Paper presented at CLS 27
April 1991

To Joe Pater, with best wish,
Elan

# CUES AND PARAMETERS IN PHONOLOGY

**B. Elan Dresher**
**University of Toronto**

I would like to begin by considering what the problem of learnability is in a general way, and then I will turn to some issues that arise in developing models of learnability for a parametric grammar. I will be looking particularly at what sort of cues a learner would need to set parameters, what Lightfoot (1989) calls the learner's 'trigger experience'. Examples will be drawn from my work with Jonathan Kaye on metrical phonology, but I believe the conclusions hold generally for other areas of the grammar.

I will assume here a 'principles and parameters' model of grammar, as proposed in Chomsky (1981) and many subsequent works. On this conception, the basic principles of grammar are fixed by Universal Grammar (UG), and hence do not have to be learned. The principles of UG incorporate a set of open parameters which may take on a limited range of values. The correct value of a parameter must be determined by the learner on the basis of experience.

## 1. Learnability: Idealization of Instantaneous Acquisition

The problem of learnability, on this view, becomes mainly the problem of how to determine the proper setting of parameters. It is useful to distinguish what Hornstein and Lightfoot (1981) have called 'the logical problem of acquisition' from what we can call 'the developmental problem of acquisition'. The logical problem, schematically diagrammed in (1), can be characterized as the problem of determining how easily one can learn the grammar G from the set of relevant input data D, given UG:

(1) The Logical Problem of Language Acquisition

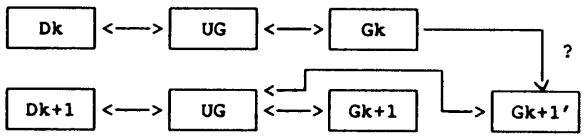
| D | <——> | UG | <——> | G |
|---|---|---|---|---|
| Data of L | | Universal Grammar | | Grammar of L |

The diagram in (1) incorporates some idealizations. One which will be of interest here is what has been called 'the idealization of instantaneous acquisition' (Chomsky and Halle (1968, Chap. 8), Chomsky (1975, Chap. 3)). We can appreciate what is meant by this idealization when we look at what it leaves out, which is the entire developmental problem of acquisition. In our terms, this problem can be taken to be that of specifying a series of diagrams like (1) representing the stages of acquisition. Thus, for some stage of acquisition k, a child has grammar Gk projected from a set of data Dk via UG.

The logical problem has been stated in terms of the final stage: the adult grammar Gf, and the final set of relevant data Df (I will assume for simplicity that UG does not change).

Let's consider the worst-case scenario: suppose that at stage k+1, a learner does not wipe the slate clean in projecting a grammar from Dk+1; rather, suppose the learner preserves some aspects of Gk, leading to some patchwork adjustments. The result is, as shown in (2), not the grammar Gk+1 we would have expected, but rather Gk+1', a grammar that deviates from the theoretically expected one (i.e. the grammar we would obtain by applying UG to Dk+1 afresh), in ways that could not be explained without taking into account earlier stages of acquisition:

(2) The Developmental Problem of Language Acquisition




The more acquisition is like this, the less tractable the logical problem as given in (1) becomes, because the real final grammar would deviate a great deal from any grammar we could obtain by directly applying UG to Df. The hypothesis suggested by the idealization of instantaneous acquisition is that the early stages do not play a crucial role in determining the final result - i.e. that we can consider acquisition *as if* it were instantaneous; the final grammar can be viewed as if projected from the final set of data, with no significant distortions caused by earlier stages.

## 2. Two Types of Learners

When we try to formulate explicit learning models, we notice that the notion of instantaneous vs. noninstantaneous learning arises in another sense, in that the model may or may not assume that learning is instantaneous at each stage. An example of an instantaneous learning model in this second sense is one in which the learner collects data for some preset amount of time, or until it decides it has seen everything important, before attempting to set any parameters. Such a model, which has access to all the relevant data, we can call a 'batch mode' learner (3a):

(3) a. *Batch Mode Learner*
Collect all data, then set parameters.

b. *Incremental Learner*
Adjust parameter settings as each datum comes in.

Now contrast this with another possible model, one which operates in 'incremental mode' (3b). An incremental learner processes data as they come in, trying to extract as much information as it can from each new datum. The choice of a batch mode or incremental learning model is orthogonal to the idealization of instantaneous acquisition, for here we are concerned not with the effects of early stages of acquisition on later stages, but in how the model works at any given stage. An incremental model appears to lend itself better to a developmental interpretation, since we think of language acquisition in real time as being incremental, but this is not so straightforward, as we shall see.

## 3. Metrical Phonology

To make any further progress in these matters we will have to look at some examples, which I will draw from the learning model for metrical phonology that I worked on with Jonathan Kaye. Metrical theory makes a good laboratory to explore questions of parameter setting, because it forms a rich domain of interacting parameters. I'd like to quickly review some basic notions of metrical phonology, then consider some issues that arise.

Referring back to the model in (1), I will sketch the various components as they are relevant to the domain of word stress, beginning with the data, D. We assume the prior operation of rules that convert the acoustic signal into words and segments. We assume also that the various acoustic cues which indicate phonological stress are mapped into one of three degrees of stress. After this preliminary processing, the data relevant to learning word stress consists of words with vowels marked as bearing main stress (2), secondary stress (1), or no stress (0). Sample forms are given in (4), where *Vancouver* has the stress contour (1 2 0), and *algebra* has the stress contour (2 0 0):

(4) Sample Input Data

| | | |
|---|---|---|
| a. | v a l n c u 2 u v e 0 r | (Vàncóuver) |
| b. | a 2 l g e 0 b r a 0 | (álgebra) |

Forms like those in (4) serve as the initial input into our model. However, they do not yet represent the input to the stress component. Stress does not apply directly to strings of segments, but is sensitive to representations built on projections from syllable structure. In many languages, stress is sensitive to syllable weight, or quantity. So the first step in the analysis of the input data in (4) involves parsing the words into syllables.

