

Chapter 5: The Case of English Metrical Phonology

5.1 The Unambiguous Data Filter

We have just seen an argument for the necessity of an unambiguous data filter, using evidence from syntactic language change modeling. The motivation for the unambiguous data filter is that unambiguous data are the most informative data in a noisy data set; learning from informative data leads to convergence on the correct target state, given standard statistical learning techniques. So, unambiguous data are highly desirable since, once identified, they completely determine the choice in the hypothesis space for the learner.

Yet, I have also outlined why identifying unambiguous data is not a simple task. In particular, the identification of unambiguous data becomes considerably harder in a system with multiple interacting parameters. One might then wonder if it's even possible for an unambiguous data filter to be successful in a more complex domain, since this situation makes unambiguous data much sparser. In short, even if unambiguous data are desirable for learning, is it feasible to use an unambiguous data filter for a system with multiple interacting parameters?

For this reason, I turn now to the domain of metrical phonology, which has several interacting parameters, some of which have one or more sub-parameters, for a total of 9 interacting parameters.³⁶ Interacting parameters provide an additional challenge for language learners because the order in which these parameters are set by the learner (sometimes called the *learning path* (Dresher, 1999)) determines whether the learner will converge on the correct adult parameter values. If the choices are not made in the correct order, the child can misconverge. A reasonable question is whether we can discover *principled* metrics that allow the child to both find unambiguous data in the input and converge on an appropriate order of parameter-setting, given the noisy situation of multiple parameter interaction.

5.1.1 Two Methods For Identifying Unambiguous Data

First, we must ensure learner convergence on the adult system by uncovering the space of possible methods there are to discover sufficient unambiguous data in the correct distributions so that the learner converges on the adult system. Two methods have been proposed to implement an unambiguous data filter: a method that uses the domain-specific representation of cues (Dresher, 1999; Lightfoot, 1999 (see previous chapter)) and a method that uses the domain-specific learning procedure of parsing (Fodor, 1998b, 1998c; Sakas & Fodor, 2001).

A cue, according to Dresher (1999), is a “specific configuration in the input” that reflects a “fundamental property” of the particular parameter value it is a cue for. Moreover, a cue is local in the sense that a learner uses the cue “without regard to the

³⁶ This is much closer to the complexity that is purported to exist in the syntactic domain. A recent implementation by Sakas (2003) has 13 interacting parameters, though this is only a fraction of the parameters posited for the adult syntactic system.

final result”, so that the learner is “not trying to match the input”. The presence of a cue unambiguously favors one hypothesis (i.e., parameter value) in the hypothesis space over another.

The parsing method relies on the learner using the parsing strategies already available for language comprehension (Fodor, 1998b, 1998c; Sakas & Fodor, 2001). The learner tries to analyze a data point with “all possible parameter value combinations” in the hypothesis space given by universal grammar. The learner is, in effect, conducting an “exhaustive search of *all* parametric possibilities” (Fodor, 1998). If only a single parameter value for a given parameter is ever present in the successful parses for a particular data point, that data point is considered unambiguous for the parameter value. Data points that can be parsed with multiple parameter values are considered ambiguous. These ambiguous data points would then be filtered out of the learner’s intake so that the update procedure is not activated when encountering them.

We will see that both these methods are successful at identifying sufficient unambiguous data to converge on the correct adult metrical phonology parameter values for English, a non-trivial task given the interactive nature of the parameters and the noise in the English data set. However, each method requires the learner to have different constraints on the order of parameter-setting.

5.1.2 Stipulation vs. Principled Derivation

Another relevant question is what pieces of each of these methods must be stipulated and what pieces can be derived in a principled fashion. As we will see, the cues method requires that we stipulate pre-specified knowledge in the form of the cues the learner uses to identify the unambiguous data. However, the constraints on the order of parameter-setting for the cues method result from properties of the learning system. In comparison, the parsing method does *not* need to stipulate extra information to identify unambiguous data – the process of assigning structure to input is already used for language comprehension. Yet, we will see that most of the constraints on the order of parameter-setting, in contrast to the cues method, must be stipulated.

5.1.3 Summary

In this chapter, I first explicitly compare the strengths and weaknesses of both the cues and parsing methods. Following that, I test both methods in the complex domain of metrical phonology, which has a set of 9 interacting parameters yielding 2^9 possible languages, and examine the potential of each method to identify sufficient unambiguous data to converge on the adult English parameter values. English is a particularly difficult case since there are actually unambiguous data in the input for the *incorrect* parameter values. Thus, the ability to converge on the correct parameter values in this exacerbated situation is a testament to the power of using either of these methods to identify unambiguous data. We will then see that both these methods can succeed and allow the learner to converge on the adult set of parameter values, providing support for the feasibility of an unambiguous data filter even in a more

complex domain. I will then observe that the constraints on the order of parameter-setting that result from using each of these methods differ, and that the cues method allows us to derive the constraints in a principled manner while the parsing method requires that we stipulate the constraints. Finally, I will argue that both methods have strengths that can be combined and speculate that a more advantageous method of identifying unambiguous data comes from using a limited form of parsing to derive cues for unambiguous data.

