set of possible URs is needed, since surface ambiguity does exist. Finally, using such an algorithm in acquisition, the learner can converge on a lexicon only if F-constraints are initially ranked above W-constraints.

[6] S's model discussed above attempts to address this problem, but fails for reasons we have discussed.

[7] Unfortunately, their initial suggestions for a solution do not appear promising, since they rely crucially on Output-output Faithfulness constraints and context-sensitive Faithfulness constraints. Both these powerful devices have been criticized on empirical and theoretical grounds — the first in Hale and Kissock (1997ab) and Hale, Kissock, and Reiss (1997ab), the second in Reiss (1996).

We will return to these learnability arguments below in section 5.4. For the moment, note that the 'emergence of the unmarked' has been touted as a property of children's speech derived from the initial ranking assumed by S and others. Given our learnability arguments thus far, we are forced to conclude that emergence of the unmarked is irrelevant to the description of children's grammars. In the next section we demonstrate that, though superficially implausible, it is in fact the case that children's grammars are faithful to observed (adult) target forms.

5 Accounting for the Disparity in Comprehension and Production

In an oft-quoted passage Chomsky (1965) characterizes the goals of linguistic theory as follows:

> Linguistic theory is concerned primarily with an ideal speaker listener, in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as *memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic)* in applying his knowledge of the language in actual performance. This seems to me to have been the position of the founders of modern general linguistics, and no cogent reason for modifying it has been offered. To study actual linguistic performance, we must consider the interaction of a variety of factors, of which the underlying competence of the speaker-hearer is only one. In this respect, study of language is no different from empirical investigation of other complex phenomena. [emphasis added—mrh&cr]

The emphasized part of this quotation characterizes no one if it does not characterize children, yet the phonological acquisition literature seems virtually unanimous in attributing the comprehension/production disparity to the grammar.[8] Why should this be so? The answer to this question seems to lie in the *nature* of children's errors. It is not simply the fact that children 'misarticulate,' since adults frequently do the same, but rather that they produce *systematic* misarticulations. Since one of the most salient attributes of grammars (and something that underlies generative theories) is exactly their systematicity, it is not surprising that researchers have analyzed child output in phonological terms. However, we believe such an approach is mistaken. Nothing in our proposed model indicates that the grammar or the process of acquiring it can account for the disparity between children's production and their comprehension. We hope to show in what follows that the empirical evidence gathered thus far on child speech does not justify appealing to natural language grammatical systems, but is instead the result of extra-grammatical factors, broadly categorized as 'performance.'

[8] Notable exceptions include Chomsky 1964, and Faber and Best 1994.

5.1 Arguments Against 'Child Phonology'

It seems preferable, *a priori*, to assume that child and adult grammars are organized by the same computational principles and contain the same sorts of processes. If this is an accurate assumption, then it is problematic that certain frequent aspects of 'child phonology' have long been known to have no parallels in adult phonological systems. One of the most widely discussed phenomenon is 'consonant harmony', responsible, for example, for the realization of 'duck' as [gək].[9] Another, less widely discussed, phenomenon is children's tendency to voice onset consonants. However, the total absence of across-the-board initial stop-voicing and place harmony in adult phonological systems indicates that such processes may not be possible in human phonological systems. Obviously, attributing them to children's grammars would be seriously misguided, if this is so. On the other hand, it has been pointed out that children often devoice codas and that similar processes occur in languages such as German and Russian. Examples such as this seem to support a supposed parallelism between acquisition and cross-linguistic tendencies and constitute the basis of much of the literature on markedness theory, from Jakobson to Stampe, as well as to the positing of new well-formedness constraints in current Optimality Theory. However, this appears to be a rather opportunistic appeal to 'markedness'. If 'markedness' is responsible for apparent parallels between child and adult phonological output (and is due to an underlying mechanism of the grammar), then we do not expect to find 'more marked' output in children,[10] let alone output that is unattested in adult language.

If the peculiarities of children's speech are to be attributed to performance effects, why are these effects sometimes paralleled by clearly phonological phenomena in adults? *Pseudophonological (i.e., performance-related) effects* (like coda devoicing) have been documented among populations other than children. Johnson, Pisoni, and Bernacki (1990) report on the intoxicated speech of the captain of the *Exxon Valdez* around the time of the accident at Prince William Sound, Alaska. They note, as do other studies of intoxicated speech, that the realization of segments may be affected by intoxication. They include the following among their list of observed effects:

(17) Some features of intoxicated speech

- misarticulation of /r/ and /l/
- final devoicing
- deaffrication

The accurate articulation of /r/, /l/, and affricates, as well as the existence of a voicing contrast in final position, all represent 'marked' features of English whose presence in the grammatical output is attributed, within OT theory, to relatively high-ranked faithfulness constraints regarding the features of the segments in question and by the absence of neutralizing rules in other theories. To account for Captain Hazelwood's output we have two options: 1) these instances of the 'emergence of the unmarked' are to be attributed to the impairment of his *performance system* by alcohol; or 2) the

[9] This example, as well as several additional ones, has been discussed in this respect by Drachman (1978).

[10] See fn. 16 below for examples of children's 'highly marked' output.

consumption of alcohol in sufficient quantities leads to constraint re-ranking or rule addition in adult grammars. Obviously, (1) is correct. We must recognize that the presence of ‘unmarked’ patterns in children’s bodily output does not permit us to attribute these patterns to the effects of the grammar, rather than to the ‘impairment’ of their performance system by immaturity.

Of course, this does not prove that none of the cited cases of ‘child phonology’ are due to the grammar, but it does demonstrate that other explanations are potentially available. Whether the output which forms the basis for the grammatical analyses proposed in work on child language represents output of the grammar or output of the ‘body’ is an empirical question, just as it is in the case of adult output.