Let us turn now to the metrical theory which serves as our representation of UG. For purposes of our project, we adopted a modified

version of the tree-based theory presented in Hayes (1981), though the trees are simplified in the manner of Hammond (1984) and Halle and Clements (1983), and approximate the bracketed grids of Halle and Vergnaud (1987).

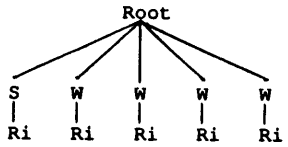
In metrical theory, stress patterns, and hence the stress levels observed in the data, are controlled by metrical structures which are built on projections from syllable structure. These metrical structures take the form, in this version, of labelled trees, where in any group of sister nodes, one node is labelled Strong and the others are labelled Weak. The various possibilities of metrical structure construction and labelling are expressed in terms of a series of binary parameters. Our model incorporates the eleven parameters listed in (5):

(5) Parameters of Metrical Theory (based on Hayes (1981), with refinements)

- P1: The word-tree is strong on the [Left/Right]
- P2: Feet are [Binary/Unbounded]
- P3: Feet are built from the [Left/Right]
- P4: Feet are strong on the [Left/Right]
- P5: Feet are quantity sensitive (QS) [Yes/No]
- P6: Feet are QS to the [Rime/Nucleus]
- P8L: There is an extrametrical syllable [No/Yes] on the Left
- P8R: There is an extrametrical syllable [No/Yes] on the Right
- P7: A strong branch of a foot must itself branch [No/Yes]
- P9: A weak foot is defooted in clash [No/Yes]
- P10: Feet are noniterative [No/Yes]

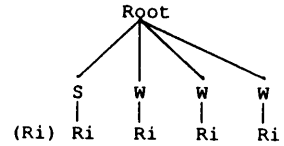
I will briefly illustrate the effects of these parameters, starting with P1. Main stress in a word is controlled by an unbounded word tree, in which either the leftmost or the rightmost node is labelled Strong. For example, the word tree in (6a) has been constructed with P1 [Left]; this gives initial stress, as in languages like Latvian, or Hungarian. Setting P1 [Right] gives fixed final stress, as in French and Farsi[1]:

(6) a. P1 [Left]: Main Stress on the Left




Main stress is not necessarily confined to a peripheral syllable, since P1 can interact with other parameters to produce different results. For example, a peripheral syllable may be designated as extrametrical by P8L or P8R, meaning it does not participate in the construction of the word-tree. Extrametricality can result in main stress falling on the second or penultimate syllable; in (6b), it falls on the second syllable, as in Lakota and Araucanian:

(6) b. P1 [Left], P8L [Yes]: Main Stress, Extrametrical Syllable on the Left

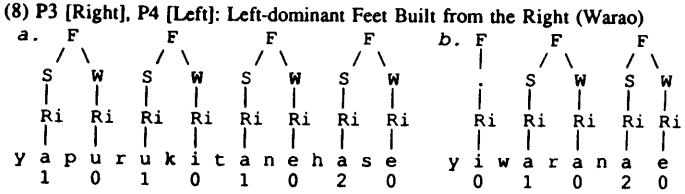



In trees such as these, only one syllable in each word is marked Strong, while all the rest are Weak. In most languages, however, syllables are first grouped together into feet, and then the word-tree is constructed on the feet. Every foot receives a stress; hence, languages with feet also have secondary stresses, though not always at the surface. If a language has feet, a number of other parameters come into play.

P2 allows feet to be at most binary, or else unbounded. Suppose we choose binary feet, which give rise to an alternating pattern of weak and strong syllables. We must then choose P3, the direction of construction, which may be either from left to right or from right to left. In addition, we must select P4, which allows each foot to be left-dominant or right-dominant. I present two illustrative examples: Maranungku (Tryon (1970)), in (7), has P3 [Left] and P4 [Left] - i.e. left-dominant feet constructed from the left; and Warao (Osborn (1966)), in (8), has left-dominant feet constructed from the right - i.e. P3 [Right], P4 [Left]:

(7) P3 [Left], P4 [Left]: Left-dominant Feet Built from the Left (Maranungku)


(8) P3 [Right], P4 [Left]: Left-dominant Feet Built from the Right (Warao)



Word trees have been omitted from the examples; however, they would be constructed on the feet, with main stress devolving upon the strongest vowel in the strong foot. Note one additional fact about Warao: in (8b), the first syllable, being alone in a foot, ought to receive a secondary stress; compare the last foot in (7a). Warao apparently does not tolerate stress clashes; hence, the non-branching foot is defooted, and the first syllable does not receive a stress. Its stresslessness is due to setting the defooting parameter, P9, to [Yes]. More precisely, destressing is controlled not by the single parameter P9, but by a series of parameters which make up another component of the phonology that is activated if P9 is set to [Yes].

The feet in (7) and (8) are not affected by the internal structure of the rimes on which they are constructed; in such languages, foot construction, and hence stress, is said to be insensitive to quantity (QI) - select P5 [No]. However, many languages have quantity-sensitive (QS) stress systems, which means that they distinguish light and heavy syllables. In such languages, a heavy syllable may not occupy a weak position in a foot:

(9) Quantity Sensitivity (QS)
A heavy syllable may not occupy a weak position in a foot.