5.2 The Cues Method and the Parsing Method

5.2.1 Cues: Strengths and Weaknesses

The cues method of finding unambiguous data has both strengths and weaknesses. Cues are attractive because they make identification of unambiguous data very simple: the data point either matches the cue, or it doesn't (though this, of course, assumes the learner can recognize the cue in the data point). In addition, a cue is designed to match a subpart of the data point, rather than the entire data point. This means the learner can glean information without understanding the structure for the entire data point. For instance, the learner can match a VO word order cue (example (1) taken from (8b) in the previous chapter) to a data point without understanding the structure for all the words in the sentence – the only words that are vital are the ones that correspond to the cue (Object, Verb, and some phrases that function as XP1 and XP2).

(1) VO word order cue: []_{XP1} []_{XP2} ... Verb Object ...

Because learners can extract information from only partial comprehension, cues offer a way to “get off the ground” when they don't know very much about the adult language.

Nonetheless, the cues method also has its weaknesses. Cues for each parameter are, by definition, a representation of domain-specific knowledge and must already be available for the learner or somehow derivable from previously available knowledge. In addition, the specificity of cues must be determined: are cues linear strings, underspecified structural pieces, or something else? Whatever the specificity of the cue, it must also be previously available for the learner or somehow derived. Thus, the cues method requires the learner to be equipped with additional knowledge (cues) to solve the language learning task.

Beyond this, some cues may require the learner to store data over time (perhaps in a summarized form) for comparison (Dresher, 1999). This is usually agreed to be an undesirable requirement in domains such as syntax because the potentially infinite number of sentences yield the possibility of unbounded storage requirements. However, the data storage requirement may fare better in domains such as metrical phonology where there are a finite number of morphemes and stress contours (even if in principle words can have infinitely many syllables).³⁷

³⁷ Note, however, that the generative procedure for assigning stress can assign infinitely many stress contours in the same way that generative syntax can assign infinitely many structures. That is,

Another potential weakness is that cues are heuristic by nature, and so may lead to false positives or false negatives that could have a detrimental effect on learning over time.³⁸ For example, if we examine the cues for the OV/VO word order parameter ((1) and (2), taken from the previous chapter), we will notice that they only take V2 movement into account.

(2) OV word order cue: []_{XP} ... Object Verb ...

However, other grammatical rules in the adult language may also impact the observable data that these cues would match. Heavy Noun Phrase Shift is one such rule: it is a movement rule that shifts an Object that precedes a Verb to a position after the Verb, provided the Object is phonologically “heavy” enough.

As an abstract example, suppose the learner encounters a data point of the form “Adverb Subject Verb Object”. This data point matches the VO cue “XP1 XP2 ... Verb Object”. Nonetheless, it could have been generated by starting with the order “Adverb Subject Object Verb” (which would have matched the Object Verb cue “XP ... Object Verb”) if the Object moved to a position after the Verb via Heavy Noun Phrase Shift. Since the observable data matches the learner’s VO cue, the learner receives a false positive for VO order by using the heuristic cue.³⁹ If this kind of interference happens sufficiently, the learner may not converge on the correct adult value.

As a more concrete example of false positives from cues, consider the case of Kannada word order (data from Tirumalesh, 1996). The basic word order for Kannada is Object Verb (OV). A learner would encounter many examples of this order, as in (3a):

(3a) OV order in Kannada

raamu	dubai-ninda	kumbaLakaayi	tand-id-d-aane
Raam _{Subj}	Dubai-abl	pumpkin_{Obj}	bring-be-npst-3sm_{Verb}
'Raam has brought a pumpkin from Dubai.'			

However, there is a rule in Kannada that will cause the observable order to have the Verb precede the Object (VO order), as in (3b). This rule applies when the meaning of the Object is surprising, and so the Object is moved after the Verb to put focus on its surprising nature.

generative procedures have no trouble coping with data of unbounded length (whether the data are words or sentences).

³⁸ Though this might help explain metrical stress change over time.

³⁹ This case also demonstrates how cues within the same language can conflict. If Heavy Noun Phrase Shift had not occurred because the Object was not “heavy” enough, the observed order would have been “Adverb Subject Object Verb.” This order matches the OV cue. So, the very same language could have cues for both OV and VO word order. The learner presumably must then decide which is more likely to be the base order for the language, given the frequency of the different cues.

(3b) VO order in Kannada, due to surprise at ‘pumpkin’

raamu dubai-ninda tandiddaane kumbaLakaayi
Raam_{Subj} Dubai-abl bring-be-pst-3sm_{Verb} pumpkin_{Obj}
"Raam has brought a pumpkin from Dubai."

