Formal and Empirical Arguments Concerning Phonological Acquisition

Mark Hale and Charles Reiss
Concordia University, Montréal

1. Introduction

This paper draws on the generative literature in phonological acquisition, as well as on the work of phoneticians and psycholinguistics, 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 given by Smolensky (1996ab). We will argue that S's model encounters two related serious difficulties — the first co n-cerning 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. We also propose a reanalysis of the Subset Principle in phonology, especially as applied to the acquisition of phonological 'inventories'. Finally, we propose, based on evidence from a variety of sources, that the resolution of the comprehe n-sion/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 a t-tempts to contribute both to research on OT and to the study of phonological acquisition ge nerally.

We believe that the study of phonology involves a characterization of mental representations and computations involving these representations. In other words, phonology is grammar. Alternative views include 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/impossible human language, not "common" (statistically preponderant) human languages. Briefly, the question of what gets counted in determining "commonness" obviously leads to insurmountable difficulties, given the standard assumption among generative linguists that the object of study is computational systems, not speech communities, for example. "Common" features are artifacts of the sampling process, phonetic factors grammaticalized through historical change/acquisition (cf. Hale 1995), etc., some of which are interesting and important domains of inquiry, but, which are strictly speaking, extra-grammatical.

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 a p-proach 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 distinctions. 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 develo p-mental 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 la n-guage provides no evidence

We will argue that both empirical evidence and learnability considerations favor the (a) hypoth e-ses—that is, we support the Strong Identity Hypothesis and the hypothesis that deviations from ta r-gets are largely due to performance effects.

2. Smolensky (1996a)

Smolensky (1996a, 1996b, henceforth S) 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 grammar. 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 if one assumes that at the initial state of the grammar, S_0, OT Well-formedness (W) constraints are ranked above Faithfulness (F) constraints which value correspondence between input forms (underlying representations) and output forms (surface phonetic representations). S's proposal is represented in (3):

(3) 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 exploits 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 can-didates which GEN provides) for a given *input/UR*

(5) COMPREHENSION: (same) OT-grammar's selects the **most** harmonic *input/UR* for a given ob-served *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 set of candidates being ge n-erated by GEN); 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). S demonstrates that the two algorithms will not always converge upon the same input-output mapping at S_0, for instance, leading to a difference in production and comprehension. 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).

In (4) we have adapted S's constraint tableaux to show how the distinction works. Compare the pronunciation of a stored lexeme /kæt/ to the comprehension of this same lexeme as pronounced by an adult.[1] 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 structure of S's valid 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.

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

(4) The grammar at the initial state following Smolensky 1996

- 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]

	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.[3] 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 velar stops stored as clicks, but given the appropriate constraint ranking, e.g. with con-

[2] 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.

[3] 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.

straints 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 to be quite irrelevant to human language.

2.2 Flaws in the Parsing Algorithm

Unfortunately, the parsing (comprehension) algorithm that S proposes, for both children and adults, suffers a serious flaw which we believe precludes the possibility of using such a model for the pu r-poses 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, or merger, in natural language. A si m-ple and well-known example will reveal this, though it is worth point out that any example which shows the effects of a phonological merger would do it as well.

German, for example, 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 co n-straint against voiced codas is ranked above constraints demanding faithfulness to underlying voicing values. The relevant aspects of German grammar are sketched in (5).

(5) German (or any other) 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.

This is, of course, a general result: in any case of surface merger, only the most 'unmarked' u n-derlying 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 unr e-solvable 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 ev i-dence from virtually every human language. Note that the same difficulty arises in the case of sy n-tactic comprehension: S's algorithm generates only the *most* harmonic parse for a given, potentially ambiguous, overt string.

2.3 Learnability

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 ("incorrect", because all W-constraints outrank all F-constraints) grammar to parse (and produce) overt phonetic forms. 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 process is more harmonic than that of the comprehension process. The production process obeys highly-ranked W-constraints such as *CODA, *DORSAL, etc. whereas the comprehension process violates these constraints (and thus is less harmonic). 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

It seems that 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 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 and therefore cannot provide a competence-based account of stable features of child speech output of this type.

3. An Alternative Parsing Model

It is clear that S's parsing algorithm must be replaced with one which generates a *set* of parses, not a single parse, if we are to account for surface ambiguity. We propose two such algorithms for parsing a surface form Φ. In (6) 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.

(6) '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; Ψ_i, i=1, . . . (b) for each UR Ψ_i GEN generates all possible surface ca n-didates; (c) for each UR Ψ_i whose optimal output is Φ, Ψ_i is a parse for Φ.

In (7) 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.

(7) 'Expanding' algorithm

- Let the set of possible parses for Φ, $I = \{\Psi_i, i=1, \ldots\}$ be equal to Φ; $I = \{\Phi\}$
- Start at the highest ranked constraint;
- 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 along precisely the dimension specified by the W-constraint. I.e. add candidates Ψ_i to the hypothesis space 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 $I = \{\Psi_i, i=1, \ldots, k\}$ represents the set of URs which will be neutralized to Φ by the grammar.

We can illustrate the operation of the algorithm in (7) by contrasting the parsing of English [rat] vs. [rad] with that of the ambiguous German [rat], assuming the URs in (8). (We have chosen the algorithm in (7) for purely expositional purposes. The same result will be obtained using the alg o-rithm in (6).)

(8) Contrastive parsing

- English /rat/ 'rot' /rad/ 'rod'
- High German /rat/ 'advice' /rad/ 'wheel'.

Since English does not have a rule of coda devoicing, we can assume that the ranking of *VoicedCoda in English is the opposite of that assumed for German, above. The operation of the parsing algorithm is sketched in (9) where a single UR is associated with surface [rat].

(9) 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 (10).

(10) German parse of [rat]: *VoicedCoda >> Faith[voice]
The candidate set consists of /rat/
The candidate set is expanded to /rat/ **and** /rad/ by the W-constraint *VoicedCoda
Voicing specification of all segments in /rat/ **and** /rad/ fixed by Faith[voice]
The overt form is ambiguous—derivable from both /rat/ and /rad/

Whichever algorithm turns out to be more useful, it is obvious that either of our proposals is s u-perior to S's, since they generate a *set* of candidate URs for a given surface form. Note that the arg u-ment 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

4.1 The Initial Ranking of Faithfulness Constraints

We now turn to a consideration of the implications of having a parsing algorithm that works (i.e. generates a *set* of underlying forms) 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, however, 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 (11).

(11) At S_0: Faithfulness constraints >> Wellformedness constraints

With the initial ranking proposed in (11) there is a single outcome to each parse at S_0. With the initial ranking proposed by S in (1) a parsing algorithm like (6), which eliminates candidates from an initially infinite set, will generate the empty set; and one like (7), which adds candidates to an in i-tially unary set, will explode the candidate set to include all possible URs. A lexicon is unacquirable under either scenario.

