The Subset Principle in Phonology: Why the *tabula* can't be *rasa*[*]

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

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

June 7, 2007

1 Introduction

This paper is an explication in terms of phonological theory of the following quotation: "In any computational theory, 'learning' can consist only of creating novel combinations of primitives already innately available" (Jackendoff 1990:40). We refer to this position as the *Innateness of Primitives Principle* (*IofPP*).[1] We will demonstrate the logical necessity of *IofPP* with a set of model languages, then show how most work on the acquisition of phonological inventories is inconsistent with *IofPP*. We then propose an alternate theory which *is* consistent with *IofPP*, which, we assume lies at the core of the Innateness Hypothesis.

The chain of reasoning we will pursue in explicating the *IofPP* can be summarized as follows. Intelligence (by which we intend merely cognition) consists of the construction and manipulation of symbolic representations. Interacting intelligently with the world requires the ability to *parse* input (assign it a representation). Learning is a form of intelligent interaction with the world, thus learning requires parsing inputs into representations. Without an innate set of representational primitives, learning cannot begin.

We demonstrate further, that our theory, unlike traditional views concerning the acquisition of phonological inventories, is consistent with the Subset Principle (SP), properly defined. We will argue that the Subset Principle, too, is a logical necessity for linguistic theory. We propose that the SP must be reconceptualized somewhat, for both phonology and syntax. In brief, the SP is to be defined over stored representations, including the representations that constitute the structural description of rules, and not over grammars or sets of sentences. The essence of the SP, as we restate it is that representations at an earlier stage of acquisition must be *more highly specified* than those at a later stage. Because of the inverse relationship between the number of features used to define a set of representations and the number of members in the set, our view of the SP will sometimes be in direct conflict with pretheoretical intuitions, as well as with the suggestions of other scholars. In the next section, therefore, we illustrate the basics of our approach using toy grammars in potentially excruciating detail. The reader's indulgence in this section, will, we hope, be repaid when we turn to real linguistic examples.

[*]To appear in the *Journal of Linguistics, 2003.* The authors are indebted to Madelyn Kissock, Ida Toivonen, Afton Lewis and two anonymous reviewers for comments on an earlier draft on this paper. Our research was supported in part through the VRQ funded project *L*e traitement de langue naturelle and the SSHRC-funded *A*symmetry Project, as well as several grants to the authors from Concordia University.

[1]This position has been formulated and defended most eloquently by Jerry Fodor (e.g. 1975), however, the earliest explicit formulation we are aware of is in Pylyshyn (1973).

2 Let's play cards

In this section we illustrate the logic behind Jackendoff's statement by using model languages consisting of sets of playing cards from a normal deck. In our analogy, cards correspond to sentences of natural languages. From our point of view as observers, a card c will be grammatical, ungrammatical or neither to a 'speaker' of G. The reason for the third possibility will become clear below. We further assume that learners of these 'card languages' are endowed with an innate card language faculty. We will explore the effect of tinkering with 'card UG' below. In general, UG will consist of types of symbols and logical operators for symbols. Our assumptions are sketched in (1).

(1) General Principles:
- Each card is a grammatical "sentence", ungrammatical "sentence" or neither.
- A grammar is a set of conditions on sentences.
- UG is a set of primitives, including:
  - types of symbols (features)
  - logical operators defined over these symbols
- A card c is 'grammatical' with respect to a grammar G iff c satisfies the conditions imposed by G. In such a case we will say, informally, that c is 'in G.'

We will now explore how the nature of 'card' UG limits the set of possible languages available to a learner.

2.1 UG1

Assume first that UG makes available to the learner the (privative) feature NUMBERCARD which characterizes cards that bear the numbers two through ten. Further assume that UG makes available the four suits: clubs, diamonds, hearts, spades ($\clubsuit, \diamondsuit, \heartsuit, \spadesuit$). These also function as privative features.[2] Finally, assume that UG makes available the logical operator AND, which allows for the conjunction of features in structural descriptions. We call this version of universal grammar UG1.

(2) UG1

- Features:

  NUMBERCARD
  $\clubsuit, \diamondsuit, \heartsuit, \spadesuit$
- Operators: AND

2.1.1 Possible grammars given UG1

Now consider some possible grammars, given the definition of UG1. Our first grammar is G_1, which is characterized as follows: G_1 = [NUMBERCARD]. This is to be read as "A sentence/card is in G_1 if and only if it is a numbercard." So, the king of diamonds is ungrammatical in G1. This is because a king is not a numbercard. On the other hand the six of diamonds and the three of clubs are both grammatical in G1.

Consider a second possible grammar G_2, characterized as follows: G_2 = [NUMBERCARD AND $\diamondsuit$]. This is to be read as "A sentence/card is in G_2 if and only if it is a diamond numbercard." In this grammar the king of diamonds is still ungrammatical, but so is the three of clubs. The six of diamonds is obviously grammatical.

Now consider G_3, defined as follows: G_3=[$\spadesuit$]. That is,"A sentence/card is in G_3 if and only if it is a spade." It is obvious what the grammatical sentences of this grammar are, but we now ask, What is the representation of 5$\spadesuit$? K$\spadesuit$? 5$\clubsuit$? The answers are [NUMBERCARD AND $\spadesuit$], [$\spadesuit$] and *[NUMBERCARD AND $\clubsuit$],[3] respectively. Only the third is ungrammatical, since it is not a spade.

[2] Note that only one of these suit features can characterize any given card. Such restrictions will not concern us further.

[3] We are assuming that the learner retains access to the UG-given features, even if these features are not used in the acquired language. Rejecting this assumption would not substantively affect the argument, but would unncessarily complicate the exposition. We are indebted to Afton Lewis for discussions on this point.

Finally, consider G_4 which is characterized by no features at all. In other words, it places no restrictions on which cards are grammatical: G_4=[ ]. That is to say, "Every sentence/card is in G_4." But now, is this completely true? The answer is that it is true of all the cards characterizable by UG1, say the fifty-two cards that can be assigned a representation given UG1. However, a tarot card or even a Joker would not be grammatical in G_4, given UG1. (Thinking ahead a bit, what would their representation be?)

2.1.2 Impossible grammars given UG1

Since any given UG delimits the set of possible grammars, it is also instructive to consider a few impossible grammars, under the assumption of UG1. Consider first (non-)grammar F_1 described as follows: F_1 = [PICTURECARD]. In other words, "A sentence/card is in F_1 if and only if it is a picturecard." Clearly this is an impossible grammar, since UG1 does not provide for a class of all and only picture cards. (Recall that NUMBERCARD is privative by hypothesis.) Similarly, consider F_2 = [NUMBERCARD OR ♢]: "A sentence/card is in F_2 if and only if it is a numbercard or a diamond (or both)." This is an impossible grammar since UG1 does not provide the logical operator OR. Next consider a potential grammar with severely limited expressive capacity: F_3=[6 AND ♠], that is "A sentence/card is in F_3 if and only if it is the six of spades." This grammar is impossible given UG1 since UG1 does not provide the means to parse a six as different from any other number.