Since it is unusual to bring something as inexpensive as a pumpkin from Dubai, the Object ‘pumpkin’ would cause surprise in answer to a question like (3c).

(3c) Question in Kannada that would produce VO order

raamu dubai-ninda eenu tandiddaane
Raam Dubai-abl what bring-be-pst-3sm
'What has Raam brought from Dubai?'

Because the Object causes surprise, it is moved after the Verb to put focus on it. However, if a Kannada learner using cues is unaware of the rule that moves a surprising Object after the Verb, this learner might consider the data in (3b) as an example of VO order since it matches the cue “XP1 XP2 ... Verb Object”. In this way, the learner receives a false positive for VO order in a language whose basic word order is OV. And again, if enough false positives (or false negatives) are encountered, the learner could fail to converge on the correct adult value of the given parameter.

Finally, some cues may require the learner to have default values among the options for a given parameter.⁴⁰ This means that the learner assumes that a given value holds unless there is evidence to the contrary. Note that the learner can still collect evidence to the contrary if data matches the cue for the non-default value, which is quite important if in fact the adult language uses the non-default value. Nonetheless, if default values are required for successful learning, these values are again an example of additional knowledge the learner requires specifically for solving the language learning task. After all, default values are representations of domain-specific knowledge that must be previously available or somehow derivable from previously available knowledge.

As an example of how the learner might derive a default value from previously available knowledge, suppose the candidate hypotheses are in a subset-superset relation, i.e. the set of data points that can be generated by one hypothesis are a subset of the set of data points that can be generated by the other hypothesis (as we saw in chapter 3 with anaphoric *one*). Under this viewpoint, the hypotheses are the opposing parameter values for the given parameter, which is knowledge the learner is assumed to already have available. The Subset Principle (Berwick, 1985; Berwick & Weinberg, 1984) then provides the learner with a principled way to derive the default value: use the subset value.

⁴⁰ This could be instantiated as a hypothesis space with non-uniform prior probabilities. The initial probability distribution would be biased towards the default value.

5.2.2 Parsing: Strengths and Weaknesses

The parsing method for finding unambiguous data has its own strengths and weaknesses. One attractive feature is that parsing is a (domain-specific) procedure the learner already has available, assuming that the learner must come equipped with a procedure that tries to assign structure to the input given the available options (Fodor, 1998b, 1998c). In addition, the parsing method only requires one data point at a time, since it extracts as much information as possible and then proceeds to the next data point. No storage of data over time is required. Third, the parsing method, as discussed, is implemented as a find-all-parses analysis; it is therefore *not* heuristic. It will only find true unambiguous data, given the relevant parameter set. Finally, since all values are used during the find-all-parses analysis of the data point, no default values are required.

While this may seem like an impressive array of strengths, the parsing method also has its pitfalls. First, identification of unambiguous data is a non-trivial task requiring more resources from the learner, either in terms of multiple simultaneous parses stored in memory or in terms of using a sensible guessing strategy (Sakas & Fodor (2001) addresses the question of a sensible guessing strategy the learner might adopt for parsing). If the learner does a full find-all-parses analysis, we must explain how the learner can feasibly do this given finite resources; if a less resource-intensive guessing strategy is used, we must explain why the learner uses this strategy.

Second, if the entire data point cannot be parsed, no information can be extracted for *any* parameter. This makes “getting off the ground” during the initial stages of learning quite difficult, when the learner may not know enough to comprehend the entire data point (see Sakas & Fodor (2001), who acknowledge these problems and propose ways to solve them in scenarios where the adult language data does not contain numerous exceptions that lead to conflicting data points).

Beyond this, if exceptions exist in the input set that violate certain adult parameter values but obey others, those data points cannot be used by the learner since the learner cannot generate a successful parse of the data point.⁴¹ In short, the parsing method does not allow information to be retrieved from subparts of a data point. One way to circumvent this problem would be for the learner to divide the data point into subparts using some sensible strategy (for instance, in syntax, the learner might divide a sentence into matrix and embedded clauses). Nonetheless, we would still need to provide a principled explanation for how the learner knows to divide up the data point in an appropriately helpful way.

Lastly, a learner using the parsing method may have difficulty finding unambiguous data if the relevant parameter set isn't sufficiently restricted (too many possible parameters value sets could fit any given data point). This is perhaps best viewed as a problem of being too exacting about classifying data as unambiguous,

⁴¹ If the learner can't parse the data point, the learner presumably throws the entire data point out for the purposes of learning, classifying it as an exception that will have to be memorized in its entirety. Cues tolerate exceptions much more easily, allowing for anomalous sub-parts to be memorized instead of requiring the entire data point be memorized.

since the consideration of too many options would prohibit any data from being classified as unambiguous.