A serious look at the data on child phonology reveals an additional problem — that is, knowing what exactly constitutes evidence. Consider, for example, the following attempts at producing adult [p^{h}ɛn] ‘pen’ collected from a 15 month old child in a thirty minute period: [mã^{ə}], [ṽ], [dɛ^{dn}], [hɪn], [^{m}bõ], [p^{h}ɪn], [t^{h}n̩t^{h}n̩t^{h}n̩] [ba^{h}], [dhauɴ], [buã] (Ferguson 1986, cited in Faber and Best 1994). The situation is actually worse than this case suggests since the transcriptions are also ‘patched’ by being filtered through the grammar of the adult listener who is also a speaker of English. Earlier studies such as Kornfeld and Goehl 1974 and current research at UCLA (Donca Steriade p.c.) indicate that transcriptions of child speech are rife with inaccuracy, in that acoustic analysis reveals subtle distinctions, for example, between supposedly merged adult /r/ and /w/, which transcribers tend not to observe.[11] *Any* use of production evidence will have to selectively cull the data (i.e., recognize the important role of ‘performance’ factors in children’s output).

The conclusion we would hope to derive from the arguments given above is that it is not at all obvious that a competence-only based account for children’s speech output is a desirable goal. Children’s output is actually not very parallel to so-called ‘unmarked’ aspects of adult languages when considered *in toto*. Furthermore, children’s output can be demonstrated to parallel systematic, yet unambiguously non-grammatical, performance effects, as seen in intoxicated speech.

It is also well-known that children’s non-grammatical abilities, such as physiological aptitude at performance and short-term memory capacity is limited. Even when their performance appears to parallel that of adults, we know, for instance, that their control of respiration and voicing differs qualitatively from that of adults (Faber and Best 1994:269). Occam’s razor demands attributing aspects of their speech output to these factors as the null hypothesis.

5.2 Some Non-Arguments Against a Performance System Account

5.2.1 Chainshifts

In his criticism of performance-based accounts for some attributes of children’s speech, S asserts that “invoking severe performance difficulties to account for the impoverishment of production relative to comprehension has several problems. Gross formula-

[11] As Faber and Best (1994:264) state, ‘[The] child may, despite the apparent lack of contrast, have acoustic differences between *red* and *wed* such that the initial consonants are perceived by adults as representing the same phonemic category.’

tions of this hypothesis, essentially claiming that children don't produce, say, a particular segment because their motor control hasn't yet mastered it" run into problems due to the chainshift data and the fact that children may imitate the relevant sounds. He cites as an example of chainshifts the following: children who produce 'thick' as [fɪk] cannot be said to be unable to produce [θ] since they produce this sound when saying 'sick' as [θɪk]. This argument has several problems. First, we must assume that the claims of such chainshifts are supported by better data than, for example, that given for a child's attempts at 'pen' cited above, i.e., we must be sure that any real generalization exists. Second, we must be sure that in transcribing 'sick' as [θɪk], there is some meaningful reason why the adult transcriber has chosen [θ], other than a mere impression that the sound the child made sounds a lot like the adult's own [θ]. In discarding the "gross formulations" of a performance-based account, S fails to consider a more coherent alternative. Clearly there is more intervening between the grammar and the utterance than just commands to the vocal tract. The merger of [θ] and [f] could be attributed to any intervening cognitive or motor process. No one, to our knowledge, has ever made the claim that children are physically incapable of moving their mouths to make a [θ]-like sound. A performance-based account holds rather that, when the performance system is given (by the grammar) the command to make a [θ], the vocal tract generates a sound *like* [f]. Given the commands to make a [s], the vocal tract may produce something *like* a [θ], resulting in an apparent chainshift.[12]

Moreover, S never actually demonstrates how an OT grammar can allow a child to produce a [θ] for an underlying /s/, but [f] for underlying /θ/. The treatment of chainshifts and other opacity effects has been one of the most difficult issues for OT. Reiss (1995, 1996) and others have demonstrated why a well-constrained OT grammar has difficulty with chainshifts. Simply put the problem is this: if the optimal output for underlying /θ/ is [f], why isn't [f] also a better output for underlying /s/ than [θ] is? Or similarly, [θ] is as well-formed (with respect to W-constraints) whether it corresponds to underlying /θ/ or /s/, and it is more faithful to /θ/; therefore, it should be the optimal candidate for the realization of /θ/.[13] At any rate, none of the existing proposals for dealing with opacity in OT (McCarthy 1994, Kirchner 1996) hold that chainshifts represent an unmarked, initial state of UG-phonology. The phenomenon of chainshifts appears, rather, to be highly marked and thus, given the assumptions made within OT concerning markedness and acquisition, must be the result of learning. S offers no plausible learning path which would lead the acquirer from the initial state to the highly marked constraint-ranking which would trigger chainshifts, and indeed it seems extremely unlikely that any such path could be coherently posited. There is *no* evidence in the PLD presented to a child acquiring English, for example, which would lead them to posit a constraint ranking which generated the [θ] > [f] but [s] > [θ] chain-

[12]Note that, when we write [s], we are writing a phonetic symbol — that is, a linguistic entity which is generated by the grammar. This is not to be confused with a physiological act which happens to look and sound like an [s], but is not an [s] *qua* linguistic representation. Sapir's famous comparison between a voiceless [ʍ] and the sound of a person blowing out a candle comes to mind.