2.2 UG2

Now imagine another species endowed with a different universal grammar called UG2, characterized by the following features: [±PICTURE], which is equivalent to having the mutually exclusive privative features [NUMBERCARD, PICTURECARD], and [±RED], which is equivalent to having the mutually exclusive features [RED, BLACK]. UG2, like UG1, provides the operator AND.

(3) UG2

- Features:
  [±PICTURE]
  [±red]
- Operators: AND

2.2.1 Some possible grammars given UG2

A possible grammar given UG2 is G_5 = [+RED AND -PICTURE]: "A sentence/card is in G_5 if and only if it is a red numbercard." What is the representation of 7♢ in this grammar? What about 7♡? And 7♠? The answers are [+RED AND -PICTURE], [+RED AND -PICTURE] and *[-RED AND -PICTURE], respectively. Since the suits are not distinguishable given UG2, the learner parses the two red cards as [+RED]. Since the numbers are indistinguishable given UG2 (as was the case with UG1) the fact that the three cards in question are all sevens is lost to the learner. They are all just [-PICTURE]. Now consider G_6 = [+RED]:"A sentence/card is in G_6 if and only if it is a red card." This grammar will include all the red cards, hearts and diamonds, number and picturecards, though of course these distinctions are not made by creatures endowed with UG2—they are only made by beings whose genetic endowment equips them to represent such contrasts.

2.2.2 Some impossible grammars given UG2

It should be easy now to see that the following two potential grammars are impossible given UG2.

- F_4 = [♠]
  "A sentence/card is in F_4 if and only if it is a spade."
- F_5 = [+PICTURE OR -RED]
  "A sentence/card is in F_5 if and only if it is a picture card or a black card (or both)."

The first is impossible since UG2 does not distinguish the suits. The second, because UG2 does not provide OR. Note however, that although F_4 is impossible assuming UG2, its specification is identical to the grammar G_3 which is allowed by UG1. So, again, the nature of UG determines the set of possible grammars.

2.3 UG3

We leave it to the reader to confirm that the following characterization of a third UG, UG3, allows for G_7, G_8 and G_9, but excludes F_6, $F_7$and F_8.

(4) Description of UG3

- Features:

  [PICTURECARD]
  [2,3,4,5,6,7,8,9,10]
  [±RED]

- Operators: AND, OR

(5) Some Possible Grammars given UG3:

- G_7 = [+RED AND 9]
  "A sentence/card is in G_7 if and only if it is a red nine."
- G_8 = [-RED AND PICTURECARD]
  "A sentence/card is in G_8 if and only if it is a black picture card."
- G_9 = [PICTURECARD OR +RED]. "A sentence/card is in G_9 if and only if it is a red card or a picture card (or both)."

(6) Some Impossible Grammars given UG3:

- F_6 = [♠]
  "A sentence/card is in F_6 if and only if it is a spade."
- F_7 = [NUMBER]
  "A sentence/card is in F_7 if and only if it is a numbercard."
- F_8 = [-RED AND Q]
  "A sentence/card is in F_8 if and only if it is a black queen."

It is worth pointing out (as one of our reviewers did) that, given UG3, it is possible to acquire a grammar which is *extensionally* equivalent to F_7: "A sentence/card is grammatical if it is [2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 10]". Of course, the goal of linguistic theory is to discovery the 'correct' model of a speaker's grammar, one that is, for example, compatible with a theory of UG that underlies all human languages. In defining I-language, a matter of 'individual psychology' as the domain of inquiry for linguistics, Chomsky (1986) argued convincingly that the fact that knowledge of language is instantiated in individual minds/brains means that there is necessarily a 'correct' characterization of a speaker's grammar (or grammars). See Reiss (2000) for discussion.

It is also worth stressing that we have demonstrated how the nature of UG limits the set of possible grammars—the set of achievable final states of the language faculty is partially determined by what is present at the initial state.

2.4 An impoverished UG4

Now imagine that UG4 provides only a single privative feature: [♢]. What happens if we expose a learner to 5♢? The learner parses (constructs a representation for) [♢]. The "5" is unparsable. It is not *linguistic* information. Now, expose the learner to "6♡". The learner parses nothing! There is *no* linguistic information

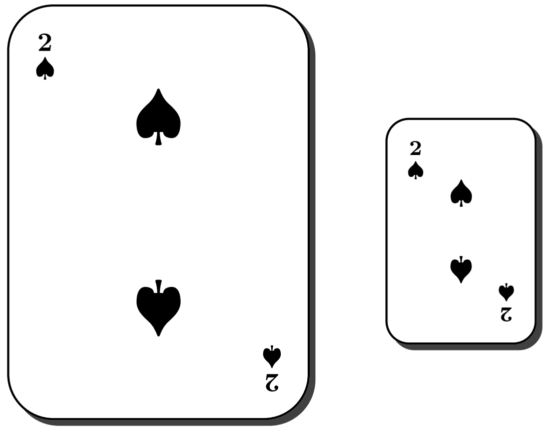


Figure 1: A Non-'linguistic' Card Contrast

in the input. (A linguistic parallel would be the parse of a belch by a human phonological system.) In fact only two grammars can be defined given UG4. G_{10}=[$\diamondsuit$] allows all and only diamond cards as grammatical utterances. G_{11}=[ ] defines, actually, a grammar which is extensionally equivalent to G_{10}, that is, the two contain the same sentences, but these sentences are generated by different grammars. The reason is that, given G_{11}, cards can either be assigned the representation $\diamondsuit$, or they are not parsed at all. So the only cards that will count as linguistic entities are the diamonds. (What happens if we instead make a binary feature [$\pm\diamondsuit$]? Hint: We again can define only two languages, but, unlike G_{10} and G_{11}, they do not contain the same sentences.)

2.5 A *really* impoverished UG5

What if UG provides nothing at all—no features and no operators? Then, no matter what we expose the learner to, nothing will be parsed. The starting point for the grammar we ultimately construct cannot be an empty slate since, to quote Jackendoff again, "Without Mental Grammar, there's no language perception—just noise" (1994:164). To reiterate: The set of primitives supplied by UG determines the set of possible grammars that can be described. Without any primitives, no grammar can be described. So the card language faculty of a creature endowed with UG5 will parse any given card in the same way as it will parse a tarot card, the Mona Lisa or the smell of pepperoni. Any innate system which parses such entities distinctly must be endowed with a mechanism for distinguishing between them. This mechanism, obviously, must be innate.

Before we move on, consider the contrast between a really large 2♠ (like a prop for a magician) and a really small one (like a card from a travel deck), as depicted in Figure 1. Obviously these two cards differ physically—one is big and one is small. They may even have different patterns on their backs and differ in many other ways. But the two cards are *linguistically identical*. They differ in the same way that whispering and shouting a given word differ, that is, they differ only paralinguistically.