5.2.3 Summary: Cues vs. Parsing Overview

Both the cues and parsing methods have a large set of strengths and weaknesses, summarized in table 5.1 below. In this chapter, I explore an additional property for comparison: the effect each of these methods has on the learning path within the given domain of metrical phonology. Specifically, I examine the potential set of order constraints for parameter-setting that are generated by using each method to identify unambiguous data.

Property	Cues	Parsing
Easy identification of unambiguous data	True	False
Can get information from sub-part of data point	True	False
Can easily tolerate numerous exceptions in the data	True	False
<i>Is heuristic</i>	True	False
<i>Requires additional prior knowledge for learner</i>	True	False
<i>Requires storage of data over time for comparison</i>	True	False
<i>Requires default values</i>	True	False
Can work even in an unrestricted large set of initial parameters	True	False

Table 5.1. A comparison summary of the properties of the cues and parsing methods. Desirable properties are in **bold**, while potentially undesirable properties are in *italics*.

5.3 The Domain of Metrical Phonology

5.3.1 Why English Metrical Phonology?

The domain of metrical phonology has several merits for an investigation about the feasibility of unambiguous data (identified with either cues or parsing). First, although the parameter set consists of several parameters that interact in a complex fashion (Dresher, 1999), the set is small enough to make a find-all-parses approach more feasible and also provides a natural restriction on the relevant parameter set for the parsing method. In addition, though the parameter set is not as large as some implementations of syntax (Sakas (2003) implemented a version containing 13 interactive parameters⁴²), it is still significantly more complex than the simplified case where the learner has only 1 or 2 interacting parameters to set.

Second, the cues method was originally proposed for this domain (Dresher, 1999), so there is some belief that it could be successful as an approach in general. It has also been used to study the acquisition of stress in English as a second language

⁴² Note that 13 parameters is likely still a very small subset of the actual available syntactic parameters.

(Archibald, 1992). Third, the English system is not a toy example, and in fact is extremely messy. There are significant quantities of unambiguous data for *both values* of any given parameter. This makes the system non-trivial to learn because of the conflicting unambiguous data; the learner is required to extract systematicity from a very noisy environment.⁴³ The noisiness of the English data forces the learner to adopt order constraints on parameter-setting so that the *correct* systematicity is posited for the system. Because the parameters interact, it is easy for the learner to converge on the incorrect parameter value for a given parameter if the learner does not use order constraints.

The difficulty of this task makes the ability of either the cues method or the parsing method to learn the English system a major accomplishment already. The success of each of these methods lends support to the feasibility of an unambiguous data filter on the learner's intake, however such unambiguous data may be discovered by the learner.

5.3.2 Metrical Phonology Parameters

Metrical phonology is the system that determines which syllables in a word are stressed and how much stress each syllable receives compared to all the other syllables in the word. Here, I will be concerned with only the parameters that determine which syllables get stressed, and not with those which determine how much stress.

5.3.2.1 Parameters vs. Probabilistic Association

Given that there are a finite number of stress contours for words of n syllables, one might reasonably wonder why a parametric system is required instead of having the learner simply associate entire stress contours probabilistically with particular words (e.g. see Skinner (1957) for an associationist view of language learning, among many others). We can also translate this question to the realm of syntax: given that there aren't infinitely many parses for a given sentence, why don't we simply probabilistically associate structures with sentences, rather than having a procedure to generate these structures? Much ink has been spilled on this subject, with primary arguments coming from the finite syntactic variations across languages and the finite range of syntactic mistakes children make during learning.

One of the main arguments for a system of metrical phonology comes from stress change over time. Suppose learners simply associated stress contours probabilistically to individual words. Then, we would expect that stress change over time would proceed in a piece-meal fashion, with individual words changing at different times. Instead, we find cases where some historical linguists posit a swift change to an underlying *system* for analyzing stress contours that are assigned to words in order to best characterize the observed language change. This is because a number of words change at the same time, which would be quite coincidental if they were not somehow related. A very direct way to relate them is to say a common

⁴³ In fact, the English data are so messy that many linguists didn't believe English had any systematicity until Chomsky & Halle (1968).

system is used to generate their stress contours, and the change occurs to this system. Dresher & Lahiri (2003), for instance, note a particular shift in stress contours in Middle English between 1400 and 1530 (a relatively short time from a language change perspective) and posit a change to one parameter in the Middle English system in order to explain it.

Second, if there were no underlying system for generating the observed stress contours, we might also expect that when change occurs, the start and end states should be close to each other from a stress contour perspective. For instance, we might expect main stress to move from the final to the penultimate syllable. However, again we find examples where the start and end states do *not* seem closely linked with respect to the observable stress contours; instead, they are only close together when viewed in terms of a parametric system for generating the observed contours (again see Dresher & Lahiri (2003) for Middle English).