The criterion for what counts as a heavy syllable is itself subject to parametric variation, controlled by the setting of P6, as indicated in (10):

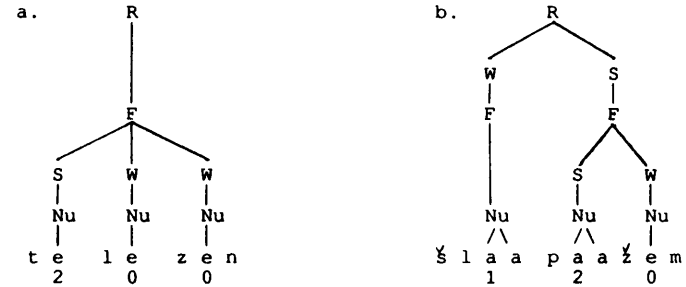
(10) QS: Parametric Variation
a. P6 [Rime]: CV is light, CVC and CVV are heavy.
b. P6 [Nucleus]: CV and CVC are light, CVV is heavy.

A short open syllable is always light, a syllable containing a long vowel or diphthong is always heavy, while a closed syllable with a short vowel can be either, depending on the setting of P6.

It follows from (9) that in quantity-sensitive stress systems, heavy syllables are normally stressed. The presence of heavy syllables can considerably disrupt the smooth alternation of stresses we have observed up to here.

The above examples have involved binary feet. In such languages, main stress can never fall more than a certain number of syllables from the edge of a word. Languages with unbounded feet, by contrast, display a different pattern. In these languages, quantity is the most important factor in stress assignment, while position is secondary. The typical formulation of main stress in such languages is: stress the [rightmost/leftmost] heavy syllable. This type of pattern can be accounted for by positing quantity-sensitive unbounded feet; an example is Eastern Cheremis (Kiparsky (1973)), shown in (11):

(11) P1 [Right]: Main Stress on the Right, Unbounded Feet (E. Cheremis)[2]




Eastern Cheremis has left-dominant unbounded feet. In words with no heavy syllables, as in (11a), a single foot extends over the whole word, as there is no preset limitation on its size. In (11b), however, the second syllable is heavy, and so must begin a new foot.

We have now seen samples of the data and the theory of UG we are assuming. The grammar of stress of a language will now just be some setting of the parameters of UG. In addition to these parameters, UG must also incorporate a learning theory which specifies to the learner how to set the parameters. For example, what in the data tells a learner that a system is quantity sensitive or not, or how to build feet? This is what Robin Clark (1989, 49) has called the Selection Problem: "given any piece of evidence, how does the learner decide which parameter is the appropriate one to set?" The Selection Problem is particularly acute in metrical phonology, because the evidence consists of stress patterns which bear on all the parameters together. Observed stress patterns are the result of the interaction of all the parameters in (5), and so unpacking them in order to determine which parameter is responsible for which bit of the pattern is not a trivial problem.

## 4. A Cue-based Learning Model

Dresher and Kaye (1990) propose a learning model for metrical theory where every parameter (or, in some cases, group of parameters) is associated with a specific cue in the input data, the presence of which triggers a change in the setting of its parameter. Such a conception of parameter setting is implicit in much work in parametric grammar (see for example Roeper and Williams (1987)), and it is a natural, though not a necessary, way to think about how systems of parameters are learned. In what follows, I will review some results that emerge about the relationship of cues to parameters, and consequences for learning models.

## 5. Subsets

One result concerns the status of subsets and the Subset Principle, formulated by Berwick (1985). Consider, for example, a simplified version of the Pro-Drop (or Null Subject) parameter. A language which does not allow Pro-Drop (say English) requires that all sentences have a lexical subject; in a language which allows Pro-Drop (say Italian), sentences may appear without overt subjects. If Pro-Drop is limited to just these facts, then we observe that the set of sentences we can generate with no Pro-Drop is a subset of the set of sentences we can generate with Pro-Drop:[3]

(12) a. *Pro-Drop Parameter (simplified)*

```
NO   ---  --  John walks  >---YES
(English)     e walks      (Italian)
```

b. *Subset Principle (Berwick 1985)*
Choose the subset language as the default parameter setting.

If a learner starts out assuming the superset value of Pro-Drop, i.e. [Yes], no positive evidence will tell it that it has overgenerated. It follows, then, that the default setting ought to be the subset, in which case positive evidence can drive the learner to change to the superset value if required. This is the Subset Principle.

From examples like these it is easy to suppose that the subset relation applies to languages, but this is not correct: relevant subsets are defined with respect to cues. This point can be simply demonstrated by considering the parameter which determines whether stress in a language is sensitive to quantity (QS) or not (QI). Recall that in a QI stress system, all syllables are counted as equal as far as stress assignment goes, whereas in a QS system, syllables are either heavy or light. Now let us consider the relation between QS and QI systems. If we look only at the output forms, there is no subset relation between them: a QI system will generate one set of stressed words, while a QS system will generate another, perhaps overlapping, set.

But consider the matter from the point of view of a learner: how does one determine if a stress system is QI or QS? The answer to this question is not obvious. In fact, there are at least two solutions we might entertain, and both of them lead to a subset relation, though a different one in each case. Suppose that $LT_s$ is a learning theory in which the cue for fixing this parameter involves a window that is one syllable in length. In QI systems, where there is no distinction between heavy and light syllables, all syllables can be either stressed or unstressed, depending on their position in the word. Similarly, light syllables in QS systems are stressed or not depending on position; heavy syllables, however, may not usually be unstressed:

(13) a. Syllable Types in QI and QS Systems

| | *QI Systems* | *QS Systems* |
|---|---|---|
| Stressed | Heavy, Light | Heavy, Light |
| Unstressed | Heavy, Light | - , Light |

b. $LT_s$: QS languages are subset of QI languages
Cue: Look for an unstressed heavy syllable.