The table in (12) 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 the assumption that all F-constraints are ranked high. Using either parsing algorithm, (6) or (7), the learner will be able to converge on a single UR for each surface form. Using (6), the high ranking of all F-constraints ensures that the optimal candidate is identical to the input form. Using (7), 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 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.

(12) Comparing HiFaith and LoFaith at S_0 using a parser that works

Surface form		Initial Hypothesis for UR	Path to adult UR
With F constraints ranked HIGH			
A.	rat	rat	Unique, correct UR is selected initially
B.	rad	rad	"
C.	rat	rat	"
D.	ratəs	rat	"
E.	rat	rat	E & F stored differently, later collapsed
F.	radəs	rad	by storing /rad/ and raising *Voiced-Coda
With F constraints ranked LOW			
A.	rat	∅ or rat, rat⊙s̥, bəbə…	There can be no learning path: each
B.	rad	∅ or rat, rat⊙s̥, bəbə…	production yields the maximally
C.	rat	∅ or rat, rat⊙s̥, bəbə…	unmarked utterance, say *ta,* as S
D.	ratəs	∅ or rat, rat⊙s̥, bəbə…	desires, but each parse yields ∅ by (6)
E.	rat	∅ or rat, rat⊙s̥, bəbə…	or else everything generated by the
F.	radəs	∅ or rat, rat⊙s̥, bəbə…	UG-given W-constraints by (7).

The bottom half of the table illustrates the problem with assuming that F-constraints are initially ranked low. As desired by S, the production mapping will generate the maximally unmarked [ta] at the initial state, but (6) will generate no parses—there is no UR which will surface as [rat] at this stage, since every UR will surface as [ta]—and (7) 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.

It is worth reiterating at this point that the reranking of constraints and the collapsing of predictable allomorphs to a single form are two aspects of a single process, despite the following suggestions to the contrary:

(13) 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 P & S 1993:§9 we find

(14) Prince and Smolensky 1993, 192

Lexicon Optimization. Suppose that several different inputs $I_1, I_2, \ldots, I_n$ when parsed by **a grammar G** [i.e. ranked constraint hierarchy—mrh&cr] lead to corresponding outputs $O_1, O_2, \ldots, O$, 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 'vicious circle' theory of language acquisition: the child needs a ranking to get URs and needs URs to get a ranking. To be fair, later work including S's model discussed above appears to address this problem, but fails for reasons we have pointed out. Constraint reranking and choice of UR are part of the same task. Using the algorithm we have proposed (which correctly produces sets of candidates in cases of surface ambiguity), the learner can converge on a lexicon **only** if F-constraints are initially ranked **above** W-constraints.

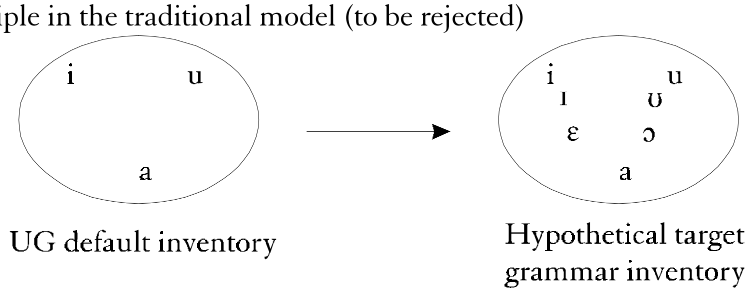
We now summarize the major points in our argument concerning parsing. First, S's parsing algorithm selects only **the most 'harmonic'** UR, so it fails to account for surface ambiguity in any human language. Second, an algorithm which associates a perceived form with a set of possible URs is needed, 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.

One conclusion we can draw from this is that the 'emergence of the unmarked' is irrelevant to the description of children's speech. Their grammars are faithful to the observed target forms.

4.2 The Subset Principle

One of the general principles known to govern the acquisition process is the so-called 'Subset Principle (SP).' We present here a radically different view of this principle, as applied to phonological acquisition, from that found in the literature (e.g., Calabrese 1988, 1995, Rice 1996). The traditional view, which we will reject, continues the same basic attitude towards children's speech output as expressed by Jakobson (1941). A standard version of this view is sketched in (15): the initial state of the grammar contains a limited number of vowels, that is, a single vowel or the three 'basic' vowels represented here; acquisition of a larger inventory leads to a superset of this initial inventory.

(15) The Subset Principle in the traditional model (to be rejected)



We consider the relevance of the SP to acquisition to be beyond question, once the assumption is made that children are not sensitive to negative evidence in the course of acquisition. This seems most plausible in phonology where explicit correction of pronunciation errors *vis-à-vis* the target form have long been known to be in vain:

> "...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 r e-spects, to have adult-like representations, which are reflected, among other things, in their vociferous rejections of adult imitations of their phonologically impoverished productions " (Faber and Best 1994: 266-7).

In other words, the SP can be viewed as a corollary to the acquisition principle of 'no negative ev i-dence'. The effect of the SP is to prevent the learner from making overly broad generalizations which cannot be corrected on the basis of negative evidence alone. We take the essence of the SP to be, therefore, a kind of restrictiveness. In other words, the initial state of the grammar, S_0, is maximally restrictive, and learning consists of relaxing restrictions. Our task, then, is to figure out how these r e-strictions are formulated (in terms of features, parameters, etc.).

Despite the fact that the SP was first formulated for phonology (Dell 1981) it has been more widely discussed in the syntactic acquisition literature, for example by Berwick (1986) and Wexler and Manzini (1987). Therefore, it may be useful to first review how the SP has been applied to a syntactic problem, as a leadup to our reinterpretation. Given our concerns, the discussion of syntactic phenomena will be informal.

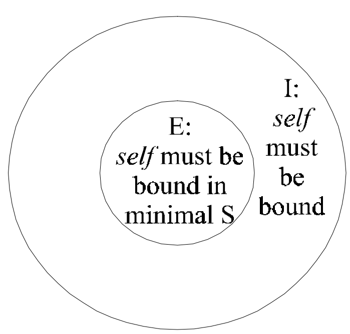
In a comparative study of acquisition of anaphora, Hyams and Sigurjónsdóttir (1990) compare the binding conditions on Icelandic *sig* and English *him/herself*. In simple terms, we can characterize the anaphors in the two languages as follows: Icelandic anaphors need to be bound; English an a-phors need to be bound in the minimal S. So, English is more restrictive, it imposes more conditions on anaphors than Icelandic does. The difference is represented by the schematic sentences in (16). In English, the anaphor can only be coreferential with the NP in the same clause, whereas in Icelandic, the anaphor can be coreferential with an antecedent in a higher clause.

(16) Anaphors in English and Icelandic

- English: John$_i$ asked Bill$_j$ to shave self$_{*i/j}$
- Icelandic: John$_i$ asked Bill$_j$ to shave self$_{i/j}$

We can represent the greater restrictiveness of English as in (17) and conclude that English corr e-sponds to the initial state (in this respect).

(17)



We can also represent the relationship of the two languages as an implicational relationship, as in

(18).