Because change can be sudden on a large scale and more easily explicable when viewed through the lens of a systematic representation for stress, it is believed that speakers represent stress contours in a systematic way that is richer than probabilistic association for individual lexical forms. Specifically for this chapter, I will assume speakers use the parameter system I outline below.

5.3.2.2 Parameters for Stress

I present a sketch of the metrical phonology parameters that are described more fully in Dresher (1999).⁴⁴ The parameter space is schematized in figure 31. Some parameters have only one level (e.g. Feet Headedness), while other parameters contain sub-parameters that become available if one option is chosen at the first level (e.g. Quantity Sensitivity).

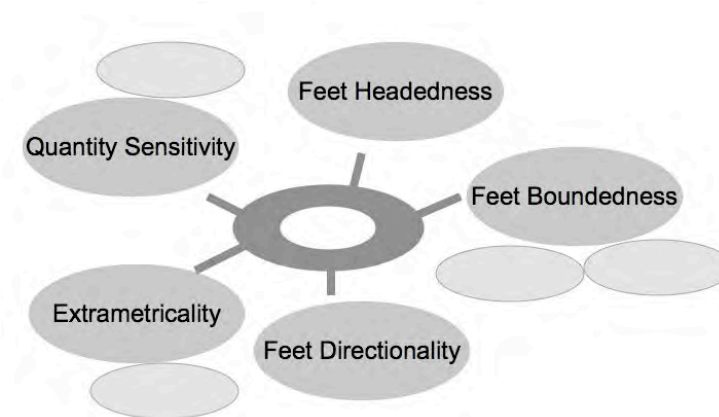


Figure 32. A schematic representation of the relevant parameters in metrical phonology, 5 main parameters and 4 subparameters for a total of 9 interacting parameters.

⁴⁴ Note that this parametric system differs from the instantiations in Halle & Idsardi (1995), Dresher (1994), and Idsardi (1992), though there are fairly straightforward mappings between the instantiation considered here and the ones considered in those studies.

A sample representation of metrical phonology structure is in figure 33.

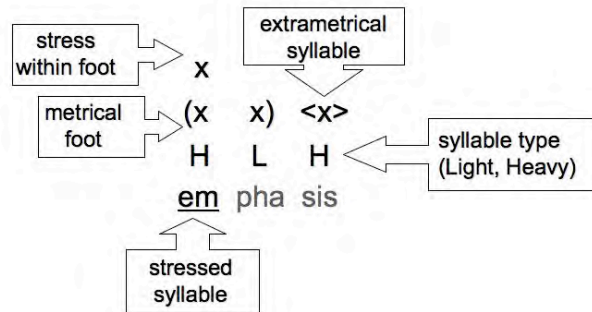


Figure 33. A sample representation of metrical phonology structure for ‘emphasis’, including terms to be described in more detail below: syllable type, metrical foot, extrametrical syllable, and stress within a metrical foot. In ‘emphasis’, the first and last syllables (‘em’, ‘sis’) are classified as Heavy, while the middle syllable (‘pha’) is classified as Light. The last syllable (‘sis’) is considered extrametrical, and not included in the metrical foot grouping. The first two syllables (‘em’, ‘pha’) are grouped into a single metrical foot, and the leftmost syllable in the foot (‘em’) is stressed.

5.3.2.2.1 Quantity Sensitivity

The first level of the quantity sensitivity parameter is whether the system is quantity-insensitive (QI) or quantity-sensitive (QS) (Halle & Idsardi, 1995; Hayes, 1980; among many others). An example language of this kind is Maranungku (Dresher, 1999). A quantity-insensitive system treats all syllables the same (represented as ‘S’ in (4)), whether they contain a long vowel as the nucleus (VV), a short vowel with a coda (VC), or a short vowel only (V). A long vowel syllable is “lu” in *ludicrous*, a short vowel with coda syllable is “crous” in *ludicrous* (the *s* is the consonant following the nucleus), and a short vowel only syllable is “di” in *ludicrous*. Note that the onset is irrelevant to syllable classification: VC, CVC, and CCVC are all classified as short vowels with codas and V, CV, and CCV are all classified as short vowels without codas. For syllables with a long vowel, the coda is also irrelevant – VV, VVC, and VVCC are all classified as long vowel syllables. In the examples below, all stressed syllables are underlined.

(4) ‘ludicrous’ analyzed in a QI system

syllable classification	S	S	S
nucleus & coda only	VV	V	VC
translation into V/C	CVV	CV	CCVC
syllables	<u>lu</u>	di	crous

A quantity-sensitive system divides syllables into (L)ight and (H)eavy⁴⁵. Examples of this kind of language include Koya, Selkup, and Khalka Mongolian (Dresher, 1999; Halle & Idsardi, 1995; Hayes, 1980; among others). Long vowels (VV) are always Heavy while short vowels (V) are always Light. A subparameter then becomes available for how to classify short vowel syllables with codas (consonants following the vocalic nucleus (VC)), since some languages classify these as Light (VC-Light), e.g. Selkup (Dresher, 1999), while others classify them as Heavy (VC-Heavy), e.g. Koya (Dresher, 1999).