Crucially, our claim is not that the contrast in card size will be *imperceptible* to an acquirer — merely that no size information will be used in the construction of the representations relevant to the 'linguistic' module. That is, given a particular card-UG, the relevance of specific contrasts that fall within the perceptual capabilities of the learner for card-grammar learning can be made explicit. The set of possible card-gramars consists precisely of those which are UG consistent. The fact that learner can perceive the difference between large cards and small ones, or between a card on the ceiling and a card on the floor, will not be relevant to the grammatical learning task. For a learner for whom these contrasts are perceptible any theory which fails to recognize innate primitives within the card-grammar domain will fail to properly constrain the set of possible grammars – i.e., the primitives of grammar construction *cannot* arise from the primitives of perception.

We have been forced to the logical conclusion that there must be something at the initial state of the grammar in order to allow learning to occur. However, one might object: "Maybe there are more basic primitives at the initial state. For example, if we are sensitive to the difference between straight and curved lines we could discover the distinction between ♢ and ♡." This is perfectly reasonable. It just means that, say, 'straight' *vs.* 'curved' are the innate primitives. But YA GOTTA START WITH SOMETHING! That something is Universal Grammar.

It should now be obvious that we are heading toward the conclusion that children must 'know' (that is, have innate access to) the set of phonological features used in all of the languages of the world. This is how the *IofPP* will be extended in this paper, but it is equally clear that the same conclusion holds for primitive operators like the AND and OR of card languages, or whatever are the operators of real grammars (both in phonology and syntax).

Obviously, we are not claiming that the set of primitives of phonology corresponds exactly to the set of distinctive features referred to in the literature. There is no question that some of the features have yet to be identified or properly distinguished from others (for some recent speculation on this matter, see Hale, Kissock and Reiss, forthcoming). In some cases a currently assumed feature may represent a conglomeration of the actual primitives of phonological representation. However, *by definition*, UG, the innate component of the language faculty, consists of the elements of linguistic representation which cannot be derived from anything else.

Consider a proposal that Q is necessary for the acquisition of human language and that Q is innate. Critics of the proposed innateness of Q must formulate their criticism in one of two ways. Either they must provide a learning path which is not dependent on Q – i.e., they must challenge the claim that Q is *necessary* for the acquisition of human language, or they must derive Q from something other more basic entities and processes (such as R), themselves available to the acquirer innately. In the absence of such alternatives, the criticism is invalid. The second alternative is the favorite of so-called constructivist theories of cognitive development. However, note that the appeal to 'general learning mechanisms', without specifying in detail what the set of actual primitives involved in any such mechanisms are, is not a responsible critique of the nativist stance.

We sympathize with one reviewer who noted that the 'card' grammar discussion in this paper appears to go into relatively painful detail about matters which should be *a priori* fairly clear and we beg the readers' forgiveness. However, we have found that for both readers of this paper and for the field in general, coming to grips with the basic insight that in order to learn over a given set of a data, the learner must possess the relevant representational primitives *within the learning domain*, has proven immensely difficult.

3 Acquisition of Phonological Inventories: The standard view

In this section we summarize and critique a theory that we see as characterizing the mainstream view on the acquisition of phonological inventories. We choose Rice and Avery (1995) because these authors are particulary lucid in their claims, and thus serve well as a basis for comparison with our proposal. The central claims of this paper are as follows:

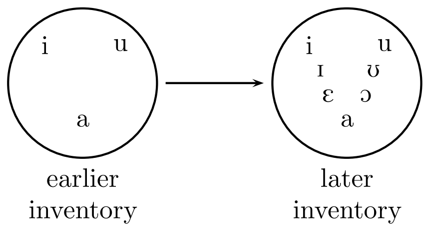
- Minimality: Initially the child's representational apparatus consists of a minimal set of primitives, say, C and V.
- Monotonicity: Representational capacity is expanded as further primitives become available.

What these claims mean is that a child's phonological representation apparatus is initially highly impoverished and is enriched over the course of acquisition. So, for example, a word like [ma] must be represented as just /CV/ at the initial stages of acquisition; later it may be represented as /[C, +son][V,-hi]/; yet later as /[C, +son, +labial][V,-hi,+bk]/; and ultimately as fully specified as the target language requires.[4]

[4]Our example uses an *ad hoc* feature system without regard for any kind of feature geometry, but the point should be clear.

Rice and Avery do not explicitly invoke the Subset Principle (SP), but their account of the learning path looks superficially like it conforms to the SP. For example, the child moves from having a small phonological inventory to having a larger one which contains all the feature contrasts, and thus all the underlying segments, present at the earlier stages. This traditional view, which we will reject, continues the same basic attitude towards children's phonological development and its reflection in their speech output as that expressed by Jakobson (1941). A standard version of this view is sketched in (7): the early states of the grammar contain a limited number of vowels, e.g., a single vowel or the three 'basic' vowels represented here; acquisition of a larger inventory leads to a superset of this early inventory.

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



Now, we must ask ourselves if it is indeed possible to get from the earlier stage to the later stage sketched above. We will argue that such a learning path is not possible given standard assumptions about the language faculty. First, assume the child is at a three vowel stage, as above. Then any vowel that the child is presented with must be parsed as one of the three, or else it will not be parsed at all. This claim follows from the definition of parsing: to parse a string is to assign it a linguistic representation.[5] A representation can only be assigned using the available representational apparatus. In the case of vowels under discussion, this gives rise to two distinct logical possibilities. A child at the three-vowel stage, presented with an [ɪ], could, under certain assumptions, parse it as an instance of the (ATR underspecified) high front vowel [i] (we return to this matter in section 5.2 below). Alternatively, the learner could fail to parse the segment as a vowel at all. No other possibilities exist under the traditional view, and neither of these possibilities will lead to an expansion of the vowel inventory. Clearly, if the learner could parse the [ɪ] as a vowel, distinct from [i], this would entail the learner having access to the feature contrast which is, by hypothesis, not available.