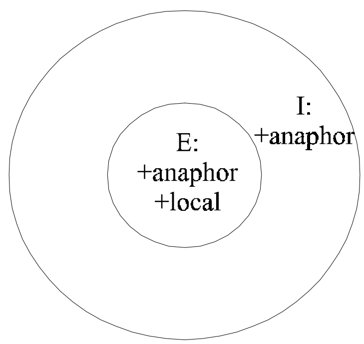
(18) The SP as an implicational hierarchy

a)Anaphor must be bound in the minimal S $\Rightarrow$ Anaphor must be bound.

b)Anaphor must be bound $\not\Rightarrow$ Anaphor must be bound in minimal S.

If we try to represent the distinction between the languages in terms of lexical features, instead of in terms of parameter settings, as has been done traditionally, we might propose the model in (19), where English anaphors are marked as [+bound, +local] whereas Icelandic anaphors are marked only as [+bound].

(19) Features for anaphors



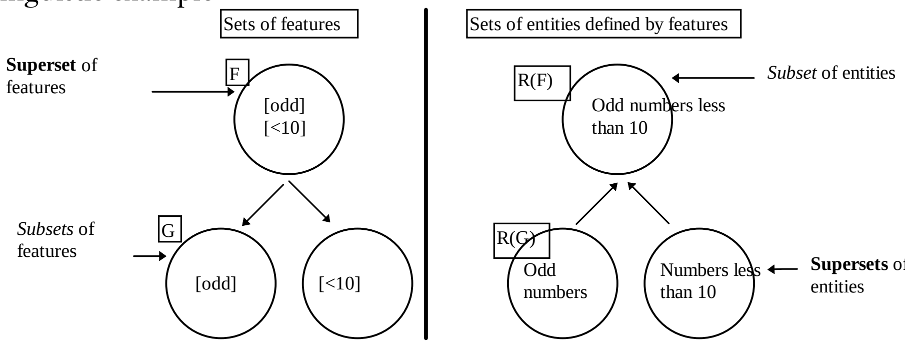
This brings us to a point, which though obvious, is crucial to our argument. Despite the simplicity of the argument, it it precisely the failure to grasp this point which has led to the misinterpretations of the SP which we will present below. The point can be formulaically stated as: fewer features equals more entities. That is, the size of a class varies inversely with the number of features used to define the class. This is stated more formally in (20).

(20) Fewer features = more entities

Let F and G be sets of features such that R(F) is the set of entities defined by F and R(G) is the set of entities defined by G. If G is a subset of F, then R(F) is a subset of R(G). That is $F \supset G \Leftrightarrow R(G) \supset R(F)$.

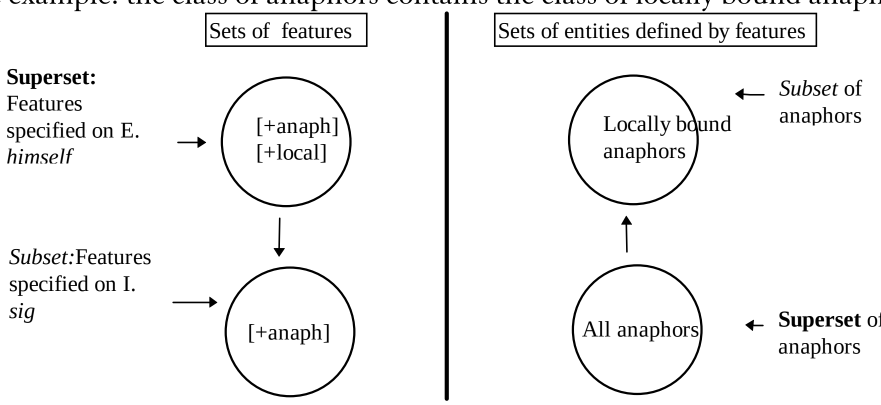
At the risk of appearing pedantic, we now present a non-linguistic example of this principle. The properties of being 'odd' and being 'less than 10' can be used to characterize, positively or negatively, subsets of the set of whole positive numbers. As shown on the left-hand side of (21), the set of properties, or features, containing both 'odd' and '$<$ 10' contains the two sets which contain only one of these features. On the right-hand side, however, we see that the containment relation goes in the other direction: the set of numbers which are both odd and less than 10 is contained within the set of odd numbers and within the set of numbers less than 10.

(21) A non-linguistic example



We can now return to our linguistic example and see that the same inverse relation holds. On the left-hand side of (22) we see a superset of features containing a subset of features, but on the right hand side we see that the interpretations associated with anaphors are in the inverse relationship.

(22) Linguistic example: the class of anaphors contains the class of locally bound anaphors:



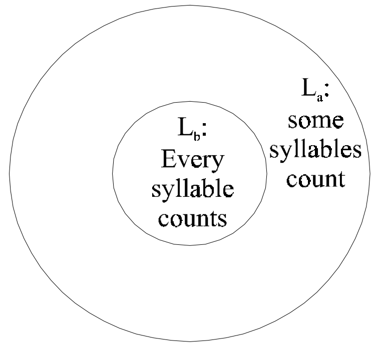
We now turn to discussion of some phonological cases from the literature, both concerning the acquisition of metrical phonology. Archibald 1995 proposes that the SP is not relevant to the acquis i-tion of English extrametr icality.

(23) Archibald 1995 on the Subset Principle:

> the two possible settings of the extrametricality parameter [+/- that is—mrh & cr] do not yield languages that are in a subset relation to each other. If English has no e x-trametricality ([-extrametrical]) we would always find penultimate stress (assuming left-headed feet). If it had obligatory extrametricality ([+extrametrical]) we would always find antepenultimate stress. The situation is a little more complex, as we have extrametricality in English that is sensitive to the grammatical category (i.e. it works differently on nouns and verbs, for instance). Thus, the Subset Principle makes no clear predictions.(86)

We propose an alternate view, namely that the initial hypothesis made by the child, the initial setting of an 'extrametricality parameter', is that **every syllable must be considered for the comput a-tion of stress**. The initial grammar, then, is **more restrictive** than any subsequent grammar which may, as in the case of English, allow certain syllables to be left out of stress computation (either based on independent computations such as foot construction or through lexical marking). This initial la n-guage, or UG, is thus true to the essence of the SP. In (24) then, L_b corresponds to the initial state. Again, we can state the relationship between S_0 and later states as an implicational one.

(24) The SP and extrametricality



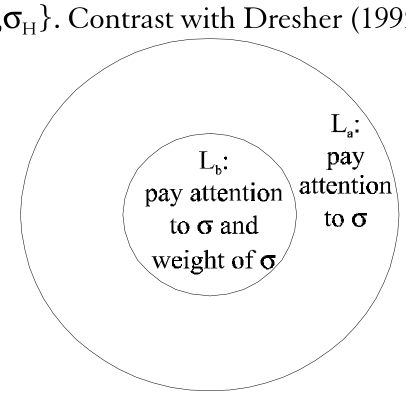
a) Every syllable counts ⇒ Some syllables count.
b) Some syllables count ⇏ Every syllable counts.