(5) ‘ludicrous’ analyzed in a QS system

(a) VC-Light

syllable classification	H	L	L
nucleus & coda only	VV	V	VC
translation into V/C	CVV	CV	CCVC
syllables	<u>lu</u>	di	crous

(b) VC-Heavy

syllable classification	H	L	H
nucleus & coda only	VV	V	VC
translation into V/C	CVV	CV	CCVC
syllables	<u>lu</u>	di	crous

Note that a syllable classified as H should have stress unless some other parameter interferes, such as extrametricality.

5.3.2.2.2 Extrametricality

Syllables in a word are grouped into larger units called metrical feet. Only syllables that are included in a metrical foot can be stressed. A syllable classified as extrametrical cannot be included in a metrical foot and therefore cannot receive stress. Only the syllable at the left or right edge of a word may be extrametrical, and only one syllable in the word may be extrametrical (both edge syllables cannot be extrametrical).⁴⁶

A system can have no extrametricality (Em-None), so that all peripheral syllables are included in metrical feet. An example of this type of language is Maranungku (Dresher, 1999). Note that metrical feet are signified by parentheses (...) in the remaining examples.

⁴⁵ Though occasionally more complex weight systems have been proposed.

⁴⁶ There are additional proposed sub-classes of extrametricality that I will not consider here, such as (1) only Light edge syllables may be extrametrical (Hayes, 1980), (2) only the final syllable of nouns may be extrametrical (Hayes, 1980), (3) only the final consonant may be extrametrical (Archibald, 1998), and (4) only the final segment of the derivational stem (as indicated in the lexicon) can be extrametrical (Harris, 1983). Excluding these subparameters is an example of restricting the relevant parameter set for parsing.

(6) An Em-None analysis of ‘afternoon’, assuming QS-VC-Light; two metrical feet

syllable classification			
& metrical foot grouping	(L	L)	(H)
translation into V/C	VC	VC	VV
syllables	<u>af</u>	ter	<u>noon</u>

A system can also have extrametricality (Em-Some), e.g. English (Dresher, 1999), and then a subparameter becomes available to decide whether the leftmost syllable (Em-Left) or rightmost syllable (Em-Right) is the extrametrical one. Extrametrical syllables are signified by angle brackets <...> in the remaining examples.

(7) Em-Some analyses

(a) An Em-Left analysis of ‘agenda’, assuming QS-VC-Heavy; 1 metrical foot

syllable classification			
& metrical foot grouping	< L >	(H	L)
translation into V/C	V	VC	V
syllables	a	<u>gen</u>	da

(b) An Em-Right analysis of ‘ludicrous’, assuming QS-VC-Heavy; 1 metrical foot

syllable classification			
& metrical foot grouping	(H	L)	< H >
translation into V/C	VV	V	VC
syllables	<u>lu</u>	di	crous

As we can see in (7) above, the syllables that are classified as extrametrical do not receive stress. This is particularly striking in ‘ludicrous’, since the extrametrical syllable ‘crous’ is classified as Heavy under QS-VC-Heavy. Under normal circumstances, a Heavy syllable is usually stressed. Nonetheless, the extrametricality of the syllable interferes here, and allows the syllable to be without stress (and conform to the observed stress contour of ‘ludicrous’). This is one example of how different parameters can interact with each other.

5.3.2.2.3 Feet Directionality

Sequences of stressed syllables can be joined together as feet (Halle & Vergnaud, 1978; Hayes, 1995; Hayes, 1980; among many others). Metrical feet can be constructed beginning from the left side of the word (Ft Dir Left) or from the right side of the word (Ft Dir Right). An example of a language constructing feet from the left is Maranungku (Halle & Idsardi, 1995). Examples of languages constructing feet

from the right are Warao and Weri (Halle & Idsardi, 1995).⁴⁷

(8a) Metrical feet from the left (in a QS-VC-Light, Em-None system): L L H; 2 metrical feet

- (i) L L H
- (ii) (L L H)
- (iii) (L L) H
- (iv) (L L) (H)
- (v) (L L) (H)

Example stress contour: L L H
 Matching word: pe rox ide 'peroxide'

(8b) Metrical feet from the right (in a QS, Em-None system): L L H; 2 metrical feet

- (i) L L H
- (ii) L L H)
- (iii) L (L H)
- (iv) L) (L H)
- (v) (L) (L H)

Example stress contour: L L H
 Matching word: ho li day 'holiday'

As (8) shows, the syllables are divided differently into metrical feet, depending on the feet directionality. Since exactly one syllable in a metrical foot can receive stress, the differing metrical foot divisions can result in differing stress contours.