[13]Given a theory of phonology which contains rules which apply in an ordered derivation, chainshifts are predicted to occur. In that sense, opacity has no status in a rule-based grammar. Opacity is just a point of logic, a possible result of applying rules in some order. This was recognized by Kiparsky and Menn (1977:73) who say that '[o]pacity is a property of the relation between the grammar and the data. An opaque rule is not more complex, merely harder to discover.'

shift. Moreover, becoming a competent adult speaker would then require readjusting the grammar (by resetting constraint parameters or by reranking complex constraints so low as to be inactive) so as to attain the adult grammar which has no chainshifts, just as the initial state of the grammar had none. Such 'Duke of York' models of the learning path are intuitively unappealing.

This is not the only way in which S's invocation of the chainshift data runs afoul of standard OT assumptions regarding markedness. The initial ranking is intended, within OT, to represent a grammar which generates only maximally unmarked output. However, the segment [s] is less marked than [θ] according to any of the standard criteria used in markedness theories — [s] is more common cross-linguistically, any language with a [θ] also has an [s], French speakers (e.g.) say [s] for English [θ], etc.

5.2.2 Imitation

The higher level of performance during direct imitation that S cites from Menn and Mathei (1992) as further evidence for his model is also contradictory to his own approach to the study of child speech output. There are two distinct accounts for what has been labeled 'imitation': (1) increased performance skill under concentration and (2) parroting. Under our account it is precisely during intense concentration on the act of performance that the child will perform better in carrying out the instructions provided by the grammar. Parroting clearly has no grammatical basis: a speaker of English can parrot a Cree sentence fairly well without acquiring a Cree grammar.

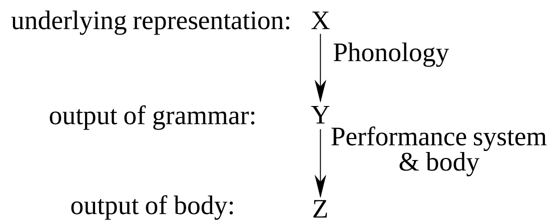
Under S's competence-only approach, neither of these types of 'imitation' can be accounted for. In the first type, since S assumes the *grammar* is responsible for, e.g., realization of [θ] as [f], increased attention to performance should lead only to a clearer hit of the target [f]. In the second type, to account for an English speaker's ability to imitate Cree, S would have to assume instantaneous acquisition of Cree constraint rankings. Clearly accounts for 'imitation' phenomena which invoke the performance system are to be preferred.

5.3 The nature of phonological acquisition

In addition to the empirical and 'learning theoretic' difficulties with S's approach to acquisition, we believe that the motivation for S's position is flawed. As we have seen, child speech output does not parallel unmarked adult speech, in general. In addition, there is clear evidence that 'there is a dramatically greater competence/performance gap for children,' (Smolensky 1996a:1) a hypothesis S rejects despite a rich body of empirical evidence to the contrary. (Note that this holds in speech as in virtually every other domain of physical activity.) We will now sketch an alternative theory which, we believe, avoids S's shortcomings. In particular, we build upon our earlier demonstration that if one adopts an OT framework at all, the initial state of the grammar must have all F-constraints ranked above all W-constraints in order to allow for the acquisition of a lexicon. We also assume, as demonstrated in Hale & Reiss (1997a), that children must have access to, and make use of, the full universal phonological feature set.

An evaluation of our hypothesis involves confronting the difficult problem of distinguishing, in children as we do in adults (in keeping with the Strong Identity Hypothesis), between an output of the phonology (a mental representation) and a real-time output of the body under some particular circumstances (those in effect at the time of utterance). The standard approach to the study of the speech of adults can be sketched as in (18) below.

(18) The Strong Identity Hypothesis assumes a single model for children and adults:



The phonology represents a mapping relationship between the underlying representation, X, and the output of the phonology for the string in question, Y (Y may be equal to X if the relevant F-constraints outrank the potentially relevant W-constraints). The performance system of the speaker, responsible for directing the body to ‘hit’ the output target in question, converts the mental representation Y to a set of articulatory commands which, filtered by contingent environmental factors (including accidental features of the body in question — its size, shape of its resonating cavities, etc. — as well as contingent features of the context in which the body finds itself at the moment of utterance — air pressure, humidity, wind speed and direction, etc.), generate an acoustic output, Z. Each of the arrows in (18) represents a mapping relationship: the grammar is responsible for the mapping of lexical entries onto output representations, the production system (and environmental factors) for mapping the output representation onto an acoustic realization. It is to be expected that these relationships, therefore, will be relatively systematic and regular; indeed, such systematicity is implicit in the notion of ‘mapping relationship.’[14] As is well-known from the study of adult phonetics, Z is highly variable (e.g., multiple articulations of /æ/ from the same speaker in the same session in the same phonological environment will still differ from one another acoustically). Given the complex set of factors determining its form (physical attributes — stable and accidental — of the speaker as well as numerous environmental effects) this is to be expected. Note that Z cannot, under any circumstances, be the same as Y: Y is a mental representation and Z is an acoustic (or articulatory) event.

It is widely acknowledged that while the speech perception skills of even very young infants are highly developed (Goodman and Nusbaum 1994), the articulatory skills of these same infants are much less sophisticated. Indeed, the general conception is that the sensitivity of the speech perception system is generally reduced over time to attend only (or primarily) to those distinctions critical for parsing the target language, while the production system moves from a state of virtually complete inarticulateness to full competence in articulating the target language. We would not be surprised, then, if the effects of an immature and generally incompetent production system on the Y of

[14] We are unlikely to be in a position to ascertain the systematicity of the effects of contingent environmental factors in most cases. The relevant information is not typically provided in studies of phonetic output.