Our next example comes from Dresher 1995 who assumes that sensitivity to syllable weight must be acquired. Dresher's claim is sketched in (25), where σ_H is a heavy syllable and σ_L is a light syllable.

(25) Sensitivity to syllable weight must be acquired, so a language which does not 'care about weight' is the initial state.

In our view this is backwards, since attending to weight restricts foot construction. Children must be innately sensitive to syllable weight. If children did not have this sensitivity, then positive evidence would never provide it to them, and weight sensitive systems would be unlearnable. In learning a language in which syllable weight does not matter, the child **learns to ignore** a distinction which must be possible in UG. We assume that the subset relation must be the contrary of what Dresher proposes. This means that L_b actually represents the initial state of UG: L_a:$\{\sigma\} \supset L_b$:$\{\sigma_L,\sigma_H\}$, so L_b is the initial state, as shown in (26), along with another implicational statement.

(26)Our proposal: L_a:$\{\sigma\} \supset L_b$:$\{\sigma_L,\sigma_H\}$. Contrast with Dresher (1995): L_a:$\{\sigma\} \subset L_b$:$\{\sigma_L,\sigma_H\}$

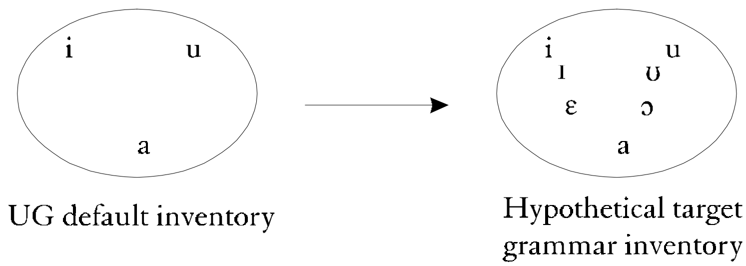


a) Paying attention to syllables and their weight ⇒ paying attention to syllables.
b) Paying attention to syllables ⇏ paying attention to syllables and their weight.

4.3 SP and segment ‘inventories’

We now return to the acquisition of phonological inventories. One of the most explicit versions of the traditional model is presented by Rice 1996, sketched in (27).

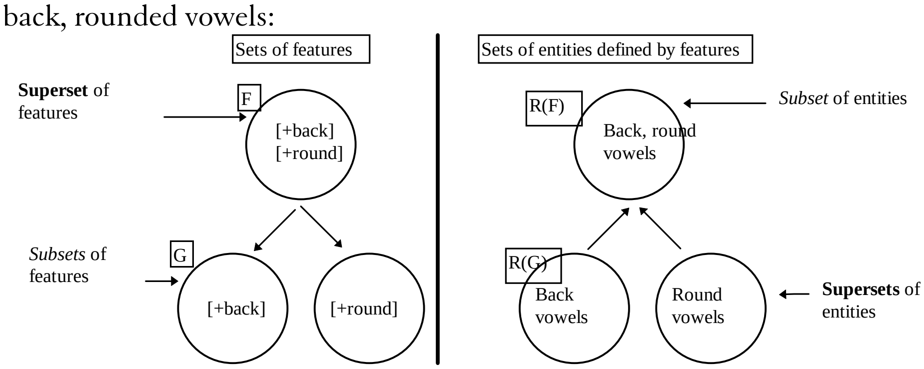
(27) The Subset Principle in the traditional model (rejected)
Central hypotheses (Rice 1996):
Minimality: Initially the child has minimal structure (not all features are available.)
Monotonicity: Inventories are built up in a monotonic [stepwise—mrh&cr] fashion as contrasts are added.



We object to this hypothesis on two grounds. First, there is no reason to expect ‘segments’ to play a role in the learning path, since features are the primitives of phonological theory. Second, the inverse relationship of features and natural classes discussed above leads to an alternative interpretation.

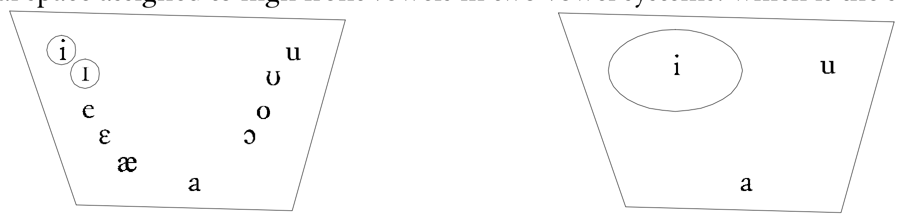
(28) Objections

- Languages (grammars) do not have /i/, /a/, /k/, etc. The “atoms” of phonology are features (cf. the *wug* test, etc).
- Fewer features = more things. The class of back vowels contains the class of back, rounded vowels:



With these objections in mind, consider the ‘phonological space’ associated with vowels in two languages, one with a rich inventory and one with a restricted inventory, shown in (29). The direction of the subset/superset relationship is not so clear when faced with two ways of looking at the problem: i) numbers of ‘segments’ and ii) phonological space.

(29) Phonological space assigned to high front vowels in two vowel systems: which is the subset?



The arguments we have offered to this point favor choosing the language with more restrictive, i.e. richer, representations and narrower phonological space associated with individual vowels as the initial state. In order to provide learnability arguments to support this proposal and justify rejecting the traditional theory, we must answer the two questions in (30). Below we provide arguments using hypothetical languages to justify the answers we provide.

(30) The tests:
a. Can the traditional view lead to a growing inventory? NO.
b. Can the proposed view lead to a shrinking inventory? YES.

In order to answer (30a) consider the acquisition of /dip/ vs. /dɪp/ in a hypothetical language which maintains the [i] / [ɪ] contrast on the surface. In the traditional system, the contrast is unlearnable, the two words will be acquired as homophones. Without access to a difference in representation, the difference between the two vowels cannot be evaluated. The so-called 'positive evidence' often invoked to allow inventory expansion is not sufficient if that evidence cannot be assigned a representation. That is, the contrast cannot be perceived linguistically if the child does not have it. This is a fundamental assumption of linguistic theory. It is equivalent to saying that a language that uses a non-human feature is unlearnable, which is tautologically true given the standard definition of UG. If a child did not have access to a feature provided by UG, then the child could not store this distinction for future use; each lexical entry would have to be relearned at each stage since each lexical entry could potentially contain the newly 'acquired' feature. [4] This is contraindicated by the acquisition evidence. On the other hand, if the distinction is available at S_0 then acquisition of contrastive lexical items is trivial. Our claim is consistent with the evidence from comprehension and perception studies (Streeter 1976, Goodman and Nusbaum 1994) and thus should represent the null hypothesis.

We now turn to (30b), loss of a 'wrong' contrast, i.e. /dip/ and /dɪp/ collapse to /dip/ in some language with a three vowel system. The traditional view will never face a problem here, since the grammar never contains more contrasts than the target language. The challenge to the theory proposed can be stated thus: How does a grammar which has more potential vowels than the target grammar end up losing irrelevant contrasts? Two cases must be distinguished.