5.3.2.2.4 Boundedness

Boundedness refers to how large a metrical foot can be (Hayes, 1980; among many others). In an unbounded system (Unb), metrical feet can be arbitrarily large. The only reason a new metrical foot is started is if a Heavy syllable is encountered when grouping syllables into metrical feet. If, as in (9c) below, there are no Heavy syllables, then there will only be 1 metrical foot. Examples of this kind of language are Selkup and Koya (Dresher, 1999).

(9) Examples of unbounded analyses

(a) QS, Em-None, Ft Dir Left system: L L L H L; 2 metrical feet

- (i) L L L H L
- (ii) (L L L H L)
- (iii) (L L L) (H L)
- (iv) (L L L) (H L)

⁴⁷ Note that the examples below contain hypothetical analyses of the English words given as examples. In other words, those analyses are compatible with the stress contours observed.

(b) QS, Em-None, Ft Dir Right system: L L L H L; 2 metrical feet

- (i) L L L H L
- (ii) (L L L H L)
- (iii) (L L L H) (L)
- (iv) (L L L H) (L)

(c) QS, Em-None, Ft Dir Left system: L L L L L; 1 metrical foot

- (i) L L L L L
- (ii) (L L L L L)
- (iii) (L L L L L)

In contrast, a bounded system places a limit on the size of the metrical foot, such that only a certain number of units are included. After that limit is reached, a new metrical foot is started. Examples of these kind of languages include Cayuvava, Warao, Weri, and Maranungku (Halle & Idsardi, 1995). Once the learner determines that the system is bounded, two subparameters become available: the size limit - 2 or 3 units (B-2 or B-3) - and what the counting units are - syllables or moras (B-Syl or B-Mor). Moras are units of syllable weight used in some languages (such as Japanese). If moras are the counting units, a Heavy syllable counts as two moras while a Light syllable counts as only one. Analyses using the various bounded options are in (10) and (11).

(10) Examples of bounded analyses: B-2 vs. B-3

(a) B-2, Em-None, Ft Dir Left: x x x x; 2 metrical feet

- (i) x x x x
- (ii) (x x x x)
- (iii) (x x) x x
- (iv) (x x) (x x)
- (v) (x x) (x x)

(b) B-3, Em-None, Ft Dir Left: x x x x; 2 metrical feet

- (i) x x x x
- (ii) (x x x x)
- (iii) (x x x) x
- (iv) (x x x) (x)
- (v) (x x x) (x)

(11) Examples of bounded analyses: B-Syl vs. B-Mor

(a1) QI, Em-None, Ft Dir Left, B-2, B-Syl: S S S S; 2 metrical feet

- (i) S S S S
- (ii) (S S S S)
- (iii) (S S) S S
- (iv) (S S) (S S)
- (v) (S S) (S S)

(a2) QS, Em-None, Ft Dir Left, B-2, B-Syl: L H L L; 2 metrical feet

- (i) L H L L
(ii) (L H L L
(iii) (L H) L L
(iv) (L H) (L L
(v) (L H) (L L)
- (a3) QS, Em-None, Ft Dir Left, B-2, B-Syl: H H L L; 2 metrical feet
- (i) H H L L
(ii) (H H L L
(iii) (H H) L L
(iv) (H H) (L L
(v) (H H) (L L)
- (b) QS, Em-None, Ft Dir Left, B-2, B-Mor: H H L L; 3 metrical feet
- (i) H H L L
x x x x
(ii) H H L L
(x x) x x
(iii) H H L L
(x x) x x
(iv) H H L L
(x x) (x x) x x
(v) H H L L
(x x) (x x) x x
(vi) H H L L
(x x) (x x) (x x)
(vii) H H L L
(x x) (x x) (x x)
(viii) H H L L
(x x) (x x) (x x)
(ix) (H) (H) (L L)

As (11a3) and (11b) demonstrate, using syllables instead of moras as the counting units can create a markedly different metrical foot structure, which then affects the observed stress contour.

5.3.2.2.5 Feet Headedness

Feet headedness refers to which syllable in a metrical foot receives stress – the leftmost (Ft Hd Left) or the rightmost (Ft Hd Right) (Hayes, 1980; among many others).

(12) Examples of analyses with Ft Hd Left and Ft Hd Right – stressed syllables

underlined

- (a) QI, Em-None, Ft Dir Left, B-2, B-Syl, **Ft Hd Left**: S S S
(S S) (S) → S S S
- (b) QI, Em-None, Ft Dir Left, B-2, B-Syl, **Ft Hd Right**: S S S
(S S) (S) → S S S

5.3.2.3 Interacting Parameters

As we can see, all the metrical phonology parameters interact in their effect on the final stress contour assigned to a given word; a change to any one of them could change the stress contour in a non-trivial fashion. An example is illustrated in (13): the change of one parameter value (Em-None to Em-Left) causes the entire stress contour to become its inverse.