(18) were more dramatic than the effects of the adult production system. That is, we would predict from general considerations such as these that Z should be more distant from Y (and more variable in its realization of a given Y) in children than it is in adults — this, indeed, is the definition of immature control of the production system.

The importance of distinguishing between the output of the body and the output of the grammar is not lessened by the difficulty of the task. The significance of the undertaking is often noted; e.g., in her survey of phonological acquisition, Macken (1995:672) notes that 'we must attempt the difficult, perhaps impossible, task of separating the grammar from the processor.' If we fail to rise to this challenge, we have abandoned the core concern of linguistics: to understand the nature of human linguistic competence.

The goal of distinguishing between these various effects can be stated as a simple question: how can we determine Y given the variability of Z? Even in the case of adults, where the relationship between Y and Z is assumed to be relatively close and stable over time, this question has proven to be a considerable challenge. We have shown that the parsing system involves essentially the same mechanisms as the generating system, applied in 'reverse,' as it were. That is, given some other person's output Z, the listener strips Z of contingent effects of the body and environment (such as the cues as to speaker identity), thereby generating a hypothesized Y. The listener then 'undoes' the effects of the phonological computational system thereby recovering an underlying representation, X, (or, in the case of ambiguous strings, a set of underlying representations) which would be expected, given the listener's grammar, to generate Y. The lexical item(s) corresponding to this X is (are) then accessed and passed to the other computational components of the grammar.

Factoring out the effects of the body and environment, i.e., converting a given Z to its corresponding Y, involves sophisticated knowledge on the part of the listener regarding the effects of different physiological properties (male vs. female, large resonating cavities vs. small, etc.) and external conditions (passing train, high wind, etc.) on the acoustic information contained in any given Z. The ability of listeners to successfully perform this conversion is key to their ability to parse a given real-time acoustic realization of [kæt] as being a reflex of 'cat' regardless of what speaker produced it under what conditions. It is important to note that, in this model of parsing, the listener's own bodily output, the listener's Z for a given Y, is completely irrelevant. The listener, who may be a 40-year old male, does not compare the output of a 12-year old female to his own bodily output. His bodily output reflects the grammar output 'augmented' by the effects of his own body and its context of utterance. Instead, he compares the output of her grammar to the output of his own grammar, his Y (which does not have speaker-specific characteristics, except, of course, for possible idiolectal features).

The Strong Identity Hypothesis would lead us to believe, in agreement with S, that it would be most sound, methodologically, to assume that precisely the same components and processes are to be posited for child speaker-listeners as we have sketched above for adults. If this is correct, it seems clear that the study of children's phonological parsing provides us with a critical tool to access information regarding Y without confronting directly the problems raised by the child's immature control of his or her output performance system. Parsing by-passes the listener's bodily output (Z), accessing instead the listener's grammar output (Y) and phonological system to arrive at an

X (or several Xs). Therefore, the study of children's parsing skills makes manifest aspects of the relationship between Y and Z which a study of children's bodily output alone may be unable to reveal.[15]

An example may help make this clear. It is well-known that at a certain stage in their development, children learning English may fail to distinguish, in their bodily output, between /s/ and /š/, producing both 'Sue' and 'shoe' as [su]. There are three likely scenarios which could account for this phenomenon, sketched in (19A-C).

(19) Three views of the s/š merger in child speech

	A		B		C		
	'Sue'	'shoe'	'Sue'	'shoe'	'Sue'	'shoe'	
underlying representation:	s	s	s	š	s	š	
	↓	↓	↓	↓	↓	↓	phonology
output of grammar:	s	s	s	s	s	š	
	↓	↓	↓	↓	↓	↓	performance
output of body:	s	s	s	s	s	s	

Under the scenario in (19A), the child has constructed identical underlying representations for both 'Sue' and 'shoe', each /s/-initial. Nothing affects these UR's in the computational component of the phonology[16] and the performance system implements the segments (more or less) accurately. Thus the body produces [s] (abstracting away from the speaker-specific and environmentally-induced effects). This hypothesis can be excluded on several grounds. First, it implies that the child's innate perceptual sensitivity to phonologically relevant contrasts in acoustic signals is of no use in the acquisition of the grammar, for it is not exploited in the construction of underlying representations. Second, it is difficult to imagine how a child who had posited the system in (A) could ever acquire English: the child accepts both adult [š] and adult [s] as realizations of his/her /s/, i.e., s/he has constructed the acoustic target space for /s/ as covering both the [s] and the [š] space of universal phonetics. The child will thus posit for the lexemes 'ship', 'sip', 'shame', 'same' the underlying representations /sɪp/, /sɪp/, /sem/, and /sem/. Since the child accepts both adult [s] and [š] as 'hits' for his/her own target [s], no positive evidence will ever reveal that the phonological system posited by the child fails to match that of the adult.[17] While the lexicon will contain more homophony than the adult lexicon, homophony must be permitted (given

[15] We assume that performance factors – e.g., distraction, shifts of attention, etc. – play a role in comprehension as well, but that they are much less dramatic than those affecting production, and thus we ignore them here. As Madelyn Kissock (pers. comm.) points out, it is standard in syntactic investigation to use grammaticality judgements, rather than output sampling, to determine a speaker's syntactic competence. Comprehension testing in children, we would argue, provides the closest parallel to grammaticality judgements possible.

[16] Note that even this scenario would require that Faithfulness to /s/ be ranked higher in the phonology than any W-constraint that would favor changing /s/, say, to [t].

[17] As we have shown elsewhere (Hale & Reiss 1996d), this is due to the Subset Principle, properly construed.