[4] It is sometimes suggested in the literature (e.g. Ingram 1995:75) that the child's grammar makes use of two kinds of representation, a phonological representation, which starts out with access to a minimal set of features, and a phonetic or acoustic representation which makes use of fully specified phonetic feature matrices. One might imagine that the child stores contrasts in the phonetic representations until the phonology is 'ready' for them, at which time previously identical phonological representations can be distinguished. This view suffers at least two difficulties: 1) the desired 'simplicity' or 'poverty' of the child's grammar is not attained, but rather just holds at one level of the grammar; 2) the standard view of the relationship of phonological and phonetic representations is that the latter are not stored in memory, but are derived from the former by the phonology.

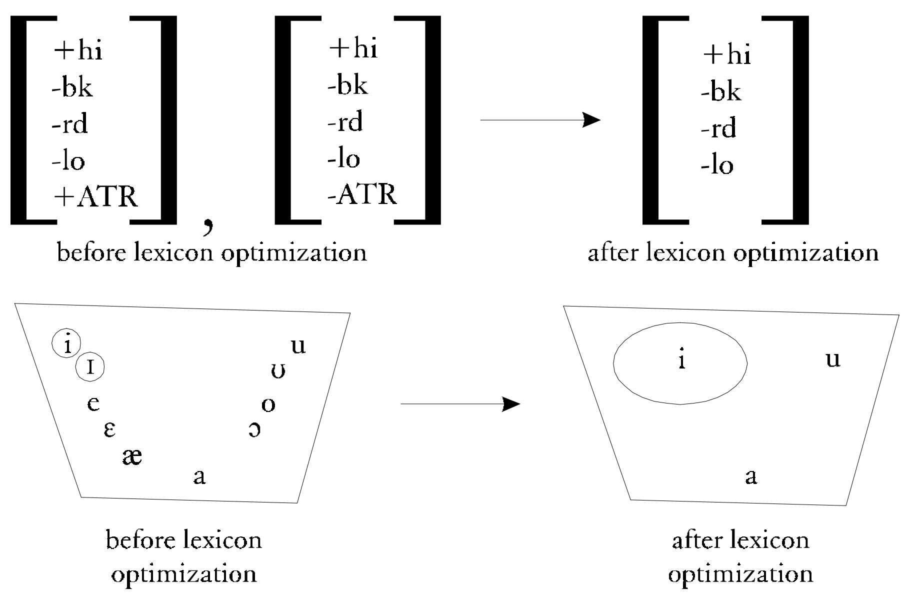
(31) Two distinct cases for the ‘collapse’ of contrast

a. Unobserved contrasts: If the target language does not present forms such as [dɪp], then there will never be any reason to remove [+ATR] from the representation of /dip/. Access to the universal feature set allows the *potential* for any contrast, not its realization. This is clearly relevant to underspecification theory in that we assume, for instance, that /i/ will never lose its [+ATR] specification without grammar internal motivation (see below).

b. Phonetic underspecification: Imagine the child hears [dip] and stores it as such. Since this child has access to all the features and since its learning conforms to the SP, it assumes that representations must be maximally restrictive (specified). This word cannot be stored with just a [-back, +hi] vowel; it must be stored as a [-back, +hi, +ATR, -round] vowel. Given the variability of articulation in a three vowel system, this child will hear phonetic [dɪp] and mistakenly posit a new lexical item, ending up with a pair of synonyms, /dip/ and /dɪp/. A process of lexicon optimization, responsible for collapsing synonyms, will determine that the feature is not relevant to the phonology of the language and thus can be excluded from representation. This proposal is very similar to the uniqueness principle of morphology (Wexler and Culicover 1980). (This algorithm does not affect all redundant features, merely those which behave as though they are *phonetically* underspecified.)

Note, at this point that the confusion is largely notational. In losing the /i/-/ɪ/ contrast the grammar moves from containing two vowels [-back, +hi, +ATR, -round] and [-back, +hi, -ATR, -round], which we happen to denote as /i/ and /ɪ/, to one [-back, +hi, -round] which we somewhat arbitrarily denote as /i/.

(32)



Clearly, this account needs to be further developed by an explicit model of lexicon optimization. Crucial questions include the nature of optimization: is it a global process which affects the whole lexicon at once?

4.4 Summary

We can summarize the argument to this point in the following way:

(33) Summary of arguments
A. Subset Principle Argument
a. The Subset Principle reflects restrictiveness in the initial state
b. Greater restrictiveness is encoded through fuller specification
∴ All features must be available for representations at S_0.
B. Learnability Argument
a. Linguistic representations contain features
b. If a feature F is unavailable at stage L_j, then positive evidence of F cannot be evaluated by the learner since the learner cannot evaluate representations with respect to F
∴ All features must be available for representations at S_0.

Since acquisition involves real-world performance in both production and comprehension, our arguments represent a useful idealization. Children will fail to acquire adult representations in some cases; the explanation for this is not to be sought in an impoverished grammar, but rather in performance factors (for further discussion, see below).

4.5 Digression on Underspecification and the SP

The approach to underspecification advocated in this paper is consistent with recent work by Sharon Inkelas and her collaborators (Inkelas 1994, Inkelas, Orhun and Zoll 1996) in the adoption of the view that underspecification is derived from the data, not imposed by principle. This position, exemplified by (34) leads Inkelas to reject philosophically based approaches to underspecification such as those represented by the (partially overlapping) categories in (35).

(34) Inkelas (1994:1) on Underspecification

> underlying representation is determined solely by optimization with respect to the grammar, not by imposing any type of constraints directly on underlying representation...[this] results in the use of underspecification only when there are alternant surface forms... (Inkelas 1994:1).

(35) ‘Grammar-blind’ approaches to underspecification rejected by Inkelas 1994, q.v. for references

- **Markedness** (universal, language-specific, or contextual); unmarked material is underspecified
- **Redundancy**; redundant feature values (determined on the basis of the segment inventory) are underspecified
- **Predictability**: predictable material is underspecified

According to Inkelas, “[t]he only motivation for underspecification is to capture alternations in the optimal way” (1994: 2). A very convincing case where underspecification can be used in an illuminating fashion comes from Inkelas’ discussion of the distribution of Turkish voiced and voiceless stops. As the data in (36) show, the two Turkish surface stops [t] and [d] show up in three different patterns.

(36) Turkish voicing alternations

a. Alternating: [0voice] (unmarked for [voice])

ḳanat ‘wing’ *ḳanatlar* ‘wing-plural’ *ḳanadIm* ‘wing-1sg.poss’

b. Non-alternating voiceless: [-voice]

sanat ‘art’ *sanatlar* ‘art-plural’ *sanatIm* ‘art-1sg.poss’

c. Non-alternating voiced: [+voice]

etüd ‘etude’ *etüdler* ‘etude-plural’ *etüdüm* ‘etude-1sg.poss’

In (a), the stop is voiceless in coda position and voiced in onset position. Inkelas proposes that this stop is underlyingly unspecified for [voice] and it receives its specification by the OT equivalent of structure-filling rules.[5] The (b) forms are stable in always showing voiceless stops and thus are underlyingly specified as voiceless. The (c) forms are consistently voiced, and thus are underlyingly marked as voiced. Inkelas, Orgun and Zoll 1996 argue convincingly that this is the best account of this type of data, rejecting, for example, the use of lexical exception features.