From the point of view of syllable types found, QS systems are clearly subsets of QI ones. The latter allow unstressed heavy syllables whereas the former do not. This implies, then, that the initial default value of this parameter in $LT_s$ should assume QS.

This particular approach to determining quantity sensitivity is not reliable, however, because there are various ways in which heavy syllables can come to lack stress even in a QS system. A heavy syllable may be extrametrical, in which case it will be stressless. Or, it may be subject to destressing, typically if it clashes with an adjacent stressed syllable. Since the source of the lack of stress is not obvious to a learner, the presence of such unstressed heavy syllables could fool it into assuming that a language has a QI stress system. In general, the lack of a stress on a given syllable is rarely a reliable cue.

There is, however, another learning theory available for dealing with these cases. Consider $LT_w$, where the test window is extended to include the whole word. Now a different cue becomes available: two words having the same number of syllables but different stress patterns can be taken as a positive cue for QS. This is because in QI systems words with the same number of syllables are all alike from the point of view of the metrical parameters. In QS systems, by contrast, there is a distinction between heavy and light syllables. We thus have the equivalence classes of word types shown in (14):

(14) a. Word Classes in QI and QS Systems

| | *QI: Syllable = S* | *QS: Syllable = H or L* |
|---|---|---|
| 2 Syllable Words | {SS} | {LL} {HL} {LH} {HH} |
| 3 Syllable Words | {SSS} | {LLL} {HLL} {LHL} {HHL} {LLH} {HLH} {LHH} {HHH} |
| 4 Syllable Words | {SSSS} | {LLLL} {HLLL} {LHLL} ... |

b. $LT_w$: QI languages are subset of QS languages
Cue: Look for two words with the same number of syllables but different stress patterns.

For example, in a QI system there is only one type of two-syllable word, while in a QS system there may be up to four, and, in general, up to $2^n$ types of words for words with n syllables. If we calculate subsets with regard to the number of equivalence classes established for words with n syllables, we see that the QI classes are a subset of the QS classes. Indeed, in $LT_W$, we must suppose that QI is the default case, since positive evidence - the existence of words with the same representations but different stress patterns - exists only for QS. A learner which starts by assuming QS will not receive positive evidence contradicting it. Rather, it would have to notice that all equivalence classes consisting of words of n syllables have the same stress pattern. Thus, $LT_W$ requires a different subset relation.

This cue is much more reliable than the syllable-based one discussed earlier, but again the learner has to be protected from false positives that could fool it. One source of such false positive cues is the presence of exceptional words. For example, stress in Polish is almost always on the penultimate syllable, except in some exceptional words where it is on the antepenultimate. We would not want the existence of a contrasting pair like *kotlína* and *ópera* to trigger the learner to assume QS, because the antepenultimate stress on *ópera* is exceptional, and has nothing to do with its syllable structure:

(15) Contrasting Stress Contours (that do not indicate QS)
- a. Polish: kotlína (regular) ~ ópera (exceptional)
- b. English: récord (noun) ~ recórd (verb)

The learner can be protected from such cases if we suppose that it must see some threshold number of a particular pattern before it considers it to be a real pattern.

Another potential confound occurs in languages where stress applies differently in different morphological classes (e.g. English noun-verb pairs like *récord, recórd)*. Such cases are presumably detectable to the learner, though we have not worked out exactly how the morphology would intervene.

$LT_W$ has another attractive property which $LT_S$ lacks. If we assume the default state to be QI, we posit that the stress learner begins with a coarse set of representations, treating all syllables as equal. The discovery that this analysis does not suffice to capture distinctions relevant to stress drives it to enrich the representations, analyzing syllables further into heavy and light. In $LT_S$, we must rather assume that the learner begins with a richer set of representations, and needs positive evidence to ignore distinctions that are already represented in its grammar, a counterintuitive procedure. We might propose in general that phonological representations are subject to this principle of increasing differentiation in the course of acquisition: distinctions are added, not subtracted.

## 6. Cross-parameter Dependencies

Let us turn now to consider some of the other parameters. When we try to determine a consistent cue for each parameter, we find that there exist pervasive cross-parameter dependencies not represented by the flat list of parameters given in (5)[4]. Parameters must be set in a particular order, indicated in (16); note that P9 is omitted, because setting the parameters of destressing follows all the stress parameters:

(16) Order in Which Parameters are Set
- P5: Feet are quantity-sensitive (QS) [Yes/No]
- P6: Feet are QS to the [Rime/Nucleus]
- P10: Feet are noniterative [No/Yes]
- P8L: There is an extrametrical syllable [No/Yes] on the Left
- P8R: There is an extrametrical syllable [No/Yes] on the Right
- P2: Feet are [Binary/Unbounded]
- P1: The word-tree is strong on the [Left/Right]
- P7: A strong branch of a foot must itself branch [No/Yes]
- P3-P4: Feet built from the [Left/Right] and strong on the [Left/Right]

To see what these dependencies are like, let us look again briefly at parameters P2 and P1. How could a learner decide whether feet are bounded or unbounded? We observed in (11) that a left-headed unbounded foot is anchored by an initial syllable, whether it is light or heavy, and then continues until it is interrupted by a heavy syllable. Hence, only heavy syllables or left-peripheral light syllables can be the heads of left-headed unbounded feet. It follows then that if a language has unbounded feet, only one peripheral light syllable in a word is eligible to receive stress. Therefore, if the learner finds a nonperipheral stressed light syllable, or stress on both a rightmost and leftmost light syllable (not necessarily both in the same word), it can conclude that feet are bounded, i.e. binary[5]. Therefore, the learner should start out assuming that feet are unbounded.

The cue for main stress utilizes the notion of a foot-sized window at the edge of a word. Main stress is confined to a peripheral foot; therefore, it should consistently appear in one of the windows which correspond to a peripheral foot. Such a cue presupposes knowledge of P5 and P6 (QS or QI) as well as P2 (Binary or Unbounded feet), for these parameters give the dimensions of the window[6].