‘night’ and ‘knight’), so homophony alone will not rule out the child’s constructed grammar. Indeed, the fact that homophony can come into being through diachronic change indicates that this type of event does indeed take place on occasion (Hale, to appear). However, it cannot hold in our case, which assumes that the child eventually converges upon the adult output. Since scenario A essentially precludes acquisition of the s/š contrast, it must be rejected.

Scenario B holds that the child, by virtue of his/her inborn sensitivity to phonetic contrasts of potential phonological relevance, has constructed accurate underlying representations for ‘Sue’ and ‘shoe’, encoding the s/š contrast correctly. However, the child has ranked the constraints in the phonology in such a manner that Faithfulness to /š/ is ranked *lower* than a well-formedness constraint such as *š (i.e., ‘do not have a [š] in the output representation’). As can be seen from the tableau in (20), the correct output (for the hypothetical child speaker) will be generated by such a scenario.[18]

(20) Hypothesis (B): the s/š-merger

	*š	FAITH
☞ šu > su		*
šu > šu	*!	
☞ su > su		
su > šu	*!	*

It should be noted that the scenario hypothesized under (B) is a standard example, indeed, it is the definition, of *structural ambiguity*: distinct underlying structures have the same representation in the output of the grammar.[19]

Under scenario (C) in (19) the child’s underlying representations are, as in the (B) scenario, set up correctly. However, under this scenario, the output of the grammar maintains the s/š contrast. The failure on the part of the child to distinguish between the two segments in his/her speech output is attributed to a shortcoming of the performance system, which responds to the instruction to produce a [š] by emitting something like a [s] instead.[20] Such mismatches are, as pointed out above, to be expected if one assumes, as everyone does, that the child is not a fully competent articulator or processor.

[18]We have simplified the set of F-constraints — which should be separated into one constraint for each feature — into a single super-constraint FAITH, following the practice of Gnanadesikan (1995). This is of course just an abbreviation of the real tableau; however, this ‘abbreviatory’ convention may have serious empirical consequences.

[19]Structural ambiguity (19B) is distinct from lexical ambiguity (19A), where identity in input structures is responsible for the ambiguity of the output, as well as from production ambiguity (19C), where forms which are distinct at the level of the output of the grammar are produced the same due to some difficulty with the performance system.

[20]A reviewer suggests that ‘markedness’ considerations must be relevant in accounting for why the two underlying fricatives merge as [s], or an [s]-like sound, rather than a [š]-like sound. If there are real tendencies in the direction of performance mergers, then such data must play a role in developing a theory of children’s *performance*. Ida Toivonen (p.c.) has provided us with tapes of a Swedish-speaking child who consistently pronounces Swedish /s/ as something that sounds like a voiceless lateral fricative. Not only is this sound absent from the ambient language, but it is surely more ‘marked’ than [s] according to any of the usual (often non-formal) metrics of markedness. The apparently widespread phenomenon of gestural overlap in children’s speech production (see Masilon and Ross 1996) gives rise to articulatory events which, if analyzed as grammatical output, would necessarily lead to the conclusion that childrens’ grammars contain *more highly marked* representations than those of the target language in a great many instances.

The question of whether or not we can distinguish between the output of the grammar and the output of the body amounts to this: is there any empirical evidence that bears on the question of whether the child has [s] or [š] as the output of the *grammar* in the word for 'shoe' at the stage of acquisition under discussion? Given (19), there is, of course, a simple test which will resolve the question. Since the child, like any adult, is assumed to use the output of his or her own grammar to parse input from other speakers, rather than the output of his or her own body, it can be seen that if the child treats other speakers' [šu] as 'shoe' and other speakers' [su] as 'Sue' then the contrast between [s] and [š] persists to the level of the output of the grammar. It is well-known that children are indeed capable of identifying accurately whether an adult has said 'Sue' or 'shoe' in spite of their own production merger at this stage. Indeed, they reject adult renderings of 'shoe' as [su] — in spite of the fact that the adult output then matches their own — strongly supporting the hypothesis that their parser is not making reference to their own bodily output form.[21]

Additional support for this conception of the child's parsing mechanism can be seen in the results of Dodd (1975). Dodd showed that children failed to parse taped versions of *their own output* when it differed significantly from the output of an adult. Only by recognizing the irrelevance of the child's Z-form, the output of their body, to the parsing procedure can this result be accounted for.

Note that, under hypothesis (19B), the consistent and accurate parsing of the adult [s]:[š] contrast by children is completely unexpected (except under S's account, which we have shown to be untenable). Some earlier approaches to the study of children's phonology, which adopted Hypothesis B, needed to invoke special 'perception' grammars, distinct from 'production' grammars (Ingram 1976, 1989a,b). These perception grammars allowed the child direct access to underlying forms (rather than having the parser operate on the output of the child 'production' grammar). As Hale and Reiss (1995) and S himself note, such a theory is far too powerful, severing the connection between the child's linguistic competence and that of adults.

5.4 The high ranking of Faithfulness Constraints in UG

"It does not matter how small you are, if you have faith and a plan of action."
Fidel Castro, NY Times, 22 April 1959

We will argue in this section that a compelling case can be made for the assumption we made in (13) that all F-constraints are ranked, at S_0, above all W-constraints.[22]

The key to understanding acquisition under OT assumptions (or, indeed, under rule-based assumptions) is centered around the answer to the question of why there are

[21] See, e.g., Faber & Best (1994:266-7), where the authors point out that '. . . these studies show that by the time infants are starting productive use of language they can already discriminate almost all of the phonological contrasts of their native language. While they cannot yet produce adult-like forms, they appear, in many respects, to have adult-like representations, which are reflected, among other things, in their vociferous rejections of adult imitations of their phonologically impoverished productions.'