While we accept fully the spirit of this argument, we find it necessary to expand the range of underspecification in an additional way. This is the case of phonetic underspecification, whereby the grammar just ‘doesn’t care’ about the setting of a certain feature. This was discussed for a hypothetical case above, where the target space for the high, front vowel in a three vowel language included both [+/-ATR] regions. The existence of such articulatory freedom in vowel articulation is an empirical issue supported by some phonetic studies of relevant languages (Manuel 1987, but see Maddieson and Wright 1995 for a potential case of a three vowel system with highly restrictive target spaces). If it turns out, however, that such variation is in fact contextually determined, then these cases of apparent ‘phonetic underspecification’ will reduce to conditioned allophony.

As a concrete example of how this approach differs from standard views on under specification consider the acquisition of the lexical representation of the /n/ in English /tɛn/ ‘ten’. Since the child starts out with full specification and removes specification only when ‘forced’ to by the grammar, representations may be grossly ‘overspecified’ (even in adult grammar). There is no motivation to leave this segment unspecified for [voice], since no alternations exist which bear on the issue, and the ambient language has voiced nasals exclusively. The theory explicitly rejects the grammar-blind principles in (20), not least on account of the severe empirical difficulties they face, as discussed by Inkelas and her references (e.g., Steriade 1994). The dental/nondental specification, on the other hand, must be dealt with in order to allow the allomorphs in [tɛn] ‘ten’ and [tɛn̪θ] ‘tenth’ to be derived from a single UR. As above (section 4.3) we must leave several crucial issues for further research. One concerns the process of lexicon optimization which accomplishes the collapse of the allomorphs in ‘ten’ and ‘tenth’. Does this proceed on a lexeme-by-lexeme basis, or is the whole lexicon scanned *en masse*? Second, what is the result of this collapsing in a given case? Does it result in specification of one value, say non-dental, and a feature changing rule (or OT equivalent) or does the dental/nondental feature specification get erased from every /n/ and get reassigned by feature-filling rules. Without going into a full discussion, we prefer to reserve feature filling operations for cases such as the Turkish voicing alternations discussed by Inkelas, that is, cases where the grammar

[5] Inkelas does not actually show how this would work in OT. For discussion, see Reiss (to appear) which makes the following point: “Since representations containing less structure are usually assumed to be ‘less marked’ than more complex structures, it is difficult to see how an underspecified representation could violate more well-formedness constraints (W-constraints) than a candidate representation which has ‘filled in’ underlying gaps (and is thus also less faithful to the input).”

requires them. A third question, related to the others, concerns the scope of 'optimization': does it affect only those representations which participate in alternations, or does the lexicon get 'optimized' according to the alternations that exist anywhere in the language? For example, what happens to the dental nasal of 'plinth', a non alternating morpheme?

Yip 1996 reaches basically the same conclusion as Inkelas 1994, both of which are compatible with the theory of rich initial representations espoused here:

> (22) This paper finds inconclusive evidence for abstract underlying representations, and concludes that the balance of the evidence suggests that learners acquire something rather close to what they hear, unless information from alternations or paradigms forces them to do otherwise (Yip 1996).

Both Yip and Inkelas argue that their conclusions follow from the nature of Optimality Theory. In our view, a principled account of acquisition and underspecification can be accomplished in a variety of frameworks.

5. Accounting for the Divergence in Comprehension and Production

It is apparent that 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. The phonological acquisition literature seems virtually unanimous in attributing this disparity to the grammar, however. 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, this is not surprising. However, we believe it is misleading. We hope to show in what follows that the empirical evidence gathered thus far on errors made by children does not support appealing to natural language grammatical systems but is instead the result of extra-grammatical factors, broadly categorized as 'performance.'

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.[6] One of the most widely discussed phenomena is 'consonant harmony', responsible, for example, for the realization of 'duck' as 'guck'. Another, less widely discussed, 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

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

current Optimality Theory. However, this appears to be a rather opportunistic appeal to 'markedness'. If 'markedness' is interpreted as 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,[7] 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 grammatical 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 among their list of observed effects

(34) 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 non-application 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 consumption of alcohol in sufficient quantities leads to constraint re-ranking or rule addition in adult grammars. 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. It is an empirical question, whether or not in each case the output which forms the basis for the grammars proposed in such work represents output of the grammar or output of the 'body'.

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 [pʰɛ̃n] 'pen' collected from a 15 month old child in a thirty minute period: [mãᵊ], [ṽ], [dɛᵈⁿ̩], [hɪn], [ᵐbõ], [pʰɪn], [tʰn̩tʰn̩tʰn̩], [baʰ], [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[8]. *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).

---

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

[8] 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."

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 (Stathopoulos and Sapienza 1993, cited by Faber and Best 1994:269). Occam's razor demands attributing aspects of their speech output to these factors as the null hypothesis.

5.2 Chainshifts & Imitation

The well-known phenomenon of chainshifts in the mapping of adult speech to child speech is something that S claims argues for a competence-only account of child phonology. An example cited by S is 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 an impression that the sound the child made sounds a lot like the adults own [θ]. Finally, S appears to be arguing against a non-existent straw-man for which no references are given. As S states, a hypothesis "invoking severe performance difficulties to account for the impoverishment of production relative to comprehension, has several problems. Gross formulations 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. S does not refer to any non-gross formulations of this hypothesis, nor to any alternate hypotheses such as that given by Hale and Reiss (1995), who argue against nearly every claim S makes. We characterized the merger of [š] and [s] in some children's output as follows:

(35) Production account of merger in child speech (Hale and Reiss 1995)

> The failure on the part of the child to distinguish between the two segments in his/her output is attributed to a shortcoming of the production [i.e. performance] system, which responds to the instruction to produce a [š] by emitting a [s] instead. 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.

Clearly there is more intervening between the grammar and the utterance than just commands to the vocal tract. The merger of the two sibilants could be attributed to any intervening cognitive or motor process. Note that the claim is not that children are physically incapable of moving their mouths to make a [š]-like sound, but rather that, when the performance system is given (by the grammar) the command to make a [š], the vocal tract generates a sound like [s]. Given the commands to make a [$t^{\check{s}}$], the vocal tract may produce something like a [š], resulting in an apparent chainshift (e.g. *chew* pronounced with a [š]-like sound and *shoe* with a [s]-like sound).[9]

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/.

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."