It is sometimes suggested (*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 from at least three difficulties. First, the desired 'simplicity' or 'poverty' of the child's grammar is not attained, but rather just holds at one level of the grammar — 'phonetic' representations are neither 'simple' nor 'impoverished'. Second, as assumed by Pinker (1996) and many others, the child's grammar should be considered to differ from the adults quantitatively, not qualitatively (since, after all, children are small humans). Having this extra lexicon would surely constitute a qualitative difference between child and adult grammars. In fact, it would appear to endow them with a *more*, not *less*, complex grammar than adults. Third, in generative grammar, surface representations are not stored. They are *generated*. In fact, the potential to create and process an unbounded number of utterances from finite means is perhaps the *raison d'être* of the generative program. Forcing the child to memorize all tokens also violates the learning

[5] A reviewer claims that we are guilty of equivocation here. His/her criticism involves an assumption that it is sufficient to allow 'parsing' to mean 'establish a *phonetic* representation', rather than a phonological one. This rich phonetic representation is then supposed to be exploitable in the phonological domain, the reviewer describing the Rice and Avery system as building on the fact that "phonological representations are stored in long-term memory and are compared with the phonetic representations in short-term memory." Our point, building on the card-grammar discussion, is that it does the learner of phonological representations simply no good to compare rich phonetic representations with a phonological representational apparatus which is impoverished in that it cannot represent, and thus cannot 'see', the richness available there. To be relevant for phonological learning, our concern here, a form must be parsed *in the phonological domain*.

theoretic correlate of generative capacity, namely that the learner is *one memory limited*. As discussed by Pinker (1996) and others, this means that the child is assumed not to store every utterance, but rather to update the *grammar* upon exposure to new evidence. The updated grammar must generate the new data, but surface utterances are not stored individually.[6] Indeed, the storage of surface forms would vitiate the need for phonology at all — the child, and the adult as well, could simply pronounce the stored surface forms.

A variant of the two lexicon hypothesis is that the special, additional lexicon that children have is not feature based, but is instead based on 'raw acoustic images' (that is, non-linguistic representations). This version of the proposal shares the three aforementioned problems, but is further burdened by the evidence that speech is recognized and processed differently from other sounds immediately from birth. Thus there is no reason to believe that children would store speech input as raw acoustic signals to a greater extent than adults do (as in the occasional case where we remember the manner in which a particular word was spoken). Note that storing raw acoustic images exclusively is not only irrelevant to language acquisition, but is in fact an impediment to the process of extracting a discrete, constant representation from vastly variant tokens.

Theories like that of Calabrese (1988) posit a universal set of markedness statements that are 'deactivated' upon exposure to positive evidence that a given featural distinction is exploited in the target language. In Hale & Reiss (2000ab) we argue that markedness is not a coherent notion, and that informal statistical arguments for the marked status of given feature configurations are flawed. In addition to such problems, a theory like that of Calabrese faces other difficulties. First, the implicational universals concerning phonological inventories are devoid of empirical content since the theory does allow for 'accidental' gaps to occur. For example, most markedness theories posit that [t] is less marked than [k]. However, there is at least one language, Hawaiian, which has [k], but not [t], due to a fairly recent sound change. In order to have any explanatory value, such a theory must predict that Hawaiian speakers would accept [t] as a possible Hawaiian sound, but be unable to provide any examples of its occurrence. Although we have not done the empirical work, we find such a scenario implausible. Now, a theory that allows arbitrarily structured inventories will obviously allow allegedly motivated ones. Therefore, a theory that proposes that phonological inventories are not constrained by markedness considerations allows for those that follow from markedness arguments, as well as those which do not. Removing markedness from phonological theory thus leads to no loss in empirical coverage.

To summarize, Rice & Avery's theory, and those like it, must be rejected as unparsimonious, incompatible with the generative program, and incapable of modelling a successful learning path. In the following section we discuss the Subset Principle (SP) of acquisition in general terms. We then apply the SP to the problem of the acquisition of phonological inventories.

4 The Subset Principle

It is worthwhile to remember that the skepticism with which the claim of 'no negative evidence' is sometimes treated is misguided. The availability of negative evidence greatly *simplifies* a learner's task. Thus, any scholar who assumes 'no negative evidence' is undertaking a harder job than one who assumes that the child does get negative evidence. If we can find a successful learning algorithm that does not rely on negative evidence, it will by necessity be successful even if negative evidence is provided. However, a learning algorithm that makes use of negative evidence may not succeed using only positive evidence. There are no explicit proposals concerning the kind of negative evidence that children get for phonology, and there are explicit arguments (Marcus 1993) that negative evidence is not used in syntactic acquisition. We therefore take upon ourselves the more difficult task of accounting for phonological acquisition without appeal to negative evidence.

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. In other words, the SP can be viewed as a corollary to the acquisition principle of 'no negative evidence'. The effect of the SP is to prevent

[6]The model proposed by Tesar & Smolensky (1998) is another example of a learning algorithm that fails to recognize this basic assumption.

the learner from making overly broad generalizations which cannot be corrected on the basis of positive 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 restrictions 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.

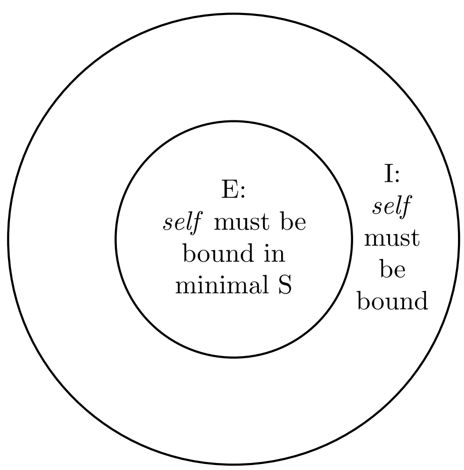
In a comparative study of acquisition of anaphora, Hyams and Sigurjónsdóttir (1990) compare the binding conditions on Icelandic *sig* and English *himself/herself*. In simple terms, we can characterize the anaphors in the two languages as follows: Icelandic anaphors need to be bound; English anaphors 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 (8). 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.

(8) Anaphors in English and Icelandic

- English: John_i asked Bill_j to shave $\text{self}_{*i/j}$
- Icelandic: John_i asked Bill_j to shave $\text{self}_{i/j}$

We can represent the greater restrictiveness of English as in (9) and conclude that English corresponds to the initial state (in this respect).

(9) Relative restrictiveness on anaphors



We can also represent the relationship of the two languages as an implicational relationship, as in (10).

(10) The SP as an implicational hierarchy

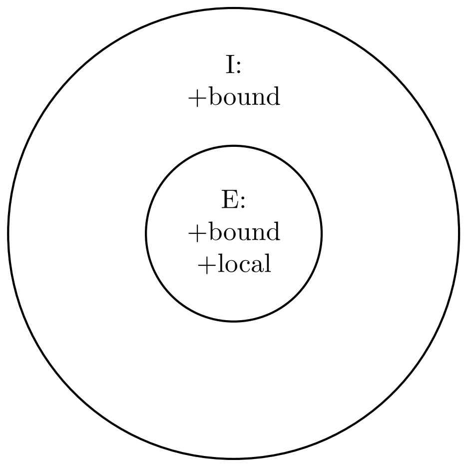
a) Anaphor must be bound in the minimal S $\Rightarrow$ Anaphor must be bound.

b) Anaphor must be bound $\nRightarrow$ Anaphor must be bound in minimal S.