[22] This still describes a rather large class of possible UG's: there is a F-constraint for every feature used in any human language and there is presumably a rather large set of W-constraints. Within its constraint class (faithfulness vs. well-formedness), we are not in possession of any evidence which would allow us to determine a unique ranking for these constraints.

well-formedness constraints (or, in rule-based systems, rules) at all. Since we assume, with most other researchers on child language, that children's perception of adult output forms is quite accurate, what prevents the simple storage of strings to which the child is exposed (reduced to linguistically-relevant featural representations) both as input form and output form? This is, of course, the question of why we assume that phonology exists at all. We will not go through the well-established experimental results which show that allophonic relations exist, nor the *wug*-test and related empirical evidence for the existence of phonological processes. It is clear, however, that two phenomena, both involving 'optimizing' lexical storage, are of potential relevance to this issue: (1) the elimination of redundancy due to 'allophonic' variation; and (2) the reduction of the size of the lexicon by linking morphologically-related forms via phonological processes (forms which may have been representationally-distinct before the linking).[23] The only 'work' the computational component of the phonological system does is to license the elimination of predictable information from the lexicon. If we do not believe in such reduction, we do not need W-constraints (or, indeed, any constraints) at all. This is because having no phonology is the equivalent of having only F-constraints. With each elimination of redundant information from lexical entries, a W-constraint must come into play to 'resupply' that information. The W-constraints do nothing else in standard models of OT phonology.

Children are innately sensitive to all possible phonological contrasts. This is dictated by every child's ability to learn any human language, as well as by experimental evidence from the study of infant speech perception, regardless of which subset of the universal feature inventory is used in that language, or what the distribution of those features in that language might be. Lexemes must be initially stored in a fully specified phonetic form (i.e., specified to the degree allowed by the universal feature system used for human languages), since only language-specific information, deducible once a reasonably-sized set of such forms has been stored, will tell the child which features are relevant within the target language and which are not.

Given a sufficient number of forms stored in this manner, a process of Grammar Optimization (which is assumed to be a constant constraint on the procedure of establishing underlying forms) will lead the acquirer to deduce that, for example, the aspiration and the voicelessness of the [s] in [k^hæts] represents lexically redundant information. It will therefore no longer be necessary to store such information in the lexicon. The output target will not change because of this, of course — it was established on the basis of the child's interpretation of the adult target. Since the phonology represents the mapping between underlying representation and output target, a change in the underlying representation without a corresponding change in the target will require a simultaneous change in the computational component linking the two. Such changes will, necessarily, reduce the role of a particular F-constraint — the devoicing of the plural marker represents 'unfaithfulness' to the underlying representation — by the promotion of a W-constraint previously dominated by the F-constraint in question. Similar elevations of W-constraints over higher-ranked F-constraints will be triggered once the Grammar Optimization procedure has sufficiently clear evidence that it can

[23]We must remain agnostic on the question of whether the functional considerations of 'saving storage space' actually drive lexicon optimization, or whether such consolidation is merely a result of the automatic behavior of pattern recognition in linguistic stimuli by humans.

posit morphophonemic alternations of the normal type. A Grammar Optimization algorithm which breaks the vicious circle inherent in the theories discussed in section 4 can be found in Hale and Reiss (1997b).

Under this conception of acquisition, F-constraints must be ranked high within UG. The elevation of W-constraints takes place as a result of grammar optimization. The 'emergence of the unmarked' is therefore to be seen as the result of learning, rather than the accidental by-product of the structure of UG or the nature of OT grammars as opposed to other types.

6 Conclusions

The competence/performance contrast is, of course, accepted in phonological circles in the study of adult phonology, where it is used to determine which aspects of adult output phonological theory needs to concern itself with and which aspects it does not. We have argued that S's criticism of theories which appeal to both competence and performance as relevant to the study of child speech production is not only idiosyncratic in its rejection of a fundamental tenet of generative linguistics, but also self-defeating, since his own theory fails to account for the very cases he invokes. Furthermore, we claimed that S's model of the different nature of production and comprehension in OT does not allow for comprehension of both more and less 'marked' underlying forms which are neutralized by the phonology in production. We sketched an alternative to the OT model of acquisition proposed by S, who posits that F-constraints are initially ranked low in UG. In our model F-constraints are initially ranked high. Because S's model does not account for 'child phonology' and implies a more difficult learning task than ours without providing any benefits, our model should be adopted as the null hypothesis. Finally, given S's model of the initial state of the grammar and his denial (in practice) of the distinction between performance and competence, we are left with no explanation for the intermediate 'stages' in children's speech output on which S's approach is predicated. We attribute these to extralinguistic maturational development, both cognitive and physiological. When all these objections are considered, it is clear that S's model is not relevant to the evaluation of the relative merits of OT versus rule-based phonology. Our theory of initially high ranking F-constraints is trivially translatable to a rule-based theory where the initial state contains no rules. The crucial aspect of our theory is its orthodoxy: performance and competence are both needed to account for human speech behavior.