These dependencies bear on the batch mode and incremental learning models mentioned before. A batch processor can simply set the parameters in the indicated order, and it will not go wrong, because it has before it all the relevant data. Therefore, by the time it has to set P1, the parameter for main stress, it will have already correctly set the values of P5, P6, and P2. Such a

learner is quite powerful: knowing that it has seen all the relevant data gives it a considerable advantage over an incremental learner. But for that reason it also appears less realistic, so let's look at how an incremental learner would deal with these dependencies.

Because an incremental learner is setting parameters on the fly, perhaps before all the relevant data have been encountered, it is important that it not make false moves from which it may not be able to recover. It is generally assumed that adhering to the Subset Principle will ensure that such a learner will not get trapped in a superset from which it would be difficult to retreat. This is true for one parameter at a time, but it is not true for a set of interacting parameters. Thus, suppose that the learner is trying to learn Eastern Cheremis (11), which is actually QS, where only a branching nucleus counts as heavy, has unbounded feet, and has main stress on the right. Suppose it has figured out that the language is QS, but has not yet seen any evidence to set P6 to [Nucleus], so P6 remains in its default state [Rime], treating closed syllables as heavy in addition to long vowels. Now, there is no problem here for P6, since the learner will eventually run across the crucial evidence to change it. In this sense, the default setting is safe. But it is not safe for the other parameters. Recall that P1, the parameter that assigns main stress, looks for main stress in a foot-sized window at the edge of a word. But when it encounters a word like *télezen*, it applies the incorrect settings of parameters P6 and P2 to discover, incorrectly, that main stress is on the left, not realizing that the whole word comprises only one foot. Many variations on this theme can be produced, all of which show that incorrect default settings can be deadly to dependent parameters.

What recourse does an incremental learner have in such situations? To keep from making a false move in setting parameters P2 and P1, it has to be sure that its values for P5 and P6 are correct. We might propose that it should hold off setting any parameters that depend on a parameter, say P5, until it is sure about the setting of P5. This is feasible if P5 is QS, because the change from QI to QS is only triggered by positive evidence, and since the parameter is binary, no further changes will occur. But what if the language being learned is really QI? In that case, no positive evidence will ever come to confirm the setting of P5.

One proposal, then, is that we set some time limit t for setting each parameter: if after time t there has been no positive evidence to move to the marked setting, then we freeze the default setting. But now the incremental learner has in fact become a batch mode learner. A second possibility is the following: allow the incremental learner to set parameters as before, but impose the principle that when a parameter changes its value, all parameters that depend on it must revert to default. No more refined procedure is possible, since we assume that the learner does not remember why it set some parameter to a particular value. Even if it could remember the crucial forms, it would not be easy to unravel the reasoning that led to every change. In the case at hand, then, the learner could set parameters P2 and P1 to various incorrect marked settings while P6 is in default state, but would have to wipe these out as soon as P6 changes.

While this modification preserves the incremental model, it brings it closer to the batch mode learner, because until P5 is at its correct setting, nothing it does with regard to the dependent parameters really matters. If this model is correct, it has a number of consequences for the developmental problem: first, this model provides a mechanism for creating superset errors in the course of acquisition which do not require negative evidence to retreat from; and second, though it assumes stages of development, it also supports the idealization of instantaneous acquisition, since erroneous parameter settings at intermediate stages of acquisition would have no effect on the grammar of later stages.

## 7. Interactions between Components: Syllable Structure and Stress

Until now we have been looking at examples of parameter interactions within one component of the grammar. Groups of parameters also interact with other components. These interactions can be more or less intricate, and I will discuss a few illustrative examples.

An interesting case concerns the relation of stress and syllable structure: we have seen that the parameters of stress depend on representations of syllable structure, which must therefore be established before stress can be learned. In some languages, a final segment (or consonant) may be extrametrical, with the result that what looks like a heavy syllable (CVC or CVV) is in fact treated as a light syllable (CV). Consider the sample forms in Hindi in (17) (I adopt here a version of the facts which come from Mohanan via Hayes (1981); Gupta (1987) gives a different version):

(17) Hindi Word Stress

| | | |
|---|---|---|
| a. ka.MAAL | b. KA.mal | c. in.SAA.ni.yat |
| raa.JIIV | RAA.jan | PA.ri.ci.taa |

The basic stress pattern is as in Eastern Cheremis: the rightmost heavy syllable receives stress; if there is no heavy syllable, the initial syllable receives stress. The forms in (a) have final superheavy syllables, which receive stress as expected; however, in (b) and (c), final heavy syllables do not receive stress. We can account for this fact if final segments are extrametrical; in that case, only the final syllables in (a) remain heavy when their final segment is discounted.

How could a learner conclude that the final segment in Hindi is extrametrical? There may be a clue from the distribution of syllable structure:

if superheavy syllables occur only in final position, that is evidence that there is an extra position of some sort there, say an appendix; and if it is a universal that all final coda segments must go into the appendix if there is one, then extrametricality would follow from universal principles, and would not have to be learned specially. But suppose that this is not universally the case. Then the only evidence for segment extrametricality is the patterning of stress. But since we need syllable representations to be established before the stress pattern can be acquired, we face the threat of circularity in the learning path.