Two of the best known attempts to deal with opacity in OT are McCarthy's "Remarks on Phonological Opacity on Optimality Theory" (1994) and Kirchner's work on chainshifts (1996). McCarthy adopts the use of parameterized constraints to account for opacity effects and suggests that the default state (i.e. initial state given by UG) for each constraint is one in which the parameters are set so as to minimize opacity. Kirchner's proposal also includes a radical modification to the original OT idea of a universal, innate set of constraints—constraints can be conjoined to generate complex constraints. Both of these suggestions rely on positive evidence and are treated as learned aspects of OT grammars. For example, McCarthy (1994, 6) says, "I will stipulate that the default form of a phonological constraint—the form in which it is represented in UG—has all of its level specific actions set to 'surface'" . Therefore, these solutions to the description of opaque systems cannot be applied to the chainshifts discussed by S since the target language provides no evidence for such shifts. In the absence of an alternative, chainshifts such as those cited by S constitute a major challenge to his own proposals. Chainshifts are not a problem for our theory because they do not require the grammar to produce the chainshifts — the child speech chainshifts are the result of performance effects.

It is also worth noting in passing that both McCarthy's and Kirchner's proposals involve major adjustments to OT, whereas opacity effects in rule-based grammar are accounted for trivially. For example, McCarthy's decision to allow parametrized constraints in Optimality Theory compromises one of the signature distinctions of OT from Principles and Parameter models. The radical nature of this departure from the original OT notion of a fixed universal constraint set is not fully recognized in the literature, even by McCarthy himself, as illustrated by the following statement: "The constraints themselves are universal, except for the fixing of particular arguments within general constraint schemata; only the ranking is language particular" (McCarthy 1994, 2). Clearly, fixing of arguments *and* constraint ranking are both language particular in McCarthy's model.

Note that positing the 'marked' (in McCarthy's sense) setting of constraints to generate a chainshift in child speech requires changing the default setting of the parameters to produce a grammar that the target language provides no evidence for, in the case of the English example S cites. Becom-

[9] Note that, when we write [s], we are writing a phonetic symbol—that is a linguistic entity which is output by the grammar. This is not to be confused with a physiological act which happens to look like 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.

ing a competent adult speaker then requires 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.

The higher level of performance during direct imitation that S cites from Menn and Mathei (1992) as further evidence for his model is actually 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.

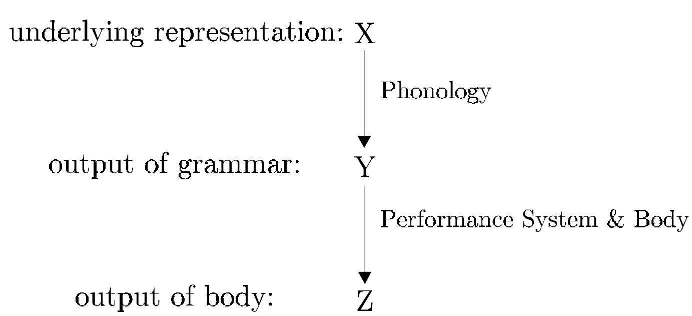
Under S's competence-only approach, neither of these types of "imitation" can be accounted for. In the first type, since S assumes the *grammar* is responsible for, e.g., realization of [š] as [s], increased attention to performance should lead only to a clearer hit of the target [s]. In the second case, to account for an English speakers' ability to imitate Cree, S would have to assume instantaneous acquisition of Cree constraint rankings.

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. 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 demonstrations that (1) if one adopts an OT framework at all, the initial state of the must have all F-constraints ranked above all W-constraints in order to allow for the acquisition of a lexicon and (2) children 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 (36) below.

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



The OT phonology (for ease of exposition we will assume a single level of phonology — i.e., a single OT) represents a mapping relationship between the underlying representation, X, and the output of the phonology for the string in question, Y (where Y may be equal to X if the relevant F-constraints outrank the potentially relevant W-constraints). The performance system of the adult, 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 cavity, 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). 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.

Each of the arrows in (36) 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.'[10]

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 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 (36) were more dramatic than the effects of the adult production system are. 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

[10] 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.

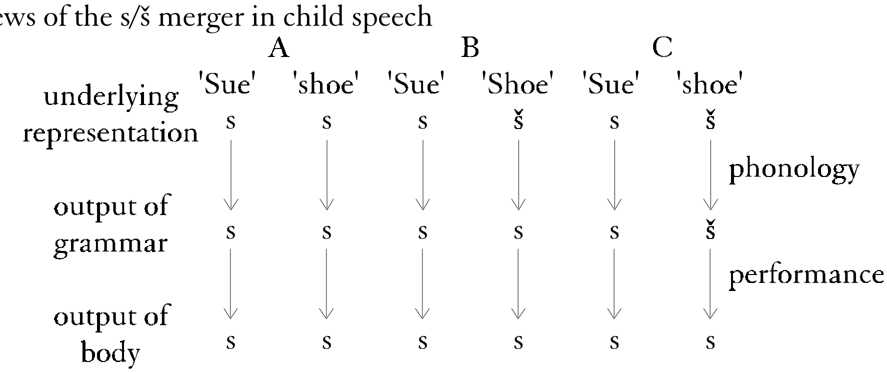
phonological computational system thereby recovering an underlying representation, X, (or, in the case of ambiguous strings, a set of underlying representations) which would be expected, given the listener's grammar, to generate Y. The lexical item(s) corresponding to this X is (are) then accessed and 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, clearly involves fairly 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 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 [k^hæt] 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 — the grammar output 'augmented' by the effects of his own body and its context of utterance — but rather to the output of his 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 performance system. Since 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), 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 may be unable, without this additional evidence, to reveal.[11]

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 'shoe' and 'Sue' as [su]. There are three likely scenarios which could account for this phenomenon, sketched in (37A-C).

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



Under the scenario in (37A), 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

[11] We assume that performance factors play a role in comprehension as well, but that they are much less dramatic than those affecting production, and thus we ignore them here.

the phonology[12] 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 does the child no good in the acquisition of the grammar, for it is not exploited in the construction of underlying representations or constraint ranking. 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.[13] While the lexicon will contain more homophony than the adult lexicon, homophony must be permitted (given *night* and *knight*), so this 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 1995). However, it cannot hold in our case, which assumes that the child eventually converges upon the adult output. Since hypothesis A essentially precludes acquisition of the s/š contrast, it must be rejected.

Hypothesis 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 (38), the correct output (for the child speaker) will be generated by such a scenario.[14]

(38) 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.[15]

Under hypothesis (C) in (37) 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

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

[13] As we have shown above, this is due to the Subset Principle, properly construed.

[14] 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.

[15] Structural ambiguity (37B) is distinct from lexical ambiguity (37A), where identity in input structures is responsible for the ambiguity of the output, as well as from production ambiguity (37C), where forms which are distinct at the level of the output of the grammar are, due to some difficulty with the performance system, produced the same.

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[16]. 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.

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 the parsing model discussed above, 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.[17]

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 *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 (37B) 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 flawed). 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

We will argue in this section, that a compelling case can be made for the assumption that all F-constraints are ranked, in UG, above all W-constraints.[18]

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 well-formedness con-

[16] 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 children's grammars contain more highly marked representations than those of the target language in a great many instances.