It is harder to maintain this distinction in L1 acquisition studies than in the study of adult native-speaker phonology because of the greater disparity between competence and performance, but it is necessary (cf. Epstein, Flynn and Martohardjono (1997) for similar arguments in L2 acquisition). Figuring out what intervenes between the grammar and the 'mouth' is difficult, but, one hopes, not impossible. As an example of how this problem has been approached in another domain of innate human behavior, consider the following result from a study of infants 'knowledge' of how to walk, discussed by Faber and Best 1994, following Thelen and Ulrich 1991: 'if the needs for balance and for ankle extension are removed, by holding infants with their feet touching a backward-moving treadmill, some infants as young as one month old will stay in place by stepping forward in the alternating pattern characteristic of adult walking.' Apparently, manipulation of their production system allowed for closer observation of their competence as walkers. In a sense, these babies *knew* how to walk like adults, but their performance was hindered by factors including physiological considerations like the relative weight of their heads to their bodies and the state of their musculature, as well as their cognitive inability to synchronize 'input from the visual and vestibular systems.' As the relative weight of head and body approaches that of adults, and as other cognitive and physiological systems mature, the child's performance system catches up with the innate knowledge of how to walk. We propose that learning to talk follows a parallel path.

S makes several claims, including those listed in (22):

(22) Major claims of Smolensky (1996a)

(a) the 'markedness' phenomena described in the child phonology literature are paralleled by phenomena of adult phonology;

(b) speech output during direct imitation and the existence of chainshifts in the mapping of adult language to child language argue for a competence- (or grammar-) based approach to child speech production;

(c) an OT analysis can maintain that children possess a single grammar and still capture the comprehension/production discrepancy in child language (which is known to be difficult in a rule-based approach); and,

(d) Faithfulness Constraints (F-constraints) must initially be ranked below Well-formedness constraints (W-constraints).

We have demonstrated that points a-d are untenable, providing new arguments based on evidence which is well-known in the linguistic, psycholinguistic and phonetic literature as well as on learnability considerations. Having challenged some widespread assumptions about the nature of phonological acquisition and its study, we have shown that the evidence actually supports the following claims:

(23) Our claims

(a) the supposed parallels between child and adult patterns of 'markedness' are illusory;

(b) improved performance under imitation is predicted only by a theory which appeals to a performance basis; chainshifts are not a problem for theories which take a 'competence *and* performance' approach (and, in fact, they remain unexplained in competence-only accounts, such as Smolensky's);

(c) the OT model of comprehension posited by S is empirically inadequate; and,

(d) within an OT framework learnability considerations favor an initial ranking in which F-constraints outrank W-constraints.

The failure of S's account also requires us to reject the claim that the proposed solution to the performance/competence debate in child phonology can be taken as evidence for the superiority of Optimality Theory over rule-based phonology.[24]

In response to some of the criticisms we have made here concerning the methodology used in many acquisition studies, we have been told that there would be nothing left for acquisitionists to do if we decide that the data is, in general, misleading. We think that this is an overly pessimistic view. There remain open several paths to a better understanding of children's phonological systems, if we are willing to devise ingenious experiments that force them to be revealed.

In order to interpret any data derived from speech behavior the performance/competence distinction must be maintained: an explicit characterization of the boundaries between the two should be one of the primary goals of phonological theory, since it defines the sphere of inquiry with which we must concern ourselves. It is clear that a more explicit theory of performance (or rather several theories) is a necessity; however, it must be predicated upon a coherent theory of grammatical competence. These issues were raised more than thirty years ago with characteristic lucidity by Chomsky (1964:39), who notes, commenting on a paper on language acquisition, that there is:

> a general tendency to oversimplify drastically the facts of linguistic structure and to assume that the determination of competence can be derived from description of a corpus by some sort of sufficiently developed data-processing techniques. My feeling is that this is hopeless and that only experimentation of a fairly indirect and ingenious sort can provide evidence that is at all critical for formulating a true account of the child's grammar (as in the case of investigation of any other real system)... I make these remarks only to indicate a difficulty which I think is looming rather large and to which some serious attention will have to be given fairly soon.

References

Chomsky, Noam. 1964. Formal Discussion in response to W. Miller & S. Ervin, 'The Development of Grammar in Child Language,' in *The Acquisition of Language*, Ursula Bellugi and Roger Brown (eds.), Chicago: University of Chicago Press.

Dodd, Barbara. 1975. Children's understanding of their own phonological forms. *Quarterly journal of experimental psychology* 27:165-172.

Drachman, Gabarell. 1978. Child language and language change: a conjecture and some refutations. In Jacek Fisiak (ed.), *Recent developments in historical phonology*, The Hague: Mouton.

Epstein, Samuel David, Suzanne Flynn and Gita Martohardjono. 1997. Second Language Acquisition: Theoretical and Experimental Issues in Contemporary Research. *Behavioral and Brain Sciences.*

[24]This is not, we point out, the same as an argument that an OT model, *in general*, is not preferable to a rule-based model, merely that this particular argument does not support such a claim.

Faber, Alice and Catherine T. Best. 1994. The perceptual infrastructure of early phonological development. In *Reality of Linguistic Rules*.

Ferguson, Charles. 1986. Discovering sound units and constructing sound systems: It's child's play. In Perkell, Joseph S; and Klatt, Dennis H (eds.) *Invariance and variability in speech processes*, 36- 53. Hillsdale, NJ: Erlbaum.

Gnanadesikan, Amalia E. 1995. Markedness and faithfulness constraints in child phonology. ms., Rutgers Optimality Archive.

Goodman, Judith and Howard Nusbaum. 1994. *The development of speech perception*. Cambridge: MIT Press.

Hale, Mark. to appear. *Theory and Method in Historical Linguistics*. Oxford: Blackwell.

Hale, Mark and Madelyn Kissock. 1997a. The phonology-syntax interface in Rotuman. *Proceedings of the Fourth Austronesian Formal Linguistics Association*, UCLA.

Hale, Mark and Madelyn Kissock. 1997b. The incomplete phase in Rotuman: a synchronic analysis. Ms., Concordia University and Oakland University.

Hale, Mark, Madelyn Kissock and Charles Reiss. 1997a. Evaluating the evidence for output-output correspondence. *BLS 23*. Berkeley: Berkeley Linguistics Society.