One of the reasons that the SP has fallen out of favor in discussions of syntactic acquisition is that a number of prominent contemporary syntactic theories (*e.g.,* Minimalism) assume an invariant syntactic component cross-linguistically. It becomes unclear how to state subset relations among languages if they all have the same syntax. In fact, the solution to this problem is quite simple in our view, and is completely compatible with a theory like Minimalism. The SP in syntax should be defined with reference to the representation of lexical entries (functional and lexical categories).[7]

If we represent the distinction between English and Icelandic in terms of lexical features of the anaphors, instead of in terms of parameter settings, as has been done traditionally, we might propose the model in (11), where English anaphors are marked as [+bound, +local] whereas Icelandic anaphors are marked only as [+bound].

(11) 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 in phonology. The point can be formulaically stated as: fewer features equals more entities. That is, assuming that feature combination is restricted to conjunction (as required by the notion 'natural class') the size of a class varies inversely with the number of features used to define the class. This is stated more formally in (12).

(12) Fewer features = more entities

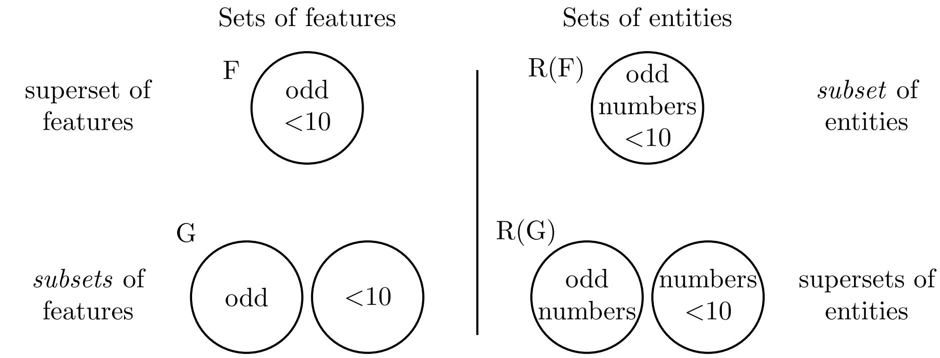
> Let F and G be sets of features such that R(F) is the set of entities defined by F and R(G) is the set of entities defined by G. If G is a subset of F, then R(F) is a subset of R(G). That is $F \supset G \Leftrightarrow R(G) \supset R(F)$.

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. Let's assume that, like linguistic features in lexical representations, these features can only be combined conjunctively. As shown on the left-hand side of (13), 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

[7]Despite the fact that the parameterization of Principle A that motivated work on the acquisition of long-distance anaphora is somewhat obsolete, we can retain it as an example. It is clearly beyond the scope of this paper to account for binding theory and all issues in syntactic acquisition, such as the 'null subject parameter'. Our revision of the SP, from parametric terms to featural terms, may very well revive some old discussions in syntactic acquisition.

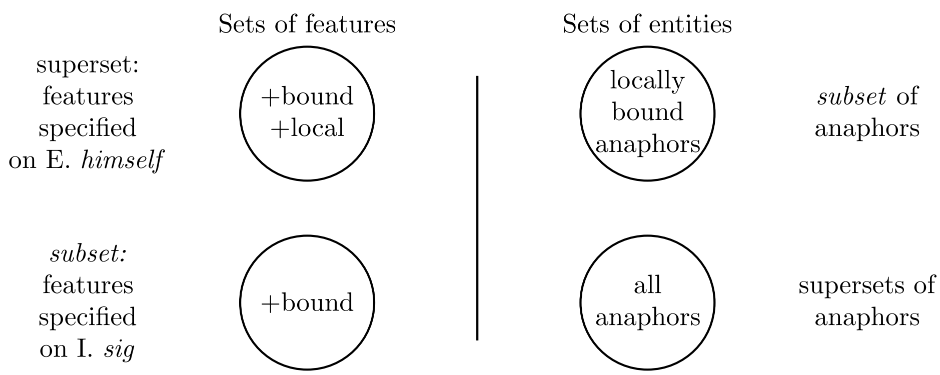
of numbers which are both odd and less than 10 is contained within the set of odd numbers and within the set of numbers less than 10.

(13) 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 (14) 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.

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



So, English represents the subset or initial state for the acquisition of anaphors by virtue of the fact that it uses a *superset* of features to restrict its characterization of pronouns.

Before we come back to phonology, we will return briefly to the card languages. In order to be consistent with the SP, what strategy should a learner endowed with UG1 employ? Recall that UG1 provides the four suits, the category NUMBERCARD and the operator AND. Compare a learner L_1 who must learn $G_1 =$ [NUMBERCARD] to one L_2 who must learn $G_2 =$ [NUMBERCARD AND $\diamondsuit$]. Since we assume that the learners belong to the same species and thus are endowed not only with identical UGs (UG1), but also with identical language acquisition devices, we must assume that they go about learning in the same way. Confronted with, say, [5$\diamondsuit$], our theory—which gives the learners full access to the primitives provided by UG1—requires both learners to construct a maximally specific parse, namely [NUMBERCARD AND $\diamondsuit$]. Both learners will first assume that all grammatical cards must be diamond numbercards. L_2 will never get conflicting data and thus the learning path is complete for this learner. L_1 *will* get conflicting information, that is, positive evidence, all of it consisting of numbercards which are *not* diamonds. Thus, L_1 will have to relax restrictions on the definition of a grammatical card to arrive at G_1. In both cases the learning strategy will be successful. And note that more features are needed to state the constraints which characterize the more restrictive G_2.

The traditional view of inventory acquisition can also be modeled using card languages. Imagine two learners, Z_1 and Z_2, with target grammars G_1 and G_2, respectively, again both genetically endowed with

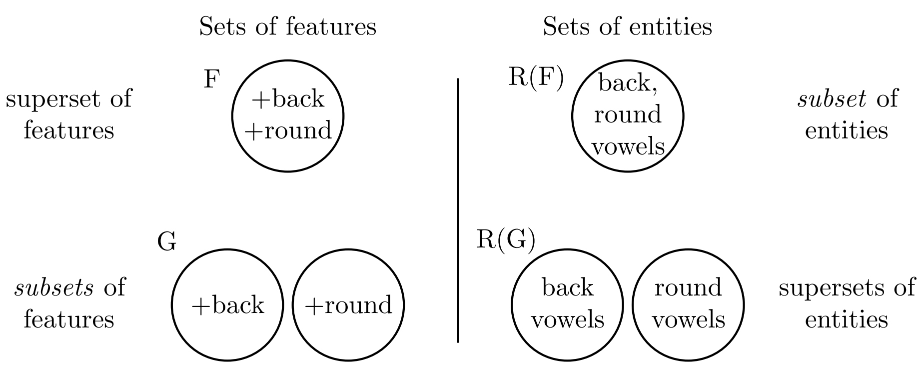
UG1. By hypothesis, these learners under the traditional view do not have access at the initial state to the full set of UG1 primitives. Their representational capacities allow them only to construct representations using a single primitive from UG1, say, [NUMBERCARD] at the initial stage. Presented with [5♢], both Z_1 and Z_2 will hypothesize that their target grammar is characterized by [NUMBERCARD], that is, that every grammatical sentences be a numbercard. Z_1 will be fine, since that is indeed the target grammar, but Z_2 will be in trouble. All future data will be consistent with this first hypothesis, since that data will consist completely of numbercards that just happen to be diamonds. Note that even when additional primitives (e.g., AND, ♢, ♠, ♣, ♡) become available, the existence of a stored lexical representation consisting solely of [NUMBERCARD] and assumed by the learner to be grammatical, precludes construction of the more restrictive target grammar.