This circle is not vicious, however. As shown by Nicholas Brownlow (1988), segment extrametricality can be established using the facts of stress, prior to an analysis of the stress system. The relevant cue is an extension of the procedure we used to determine whether or not a system is QS. If we sort the words in (17) into word classes on the basis of their syllable structures, we obtain the classes in (18):

(18) Hindi Word Classes (with no segment extrametricality)

| | | |
|---|---|---|
| a. L H (0 2) | b. L H (2 0) | c. L H L H (0 2 0 0) |
| H H (0 2) | H H (2 0) | L L L H (2 0 0 0) |

These representations produce a contradiction between (a) and (b), however, since we now have words with the same representations, but with different stress patterns. Brownlow proposes that segment extrametricality parameters intervene here. The representations are returned to the syllable parser, which adjusts them to include segment extrametricality, deriving the classes in (19):

(19) Hindi word classes with segment extrametricality

| | | |
|---|---|---|
| a. L H (0 2) | b. L L (2 0) | c. L H L L (0 2 0 0) |
| H H (0 2) | H L (2 0) | L L L L (2 0 0 0) |

The representations in (19) contain no contradictions, in that there are no longer identical forms with conflicting stress contours. It follows, then, that segment extrametricality is deducible from the facts of stress before any metrical parameters have been learned. The stress learner can now proceed on the basis of the representations in (19).

This learning theory makes an empirical prediction, albeit one that is difficult to test. It claims that conflicts of the sort found in (18a,b) are the crucial cues to diagnosing segment extrametricality. Notice that the quadri-syllable words in (c) do not exhibit such conflicts; if the learner were exposed to only the forms in (c), it would not detect segment extrametricality, and could run into difficulties in trying to set the metrical parameters. In these circumstances we might expect some change in the grammar. We expect, then, that there could be situations where a linguist could deduce that a grammar has a certain property, but where a learner, lacking the crucial cue, does not arrive at the same solution.

## 8. Interaction between Components: Stressing and Destressing

Another interesting case of interaction among components is that between stressing and destressing. Destressing is a common phenomenon which can create disturbances to expected stress contours which can trip up an unwary learner. By way of illustration, let us return to Warao, in (8). Suppose that a learner has already determined that it has quantity insensitive (QI) binary feet, and let us consider how it might now arrive at the correct value of parameters P3 and P4: i.e., how could it determine that Warao has left-dominant feet built from the right edge of the word?[7]

There are four configurations generated by P3 and P4: feet can be built from the left or from the right, and the feet can be left-dominant or right-dominant. Each one of these four configurations corresponds, in its ideal form, to a characteristic pattern of alternating stresses (we can ignore the difference between main and secondary stresses here). Suppose, then, that we simply try to fit the stress patterns of the input words to each of these patterns in turn until we get a consistent fit. Thus, the word in (8a) is consistent also with left-dominant feet built from the left; but (8b) is not. Forms like (8b) lead the learner to reject this configuration, and try another one.

The problem with this cue is that the correct answer - left-dominant feet built from the right - does not match the pattern of (8b), either. The expected pattern is given in (20a):

(20) Left-dominant Feet Built from the Right (V=stressed, v=unstressed)

| | *Even* | *Odd* |
|---|---|---|
| a. Basic pattern: | V v V v V v | V V v V v |
| b. Defooting 1: (Warao) | V v V v V v | v V v V v |
| c. Defooting 2: (Garawa) | V v V v V v | V v v V v |

In words with an odd number of syllables, we expect the first syllable to be stressed; in (8b), it is unstressed. Now, we know why that is: it's because the initial syllable is defooted, causing the loss of its stress. This change in the pattern, though relatively minor, is enough to derail a learner looking for the pattern of (20a). The problem is made worse by the fact that destressing can apply in several different ways. Thus, in languages like Garawa (Furby (1974)) and Italian, a stress clash of the type (20a) is resolved as in (20c), by keeping the stress on the initial syllable and destressing the second syllable:

(21) P3 [Right], P4 [Left]: Left-dominant Feet Built from the Right (Garawa)[8]

```
a.      F              F
       / \            / \
      S   W          S   W
      |   |          |   |
      Ri  Ri         Ri  Ri
      /\  /\         |   |
   w a t j i m   p a ŋ u
   2   0         1   0

b.  F     F         F         F         F
    |    / \       / \       / \       / \
    .   S   W     S   W     S   W     S   W
    |   |   |     |   |     |   |     |   |
    Ri  Ri  Ri    Ri  Ri    Ri  Ri    Ri  Ri
    |   |   /\    |   /\    |   |     |   |
  n a   ř i ŋ i n   m u k u n   j i n a   m i ř a
  2   0   0         1   0       1   0     1   0
```

It turns out that the stress learner can abstract away from destressing completely: an effective cue for foot construction simply ignores unstressed syllables in strong positions; then, only the presence of a stressed syllable in what should be a weak position will count as a violation:

(22) A Robust Cue for P3 and P4

For each setting of P3 and P4, scan across the word; the presence of a stressed syllable in what should be a weak position rules out a setting. The presence of an unstressed syllable in a strong position does not count.

Separating out destressing, and ordering it after stressing, accords well with the way destressing has usually been thought of, as a series of operations that apply to the output of the stress rules. It is interesting, therefore, that the parameters of stress are learnable in the absence of knowledge about the details of destressing.

## 9. Dealing with Failure

A further set of consequences arise from the fact that parameters and modules interact, and these have to do with how the learning model deals with failure, or apparent failure. In the learning model described by Dresher and Kaye (1990), parameters are fixed in turn on the basis of local cues; the resulting set of parameter values is then passed to an applier, which uses them to construct metrical trees and derive stress patterns. The applier can also function as a tester, checking that the stress patterns it obtains using the acquired parameter values match input stress patterns. When they match, the applier can verify that the acquired grammar is correct.