[17] See the quotation from Faber and Best (1994) in Section 4.2 above.

[18] 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.

straints (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. 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 central to this issue: (1) the reduction of the amount of information stored for a given segment because of distributional restrictions on the segment in question which render some phonetic features of the grammar-output representation of that segment predictable (e.g., in the case of allophones); and (2) the reduction of the size of the lexicon by linking via phonological processes morphologically-related (but possibly representationally-distinct, before the linking) forms.[19] 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. 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.

The Subset Principle requires that children be innately sensitive to all possible phonological contrasts, as dictated by every child's ability to learn any human language, regardless of which subset of the universal feature inventory is used in that language, or what the distribution of those features in that language is. 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), for 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. A child who hears [kʰæt], posits this form as the target output for his or her grammar, and stores this as his or her underlying form has — through that act — posited high-ranking for faithfulness to all the features in [kʰæt].[20] Put another way, only if UG had high-ranking for F-constraints could the child, using the computational system of UG as its sole guide to the relationship between target and input form, initially posit /kʰæt/ as the underlying form for a perceived target of the shape [kʰæt].

Given a sufficient number of forms stored in this manner, the process of lexicon 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 on [kʰæt] represents lexically redundant information in that it is predictable. It will therefore no longer be necessary to store such aspiration 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 OT 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 addition of aspiration 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 lexicon optimization procedure has sufficiently clear evidence that it can posit morphophonemic

---

[19] We 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 stimulus by humans.

[20] Additional phonetic detail to which the child certainly may attend is left out here for ease of exposition.

alternations of the normal type.

Under this conception of acquisition, F-constraints must be ranked high within UG. The elevation of W-constraints takes place as a result of lexicon 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 we assume that our theory 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 (which we attribute to extralinguistic cognitive and physiological maturational development) on which S' approach is predicated. 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 increased influence of extragrammmatical factors, but it is necessary (cf. Epstein, Flynn and Martohardjono (to appear) 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 innately programmed 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 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 performance-related facts 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 (39):

(39) 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 argued against points a-d, providing new arguments based on evidence which is well-known in the linguistic, psycholinguistic and phonetic literature. After sketching some basic assumptions about the nature of phonological acquisition and its study, we showed that the evidence actually supports the following claims:

(40) 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. Our own analysis is set in an OT framework, but this is mainly to facilitate comparison with S’s paper.

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 accompanied by a coherent theory of grammar which is consistent with fundamental assumptions of the field.

References

Archibald, John. 1995. *Phonological Acquisition and Phonological Theory.* Hillsdale, NJ: Lawrence Erlbaum.

Barton, David. 1978. *The role of perception in the acquisition of phonology*. Bloomington, IN: IULC.

Berwick, Robert. 1986. The acquisition of syntactic knowledge. MIT Press, Cambridge, MA.

Calabrese, Andrea (1988) *Towards a theory of phonological alphabets.*Ph.D. dissertation, MIT.

Dell, François (1981) "On the learnability of optional phonological rules," *Linguistic Inquiry* 12, 31-37.

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.

Dresher, Elan. 1996. Charting the learning path: Cues to Parameter Setting. Talk presented at *MOT*, McGill University. Ms. University of Toronto.

Epstein, Samuel David, Suzanne Flynn and Gita Martohardjono. To appear. Second Language A c-quisition: Theoretical and Experimental Issues in Contemporary Research. *Behavioral and Brain Sciences*.

Faber, Alice and Catherine T. Best. 1994. The perceptual infrastructure of early phonological deve l-opment. 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. 1995. *Theory and Method in Historical Linguistics*. ms., Concordia University.

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

Hale, M. and C. Reiss. 1996a. How to Parse (and How Not to) in OT Phonology. Paper presented at NELS 27, McGill University. Ms. Concordia Unive rsity.

Hale, M. and C. Reiss. 1996b. The Comprehension/Production Dilemma in Child Language: A R e-sponse to Smole nsky. *ROA*.

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

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

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

Hyams, N. and S. Sigurjónsdóttir (1990) "The development of 'long distance anaphora,': a cross-linguistic study with special reference to Icelandic," *Language Acquisition* 1, 57-93.

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).

Ingram, David. 1995. The Acquisition of Negative Constraints, the OCP, and Underspecified Repr e-sentations. In Archibald.

Inkelas, Sharon. 1994. The Consequences of Optimization for Underspecification. Ms. UC Berkeley. ROA.

Jakobson, Roman. 1941. Kindersprache, Aphasie und allgemeine Lautgesetze. Selected Writings, vol. 1, 328-401. The Hague: Mouton, 1971.

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

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

Kirchner, Robert. 1995. 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.

Manuel, Sharon. 1987. *Acoustic and Perceptual Consequences of Vowel-to-Vowel Coarticulation in three Bantu Languages.* PhD. dissertation, Yale University, New Haven.

Maddieson, Ian and Richard Wright. 1995.The vowels and consonants of Amis — a preliminary phonetic report. In *Fieldwork Studies of Targeted Languages III: UCLA Working Papers in Phonetics, Vol 91*. Department of Linguistics, UCLA.

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

McCarthy, J. 1994. Remarks on Phonological Opacity in O T. 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, MIT Press, Cambridge, MA.

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 phonlogy: 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 Gra m-mar. Technical Report Rutgers Center for Cognitive Science, Rutgers University, New Brun s-wick, N.J.

Reiss, Charles. (to appear). Unifying the Interpretation of Structural Descriptions. Proceedings of *WCCFL XIV*. CSLI Publications/Cambridge University Press.

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

Reiss, Charles. 1995b. Deriving an implicational universal in Optimality Theory. *Proceedings of NELS 26*.

Rice, Keren. 1996. Phonological Variability in Language Acquisition: A Representational Account.

Ms. University of Toronto.
Rice, Keren and Peter Avery. 1995. A Theory of Segmental Elaboration. In Archibald.
Smolensky, Paul. 1996a. On the comprehension/production dilemma in child language. *LI* 27.
Smolensky, Paul. 1996b. The Richness of the Base. *ROA*.
Stathopoulos and Sapienza 1993
Steriade, Donca 1994. Underspecification and Markedness. In J. Goldsmith (ed.) *A Handbook of Phonological Theory*. Cambridge: Blackwell.
Streeter, L. A. 1976. Kikuyu labial and apical stop discrimination. *Journal of phonetics* 4:43-9.
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].
Wexler, Kenneth and Peter Culicover. 1980. *Formal Principles of Language Acquisition*. Cambridge, MA: MIT Press.
Wexler, Kenneth and M. Rita Manzini. 1987. Parameters and Learnability in Binding Theory. P a-rameter Setting, ed. by Thomas Roeper and Edwin Williams, 41-76. Reidel, Dordrecht.
Yip, Moira. 1996. Lexicon Optimization in Languages without Alternations. Ms. University of Cal i-fornia, Irvine. ROA.