5 SP and segment 'inventories'

5.1 Further problems with the traditional view

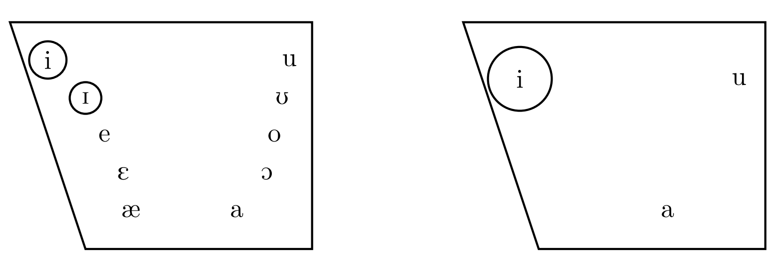
Recall that the traditional view of inventory acquisition appears to conform to the SP: earlier states have fewer segments than later states. There are two reasons to be skeptical of this superficial impression. First, there is no reason to expect 'segments' to play a role in the learning path, since features are the primitives of phonological theory. Second, the inverse relationship of features and natural classes discussed above leads to an alternative interpretation, as (15) shows.

(15) The class of back vowels contains the class of back, rounded vowels.



The final issue concerns the phonetic space. Consider two languages, one with a rich vowel inventory and one with a restricted inventory, shown in (16). 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.

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



The traditional view of inventory acquisition in (7) fails to take account of the fact that the *space* occupied by a high, front, nonround vowel underspecified for ATR may be identical to that occupied by the two fully specified vowels [i] and [ɪ]. The traditional view of acquisition sees the acquirer as moving from state A to state B, however it is clear that the subdivision of the underspecified high front unround vowel space into two more restrictive target spaces does not involve going from a subset to a superset, unless one defines the relevant sets over IPA symbols, rather than over phonetic space or phonological features.

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 other words, the representational primitives of phonology must all be innately available to the learner.

5.2 The learning path

In order to provide learnability arguments to support this proposal and justify rejecting the traditional theory, we must answer the two questions in (17). Below we provide arguments using hypothetical languages to justify the answers we provide.

(17) The questions:

a) Can the traditional view lead to a growing inventory? NO, so it must be rejected.

b) Can the proposed view lead to a shrinking inventory? YES.

In order to answer (17a) 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 phonetic 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 parsed linguistically if the child does not have the appropriate representational apparatus. This is a fundamental assumption of linguistic theory. It is equivalent to saying that a language that uses a feature which is not available to humans 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. 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.

We now turn to (17b), loss of a 'wrong' contrast, i.e., /dip/ and /dɪp/ collapse to /dip/ in some language with a three vowel system. The challenge to the theory proposed in this paper 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.

(18) Two distinct cases for the 'collapse' of contrast

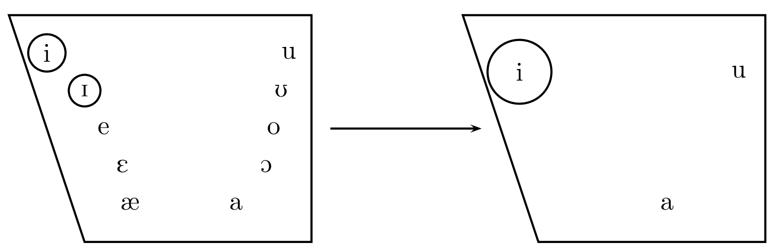
a) Unobserved contrasts: If the target language does not present forms such as [dɪp], then there may never be any reason to remove [+ATR] from the representation of /dip/. Access to the universal feature set allows the *potential* for any contrast, not its realization. This is clearly relevant to underspecification 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 some three vowel systems, this child may also hear phonetic [dɪp] for what the speaker intends to be the same lexical item as the acquirer's existing [dip] representation. The child, under our proposal, will then 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 representations. 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/.

(19) Lexicon optimization

$$\underbrace{\begin{bmatrix} \text{+hi} \\ \text{-bk} \\ \text{-rd} \\ \text{-lo} \\ \text{+ATR} \end{bmatrix}, \begin{bmatrix} \text{+hi} \\ \text{-bk} \\ \text{-rd} \\ \text{-lo} \\ \text{-ATR} \end{bmatrix}}_{\text{before lexicon optimization}} \longrightarrow \underbrace{\begin{bmatrix} \text{+hi} \\ \text{-bk} \\ \text{-rd} \\ \text{-lo} \\ \end{bmatrix}}_{\text{after lexicon optimization}}$$



before lexicon optimization after lexicon optimization

Clearly, this account needs to be further developed by an explicit model of lexicon optimization. See Hale & Reiss (2002) for some suggestions.

It is of value to examine in some detail just how the 'traditional' analysis—which assumes that not all of the features are available to the acquirer at the initial state—fares in this 'shrinking inventory' task. The ultimate target in the hypothetical case sketched above is the representation of a high, front, non-round vowel underspecified along the ATR dimension. Assume that the 'marked' feature ATR is not available at some early stage in the acquisition task. Given the broad phonetic target space of our hypothetical three-vowel system, the acquirer will hear both [dip] and [dɪp], however, given the child's limited representational apparatus, s/he will posit representations which are underspecified along the ATR dimension. These are, in fact, the correct representations and one might therefore assume that the child's learning task with respect to these vowels would be complete—and that the swift and direct nature of the acquisition in this case would lend support to the traditional model. But how could the child *know* s/he was done? An acquirer would also assume, at this stage, underspecified representations in the case of a different ambient language which *did* make use of ATR features. That is, the fact that all high, front, non-round vowels are underspecified in our hypothetical case is a language-specific property which must be learned by explicit evidence. In fact, when the ATR feature becomes available for the construction of representations, the learner will begin to (wrongly, *vis-à-vis* the target language) construct fully-specified representations (just as s/he must do so correctly when exposed to a language such as English, in which ATR is contrastive). This is because, in the hypothetical case under discussion, the broad target space for the ATR-underspecified vowel allows for hits in both the [dip] and [dɪp] space. The adults providing the evidence for the target language will therefore produce hits throughout the broad target space. Of course, the child cannot know in advance if s/he is learning a language that uses ATR contrastively.