The criteria for what counts as a match must allow for the fact that, in the first pass, the tester does not have any information about destressing. To return to our example languages, we have seen that the actual stress pattern of five-syllable words in Warao is [0 1 0 2 0], as in (23a) (Warao has main stress on the right); however, the parameters sent by the parameter setter to the applier produce the pattern [1 1 0 2 0]:

(23) Acceptable Mismatches between Input and Derived Forms

| | *Input* | *Derived* | *Destress* |
|---|---|---|---|
| a. Warao | 0 1 0 2 0 | 1 1 0 2 0 | 0 1 0 2 0 |
| b. Garawa | 2 0 0 1 0 | 2 1 0 1 0 | 2 0 0 1 0 |

Similarly, the the actual pattern for a five-syllable word in Garawa, which has main stress on the left, is [2 0 0 1 0], though the learner derives [2 1 0 1 0]. In each case, the discrepancy (0 instead of 1 stress) is ignored by the tester, which considers the forms to match. The details of destressing are then learned by another component dedicated to that, which ultimately produces the right output.

While the discrepancies between derived and input forms involved in this type of destressing are relatively minor, these examples show that the role of the tester in a cue-based learning system is not straightforward: discrepancies between input and derived forms do not necessarily indicate that the parameters are incorrectly set; the fault could lie elsewhere.

Suppose now that some parameters have in fact been incorrectly set, perhaps because some cues failed. Although parameters are set on the basis of local cues, their effects, when combined with other parameters, are global, and often unpredictable. Therefore, if certain parameters are incorrectly set, there is no obvious way for the tester to determine where the error lies. Even worse, there is no clear indication in which direction to modify the system so as to achieve a better result. Therefore, it is not clear how parametric systems of this type can be learned on the basis of models which crucially rely on some overall measure of goodness-of-fit.

One interesting model of this kind has been proposed by Robin Clark (1990). Clark proposes a genetic algorithm for setting syntactic parameters. In this type of model, a number of candidate hypotheses are selected at random, where each candidate consists of a full set of all relevant parameters. The candidates are applied to data, and their relative success is determined by a measure of goodness of fit. In Clark's case, the measure consists of how many nodes each candidate hypothesis succeeds in parsing. After each trial, the more successful candidates are selected and the less successful are weeded out in a Darwinian process. A new generation of candidate hypotheses is derived from

the successful ones, by randomly changing the values of a small number of parameters, and the process repeats for a number of generations, until the system converges on the correct set of parameter values.

The problems of adapting this system to our case have already been raised above: we have no measure of overall goodness-of-fit; even if we did, it is not clear how such a measure could guide a learner to modify its parameter settings in the right direction.

It follows that a cue-based learning system has little alternative but to rely on its cues, and not on a tester which checks for overall correctness. That is, parameters set on the basis of cues must be assumed to be correct, however great the apparent mismatch this leads to. By the time an incorrect parameter value has been passed to the tester, it is already too late to fix it. The time to discover a problem is at the stage where cues are in play. It is important, therefore, for a learner to be able to spot problems locally.[9]

It is instructive to observe how the stress learner deals with stress patterns that it is guaranteed not to be able to learn. An example of such a language is Cayuvava (Key (1961); Levin (1985)), which exemplifies a kind of ternary pattern:

(24) Cayuvava Ternary Pattern

```
/   /       /         /    \    /      \    /
V v, V v v, v V v v, v v V v v, V v v V v v, v V v v V v v, ...
```

However this pattern is analyzed, a learner limited to the parameters in (5) is simply incapable of learning this stress system. It is interesting, then, that our cue-based learner is nevertheless able to set most of the Cayuvava parameters correctly. It learns that Cayuvava stress is not sensitive to quantity, that main stress is on the right, and that feet are bounded. It is only when it tries to determine what kind of feet there are (i.e. parameters P3 and P4) that it runs into trouble. What is encouraging, though, is that it realizes that there it has a problem here. Indeed, it is just at this point that it is missing a parameter that expands the type of bounded feet, and it is just at this point that the learning path must be elaborated.

In this case, we can fix the learner's UG and supply it with the required parameter. Other cases may not admit of the same type of solution. It is possible, for example, that some languages have truly idiosyncratic versions of certain parameters, which cannot be built into UG. While it is not clear how to modify the learner so as to enable it to improvise on its own in such cases, the ability to detect where a problem lies is a key element to an eventual solution.

## 10. Rules and Representations

The kind of interaction between components we have observed above may extend to other areas of the phonology. Recent work in phonology has tended to place the burden of explanation on the system of representations rather than on rules, raising the prospect that acquisition of phonology might be mostly a matter of acquiring representations; the operation of rules would then be largely determined, and there would be little left to learn. While this is an appealing scenario, there is again a risk of circularity, since the richer representations become, the more they incorporate variable elements, i.e. parameters which must be learned on the basis of evidence supplied by rules. A case in point is Glyne Piggott's analysis of a type of nasal harmony: he proposes that cross-language differences in the operation of nasal harmony derive from the variable specification of segments for a Soft Palate node (e.g. Piggott (1987; 1989)). Jila Ghomeshi (1990) argues that the posited segmental representations cannot be learned without taking into account the facts of nasal harmony; the rule of nasal harmony, however, can be learned after the representations are in place. This interaction between rules and representations resembles that between the metrical parameters and syllable structure. Phonology is thus, as has been said, a system of systems, whose components interact in intricate and structured ways.


### Acknowledgements

This article draws on joint research done with Jonathan Kaye. For a fuller discussion of the learning model and related issues, see Dresher and Kaye (1990). This research was funded in part by grants from the Social Science and Humanities Research Council of Canada (#410-84-0319 and 410-88-0729) and from IBM Canada in partnership with the Centre for Computing in the Humanities at the University of Toronto. I would also like to thank Lise Dobrin and the other organizers of CLS 27 for their assistance.


### Notes

1. See Hayes (1981) and Halle and Vergnaud (1987) for further details and sources concerning the languages mentioned here and below.