Hale, Mark, Madelyn Kissock and Charles Reiss. 1997b. What is output? Output-output correspondence in OT. *WCCFL 16*. Stanford: CSLI.

Hale, Mark and Charles Reiss. 1995. On the Initial Ranking of OT Faithfulness Constraints in Universal Grammar. Poster presentation at Stanford Child Language Research Forum. Also available in *Concordia University Working Papers in Language and Linguistics* and Rutgers Optimality Archive.)

Hale, Mark and Charles Reiss. 1996a. How to Parse (and How Not to) in OT Phonology. Proceedings of *NELS* 27, McGill University. Ms. Concordia University.

Hale, Mark and Charles Reiss. 1996b. The Comprehension/Production Dilemma in Child Language: A Response to Smolensky. ROA.

Hale, Mark and Charles Reiss. 1996c. Evidence in Phonological Acquisition: Implications for the Initial Ranking of Faithfulness Constraints. *Proceedings of CLRF 28*. Stanford: CSLI.

Hale, Mark and Charles Reiss. 1996d. The Subset Principle and acquisition of the phonological inventory. ms, Concordia University. [Presented at FLSM, Ohio State University.]

Hale, Mark and Charles Reiss. 1997a. Phonological underspecification and the subset principle. Proceedings of the Western Conference on Linguistics, UC-Santa Cruz.

Hale, Mark and Charles Reiss. 1997b. Grammar Optimization: The simultaneous acquisition of constraint ranking and a lexicon. Ms., Concordia University.

Hale, Mark and Charles Reiss. Markedness, change, and acquisition: What is phonology a theory of? In preparation.

Ingram, David. 1976. Phonological analysis of a child. *Glossa* 10:1,3-27.

Ingram, David. 1989a. *First language acquisition: method, description, explanation.* Cambridge: Cambridge University Press.

Ingram, David. 1989b. Phonological disability in children. London: Cole & Whurr (2nd edition).

Jackendoff, Ray. 1987. *Consciousness and the Computational Mind.* MIT Press: Cambridge, MA.

Johnson, Keith, David Pisoni and Robert Bernacki. 1990. Do voice recordings reveal whether a person is intoxicated? A case study. *Phonetica* 47:215-237.

Kiparsky, Paul and Lise Menn. 1977. On the Acquisition of Phonology. In *Language, Learning and Thought*. Ed. J Macnamara. New York: Academic.

Kirchner, Robert. 1996. Synchronic Chain Shifts in Optimality Theory. *LI* 27:341-50.

Kornfeld, J. R. and H. Goehl. 1974. A new twist to an old observation: kids know more than they say. In *Papers from the Parasession on Natural Phonology*, A. Bruck, R. Fox and M. LaGaly (eds.). Chicago:CLS.

Macken, Marlys A. 1995. Phonological acquisition. In John Goldsmith (ed.), *Handbook of phonological theory*. Oxford: Blackwell.

Masilon, Todd and Kie Ross. 1996. Weak syllable deletion: an Articulatory Phonological Account. ms. UCLA Linguistics Department.

McCarthy, J. 1988. Feature geometry and dependency: a review, *Phonetica* 45:84-108.

McCarthy, J. 1994. Remarks on Phonological Opacity in OT. Ms. UMass, Amherst. To appear in *Proceedings of the Second Colloquium on Afro-Asiatic Linguistics.* J. Lecarme, J. Lowenstamm and U. Schlonsky, eds.

McCarthy, John and Alan Prince. 1993. *Prosodic Morphology I: constraint interaction and satisfaction*. Ms. University of Massachusetts, Amherst, and Rutgers University, New Brunswick, NJ. To appear as Linguistic Inquiry Monograph, Cambridge, MA: MIT Press.

McCarthy, John and Alan Prince. 1994. The emergence of the unmarked: optimality in Prosodic Morphology. *NELS* 24:333-379.

Menn, Lise and Edward Mathei. 1992.The two-lexicon approach to child phonology: looking back, looking ahead. In *Phonological Development: Models, Research, Implications*, ed. Charles A. Ferguson, Lise Menn, and Carol Stoel-Gammon. Timonium, Md.: York Press.

Prince, A. and P. Smolensky. 1993. *Optimality Theory: Constraint Interaction in Generative Grammar*. Technical Report Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, N.J.

Reiss, Charles. (1997). Unifying the Interpretation of Structural Descriptions. *Proceedings of WCCFL XIV*:413-428. CSLI Publications/Cambridge University Press.

Reiss, Charles. 1995. Deriving an implicational universal in two theories of phonology. ms., ROA-102.

Reiss, Charles. 1996. Deriving an implicational universal in Optimality Theory. *Proceedings of NELS 26*.

Smolensky, Paul. 1996a. On the comprehension/production dilemma in child language. *LI* 27.

Smolensky, Paul. 1996b. The Richness of the Base. ROA.

Tesar, Bruce. 1997. Multi-recursive constraint demotion. ROA-197.

Tesar, Bruce and Paul Smolensky. 1993. *The learnability of Optimality Theory: An algorithm and some basic complexity results.* Technical Report, Computer Science Department, University of Colorado, Boulder (also ROA-156).

Tesar, Bruce and Paul Smolensky. To appear. The learnability of Optimality Theory. *LI*.

Thelen, Esther and Beverly D. Ulrich. 1991. *Hidden Skills: A Dynamic Systems Analysis of Treadmill Stepping During the First Year*. Monographs of the Society for Research in Child Development, No. 223, vol. 56, No. 1 [with commentary by Peter Wolff].