At this point, the child will face the same learning task as the learner initially faced under our assumptions regarding the learning path (as outlined above)—i.e., s/he will need access to some type of 'lexicon optimization' in order to learn that the [i] : [ɪ] difference now posited for certain lexemes is in fact non-contrastive. In other words, the learning path for such a language under the traditional scenario consists of the learning path under our proposed scenario preceded by an additional stage that is of no use to the learner. The initial lack of availability of ATR as a feature to be used in the construction of lexical representations thus works to the acquirer's advantage *neither* in the 'growing inventory' nor in the 'shrinking shrinking' scenarios.

5.3 Empirical evidence

In addition to being logically consistent with standard assumptions of learnability theory and linguistic theory, our proposal of innate access to full representational apparatus is also plausible in light of well established empirical studies. For example, psycholinguistic experiments show that even newborns can distinguish sounds that constitute possible phonetic contrasts in the languages of the world and that by 10 months of age they have lost some of their power of discrimination. So language acquisition is, in a very real sense, a process of loss—we are "deafened" by our experience.[8] Given that very young infants appear to be sensitive to any contrast that is used in the languages of the world, a theory of phonological acquisition that reflects their innate phonetic capacities seems preferable to one that must grant these well supported results (*e.g.*, Streeter 1976, Goodman and Nusbaum 1994), but then claim that children's phonological capacities are severely impoverished.[9] Our proposal represents the null hypothesis.

Innate access to categorical, rather than gradient, phonological features is also consistent with the well established studies of categorical perception of speech in infants (e.g. Miller & Eimas 1983). Indeed, it is possible to define the phonetics/phonology boundary as consisting of a distinction between categorical and gradient phenomena.

Finally, we know that children's representational ability is always far ahead of what we perceive to be their speech output:

> ...they appear, in many respects, to have adult-like representations, which are reflected, among other things, in their vociferous rejections of adult imitations... (Faber and Best 1994: 266-7).

For further discussion of the competence/performance gap in child language see Hale & Reiss (1998).

The model we have constructed is strikingly paralleled by Patricia Kuhl's work on the perceptual magnet effect (e.g., Kuhl & Iverson 1995; see Hawkins 1999 for a useful overview). According to this theory, the auditory space of infants is divided by a set of natural (innate) boundaries that provide an upper limit on the number of possible vowel contrasts in human language. Early exposure to speech provides the child with information about which of the categories defined by these boundaries are exploited in the ambient language. Some of the innately available categories which are not used are still discriminated in experimental situations. During development, the vowel space is reorganized in that certain of the innate boundaries come to be ignored in linguistic processing. The regions of the vowel space on either side of these boundaries are merged into a single category from the linguistic perspective.

The parallel between the innately provided maximum number of, say, vowel distinctions and an innate set of phonological features that can define vowels is obvious. It is also obvious that the auditory system imposes upper limits on the number of contrasts made by speech perception modules, which in turn imposes limits on the number of constrasts made by the grammar. As Pylyshyn (1984) points out, the "*computationally*

[8]These are the so-called phoneme discrimination tests. The trouble with the term *phoneme discrimination* is that phonemes can only be defined on the basis of lexical contrasts (such as minimal pairs). Since a one-month old infant is not distinguishing sounds on the basis of a phonemic contrast in his or her own lexicon, a better term would be *discrimination of potential phonemes*: "In general, it should be observed that 'minimal pair' is not an elementary notion. It cannot be defined in phonetic terms, but only in terms of a completed phonemic analysis" (Chomsky 1964:97).

[9]For an extensive criticism of the widely misunderstood results of Maye, Werker and Gerken (2002) in this regard, see Kissock (2002), with which we are in full agreement.

relevant states are a tiny subset of [a system's] physically discriminable states", and the "former are typically a complex function of the latter" (150). So, it is not necessary to come to a decision about which layer of Kuhl's model best corresponds to the level of featural (linguistic) representations. We can just note that the model of the learning path proposed here converges nicely with Kuhl's model.

5.4 Summary

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

(20) 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.

C. Empirical Argument

  a. Infants appear to be innately sensitive to any possible phonological contrast.

  b. Phonological contrasts are parsed and represented in terms of features.

  ∴ 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).

6 Innateness and Maturation

It has been suggested that the claims of innateness argued for in this paper can be replaced by appeal to maturation. There are several reasons why this approach is less appealing than ours. First of all, if a given featural distinction is not available at a certain stage of acquisition, then, by definition, the child cannot differentiate representations that are distinguished in the adult grammar by means of this feature. So, such representations could only be acquired as identical ('homophonous'). Once the relevant feature is made available by maturation, it cannot just be 'plugged' into the learner's lexicon in the appropriate places with respect to the adult lexicon. Instead, the child must *relearn* the relevant forms (by exposure to them). In other words, the maturation view just consists of a sequences of learning and unlearning, and the last of these learning stages, that is, at the end of the relevant maturation sequence, corresponds to the stage of full access to the universal feature set. This is our initial stage and the state required to lead to full acquisition.

Second, appeal to maturation in representational capacity is based on superficial impressions of children's *speech*, and not a principled analysis of their linguistic abilities. Such an analysis should make reference to careful acoustic and articulatory examination of their speech, as well as to their comprehension abilities. We have gone through the relevant facts and arguments in some detail in Hale & Reiss (1998), so we provide only a brief overview here. Studies on the acoustics of child speech has shown that sounds that are perceived as identical (merged) by adult transcribers are, in fact, distinct (Kornfeld & Goehl 1974, Gibbon 1990). For example, children who sound to adults as if they are merging initial [r] and [w], in fact are not doing so. Therefore, there is no reason to believe that the child lacks the representational apparatus necessary to encode such distinctions.[10] Instead the child merely lacks the ability to articulate the relevant distinction with sufficient clarity. This inability can be due to physiological immaturity, overloaded resources in motor planning, attentional deficits, lexical access delays, etc. The fact that children's pronunciation typically improves when they focus on the task of articulation supports the 'performance' view we espouse.

If children's speech output were a reliable indicator of the state of the grammar, we would expect that they would be able to parse their own speech. However, Dodd (1975) showed that, at a point when they successfully parse recorded speech of unfamiliar adults, children may fail to parse their own recorded speech if it deviates from adult output.

A third argument against the appeal to maturation is one of parsimony. The infant speech perception studies cited above suggest that children innately possess an ability which is relevant to the categorization of speech sounds. The maturation hypothesis entails poor representational capacity at the stage when discrimination abilities are most sharp and increased representational capacity just as discrimination abilities are being attenuated.