2. We assume secondary stresses in E. Cheremis, though we have no information as to whether they appear on the surface. If they do not, the learning problem is more difficult, since the learner is presented with less evidence as to the distribution and relative prominence of feet. Our model can learn the correct placement of main stress in these cases whether or not there are secondary stresses.

3. For sake of the illustration, I omit here other possible cues or effects of Pro-Drop which could substantially alter this picture; see Jaeggli and Safir (1989) for discussion.

4. Wexler and Manzini (1987) propose what they call the 'Independence Principle', the intention of which is to guarantee that parameters can be set independently of each other. It should be noted that the dependencies discussed here, though technically adhering to the Independence Principle, nevertheless show that parameters cannot be correctly set in this way. It follows that the Independence Principle does not guarantee parameter independence in the wider sense.

5. The actual operation of this cue can be complicated by extrametricality, which creates uncertainty as to where the operative edge is.

6. Again, I omit complications caused by extrametricality.

7. Depending on how the cue for extrametricality is ordered, we could also analyze Warao as having right-dominant feet built from right to left, with a final extrametrical syllable. We assume, contrary to Halle and Vergnaud (1987, 19n.), that a solution which makes no recourse to extrametricality is preferred in this type of case; however, this assumption is not crucial to the argument.

8. Hayes (1981, 54-55) proposes a different analysis which requires different parameter settings for main and secondary stress. This complexity is not required in our analysis; a similar analysis is proposed by Halle and Vergnaud (1987, 19). We ignore here the difference between secondary and tertiary stresses.

9. See Berwick and Weinberg (1984, 231ff.) for discussion of this idea in the context of parsing and acquisition of syntax.

## References


Berwick, R. C. (1985) *The Acquisition of Syntactic Knowledge*, MIT Press, Cambridge, Mass.

Berwick, R. C. and A. S. Weinberg (1984) *The Grammatical Basis of Linguistic Performance*, MIT Press, Cambridge, Mass.

Brownlow, N. D. (1988) "A Computer-based Learning Theory for Final Segment Extrametricality," ms., University of Toronto and Carnegie Mellon University.

Chomsky, N. (1975) *Reflections on Language*, Pantheon, New York.

Chomsky, N. (1981) "Principles and Parameters in Syntactic Theory," in N. Hornstein and D. Lightfoot, eds., *Explanation in Linguistics*, Longman, London.

Chomsky, N. and M. Halle (1968) *The Sound Pattern of English*, Harper & Row, New York.

Clark, R. (1989) "On the Relationship between the Input Data and Parameter Setting," *NELS* 19, 48-62.

Clark, R. (1990) "Some Elements of a Proof for Language Learnability," ms., University of Geneva.

Dresher, B. E. and J. D. Kaye (1990) "A Computational Learning Model for Metrical Phonology," *Cognition* 34, 137-195.

Furby, C. (1974) *Garawa Phonology (Pacific Linguistics Series A, #37)*, Australian National University, Canberra.

Ghomeshi, J. (1990) "A Parametric Approach to Nasal Harmony," ms., University of Toronto.

Gupta, A. (1987) "Hindi Word Stress and the Obligatory Branching Parameter," *CLS* 23, Part Two, 134-148

Halle, M. and G. N. Clements (1983). *Problem Book in Phonology*, MIT Press, Cambridge, Mass.

Halle, M. and J.-R. Vergnaud (1987) *An Essay on Stress*, MIT Press, Cambridge, Mass.

Hammond, M. (1984). *Constraining Metrical Theory: A Modular Theory of Rhythm and Destressing*, Ph.D. dissertation, UCLA. Distributed by IULC, Bloomington, Indiana. Published 1988, Garland Publishing, New York.

Hayes, B. P. (1981) *A Metrical Theory of Stress Rules*, Ph.D. dissertation, MIT. Distributed by IULC, Bloomington, Indiana.

Hornstein, N. and D. Lightfoot (1981) "Introduction," in N. Hornstein and D. Lightfoot, eds., *Explanation in Linguistics*, Longman, London.

Jaeggli, O. and K. J. Safir, eds. (1989) *The Null Subject Parameter*, Kluwer, Dordrecht.

Key, H. (1961) "Phonotactics of Cayuvava," *IJAL* 27, 143-150.

Kiparsky, P. (1973) "'Elsewhere' in Phonology," in S. R. Anderson and P. Kiparsky, eds., *A Festschrift for Morris Halle*, Holt, Rinehart and Winston, New York.

Levin, J. (1985) "Evidence for Ternary Feet and Implications for a Metrical Theory of Stress Rules," unpublished ms., University of Texas, Austin.

Lightfoot, D. (1989) "The Child's Trigger Experience: Degree-0 Learnability," *Behavioral and Brain Sciences* 12, 321-375.

Osborn, H. (1966) "Warao I: Phonology and Morphophonemics," *IJAL* 32, 108-123.

Piggott, G. (1987) "On the Autonomy of the Feature Nasal," *CLS* 23, Part Two, 223-238.

Piggott, G. (1989) "A Parametric Approach to Nasal Harmony," in H. van der Hulst and N. Smith, eds., *Features, Segmental Structure and Harmony Processes*, Foris, Dordrecht.

Roeper, T. and E. Williams, eds. (1987) *Parameters and Linguistic Theory*, D. Reidel, Dordrecht.

Tryon, D. T. (1970) *An Introduction to Maranungku (Pacific Linguistics Series B, #14)*, Australian National University, Canberra.

Wexler, K. and M. R. Manzini (1987) "Parameters and Learnability in Binding Theory," in T. Roeper and E. Williams, eds., *Parameters and Linguistic Theory*, D. Reidel, Dordrecht.