The final justification for our approach, as opposed to a maturational view, is adopted basically from Pinker (1984/95). When we find that children's speech output differs from that of adults we can assume that the difference is due to a discrepancy in performance systems, or to a hypothesized discrepancy in representational capacity, or to both. We know that children's (cognitive and physiological) performance systems are extremely underdeveloped. Therefore, even in the absence of logical arguments against such a discrepancy in representational capacity, we would be forced by Occam's Razor to assume as the null hypothesis that the discrepancy in output is due *only* to the immaturity of the performance systems. Of course, this hypothesis is subject to revision, but it represents the most constrained model available. Now, the only reason to posit stages of maturation in representational capacity is presumably to account for observed differences in speech output over time. However, since we know that the performance systems are becoming more and more adult-like over time, we can, and must, attempt to derive developmental stages from this (non-grammatical) maturation which is independently necessary for a full account of development.

7 Conclusion

We believe that our model also provides some insight into the history of phonological theorizing. In the context of the Jakobsonian model of initially impoverished representations, it made sense to assume that languages ended up with high degrees of underspecification. Adding fewer contrasts represented a shorter learning path, and thus a more elegant acquisition model. We are not aware of this rationale being made explicit anywhere but it seems relevant. However, in the context of a model that strips away representational matter when forced to by the nature of the evidence, excessive underspecification represents an unmotivated extension of the learning path. Thus our model actually requires a phonology with little underspecification.

We hope to have explicated Jackendoff's statement and thus shown that the *Innateness Hypothesis* is perhaps a misnomer for a logical necessity, the *Innateness of Primitives Principle*. Given its generality, we look forward to extending the reasoning used here to other areas of linguistic theory, such as the acquisition of the syntactic features that distinguish lexical items. In addition to the logical arguments we provide, we show that our view converges strongly with empirical work concerning children's speech comprehension capabilities. We also find that our model more accurately mirrors the child's learning path, since children do not appear to need to relearn each lexical item when they begin to show a new contrast in pronunciation—which is what a view that denies early access to all features would predict.

We took pains to illustrate the logic of the *Innateness of Primitives Principle*. While several readers of this paper objected to the length of this discussion, we have chosen to maintain it in its current form because of the oddness of the conclusion it leads us to. The notion that infants must have access to all the

[10]This is especially true given their ability to distinguish them in comprehension.

representational apparatus of UG *is* odd, but as Zenon Pylyshyn (1984:xxii) has pointed out:

> [I]f you believe P, and you believe that P entails Q, then even if Q seems more than a little odd, you have some intellectual obligation to take seriously the possibility that Q may be true, nonetheless.

We see this paper as a step towards fulfilling this intellectual obligation.

References

Archibald, John, (ed.). 1995. *Phonological Acquisition and Phonological Theory.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

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

Calabrese, Andrea. 1988. *A theory of phonological alphabets.* PhD dissertation, MIT.

Chomsky, Noam. (1964). Current issues in linguistic theory. In Fodor, J. & Katz, J. (eds.) *The structure of language.* Englewood Cliffs, N.J.: Prentice-Hall. 50-118.

Chomsky, N. 1986. *Knowledge of Language.* Westport, CT: Praeger.

Dell, Franois. 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.

Faber, Alice and Catherine T. Best. 1994. The perceptual infrastructure of early phonological development. In Susan D. Lima, Roberta L. Corrigan, Gregory K. Iverson, (eds.). *Reality of Linguistic Rules.* Philadelphia: J. Benjamins. 261-280.

Fodor, J.A. 1975. *The Language of Thought.* Sussex: Harvester Press.

Gibbon, F. 1990. Lingual activity in two speech-disordered children's attempts to produce velar and alveolar stop consonants: evidence from electropalatographic (EPG) data. *British Journal of Disorders of Communication* 25: 329-340.

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

Hale, M. & C. Reiss. 2000a. Substance abuse and dysfunctionalism: Current trends in phonology. *Linguistic Inquiry* 31:157-169.

Hale, M. & C. Reiss. 2000b. Phonology as cognition. In N. Burton-Roberts, Philip Carr & Gerry Docherty (eds.), *Phonological Knowledge.* OUP: Oxford. 161-184.

Hale, M. and C. Reiss. 1998. Formal and Empirical Arguments Concerning Phonological Acquisition. *Linguistic Inquiry 29*:656-683.

Hale, M. and C. Reiss. 2002. Grammar Optimization. Manuscript, Concordia University.

Hale, M., M. Kissock, and C. Reiss. 2002. Microvariation, Variation and the Features of Universal Grammar. Submitted to *Language and Speech.*

Hawkins, Sarah. 1999. Reevaluating assumptions about speech perception. Chapter fifteen in J.M. Pickett (ed.) *The acoustics of speech perception.* Boston: Allyn and Bacon. 232-288.

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. 1995. The Acquisition of Negative Constraints, the OCP, and Underspecified Representations. In J. Archibald (ed.) 1995a: 63-79.

Jackendoff, Ray. 1994. *Patterns in the Mind: Language and Human Nature.*New York: BasicBooks.

Jackendoff, Ray. 1990. *Semantic Structures.* Cambridge, MA: MIT Press.

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

Kissock, Madelyn. 2002. Do Experimental Studies Inform Theoretical Research? Manuscript, Oakland University.

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. 210-219.

Kuhl, P. & G. Iverson. 1995. Linguistic experience and the perceptual magnet effect. In W. Strange (ed.) *Speech perception and linguistic experience.*Timonium, Md.: York Press.

Marcus, Gary F. 1993. Negative evidence in language acquisition. *Cognition 46*: 53-85.

Maye, Jessica, Janet Werker, and LouAnn Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition* 82: B101-B111.

Miller, J. & P. Eimas. 1983. Studies on the characterization of speech by infants. *Cognition* 13:135-165.

Pinker, S. 1996. *Language Learnability and Language Development* (2nd edition with new commentary by the author; 1st edition 1984). Cambridge, MA: Harvard University Press.

Pylyshyn, Z. 1984. *Computation and cognition: toward a foundation for cognitive science.* MIT Press, Cambridge, MA.

Pylyshyn, Z. 1973. On the role of competence theories in cognitive psychology. *Journal of Psycholinguistic Research.*

Rice, Keren and Peter Avery. 1995. Variability in a deterministic model of language acquisition. In J. Archibald (ed.) 1995a: 23-42.

Reiss, C. (2000). Optimality Theory from a Cognitive Science Perspective. *The Linguistic Review* 3-4: 291-301.

Reiss, C. (2001). L2 Evidence for the Structure of the L1 Lexicon. *International Journal of English Studies* 1: 219-239.

Reiss, C. (2003). Deriving the feature-filling / feature-changing contrast: An application to Hungarian vowel harmony. *Linguistic Inquiry* 34:199-224.

Streeter, L. A. 1976. Kikuyu labial and apical stop discrimination. *Journal of Phonetics 4*:43-9.

Tesar, Bruce and Paul Smolensky. 1998. The learnability of Optimality Theory. *LI 29*:229-268.

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. Parameter Setting, ed. by Thomas Roeper and Edwin Williams, 41-76. Reidel, Dordrecht.