(13) A change to one parameter can drastically affect the stress contour assigned

- (a) QI, **Em-None**, Ft Dir Left, B-2, B-Syl, Ft Hd Left: S S S S S
3 metrical feet
(S S) (S S) (S) → S S S S S
Example: Maranungku ‘langkaratati’ → lang ka ra ta ti

- (b) QI, **Em-Left**, Ft Dir Left, B-2, B-Syl, Ft Hd Left: S S S S S
2 metrical feet
< S > (S S) (S S) → S S S S S
Example: Maranungku ‘langkaratati’ – incorrect stress pattern
→ lang ka ra ta ti
Example: English ‘communication’ – correct stress pattern
→ co mmu ni ca tion

Moreover, ambiguity can also easily arise – a single stress contour could be covered by multiple combinations of different parameter values, as shown in (14). Note that these combinations yield identical stress contours for ‘communication’, but these combinations may well yield differing stress contours for other words. Thus, the collection of combinations that produce the observable stress contour for any given word will vary from word to word.

(14) Multiple analyses of a single stress contour:

some analyses of ‘communication’ = co mmu ni ca tion

- (a) QI, Em-Left, Ft Dir Left, B-2, B-Syl, Ft Hd Left
2 metrical feet
< S > (S S) (S S) → S S S S S
- (b) QI, Em-Right, Ft Dir Left, B-2, B-Syl, Ft Hd Right
2 metrical feet
(S S) (S S) < S > → S S S S S

- (c) QS, QSVCH, Em-Right, Ft Dir Right, B-2, B-Syl, Ft Hd Right

2 metrical feet

(L H) (L H) < H > → L H L H L

(d) QS, QSVCH, Em-Right, Ft Dir Left, B-3, B-Mor, Ft Hd Right

2 metrical feet

(x x x) (x x x) x x
(L H) (L H) < H > → L H L H L

Converging on the correct values for the adult system with its interacting parameters is thus not a simple task. Because the parameters all combine to produce the observable stress contour, identifying unambiguous data for a *single* parameter value is not easy. Nonetheless, this is precisely what the cues and parsing methods are proposed to do. I will now describe how both methods would identify unambiguous data for each of the values of each of these parameters, thereby instantiating the unambiguous data filter on the learner's intake.

5.4 The Cues Method for Finding Unambiguous Data

The cues method makes identification of unambiguous data simple, provided the learner knows the relevant cues and can match them to the data encountered. Recall that the cues method was originally proposed by Dresher (1999) for the metrical phonology domain, and he described a set of potential cues for each of the parameters. One property of his cue set is that it assumes some parameters values are the default, and cues are only for the marked values. As I noted previously, this could be perceived as a pitfall since it requires the learner to have pre-specified domain-specific knowledge (perhaps as a non-uniform prior probability distribution biased towards the default value). Dresher (1999) suggests a way to derive this knowledge: learners begin with simple representations and must be driven to more complex representations (in the spirit of Chomsky & Halle (1968)). He proposes that his default values are simpler representations than their marked counterparts.

Nonetheless, since default values are an additional stipulation about the learner's knowledge, I provide an alternate set of cues that does not require defaults; each opposing parameter value has its own cue. I will compare the performance of these two cue implementations on the metrical phonology data.

5.4.1 Quantity Sensitivity

In the cue set proposed by Dresher (1999), the value where the syllables are undifferentiated (QI) is the default value. The cue for QS (where the syllables are classified as either Light and Heavy) is to compare words with the same number of syllables. If they have different stress contours, then the system is QS.

(15) Dresher cues for quantity sensitivity

(a) QI: default value (no cue required)

(b) QS: 2 words with n syllables that have different stress contours

Ex: $n = 2$, word 1: VV V, word 2: VV VV

An alternate cue set has cues for both QI and QS, as well as for the subparameters of QS (QS-VC-Light, where a VC syllable is treated as Light, and QS-VC-Heavy, where a VC syllable is treated as Heavy). The cue for QI is to find an unstressed internal VV syllable (which would be Heavy in a QS system, and therefore likely to be stressed) (16a). The cue for QS is to find a 2 syllable word with 2 stresses (or a 3 syllable word with 2 adjacent stresses if the system is known to be extrametrical already) (16b). Once the system is known to be QS, the cue for QS-VC-Light is an unstressed internal VC syllable (16c) while the cue for QS-VC-Heavy is a 2 syllable word with 2 stresses, where at least one syllable is VC (or the 3 syllable variant if extrametricality is known to apply) (16d).

(16) Alternate cues for quantity sensitivity

(a) QI: unstressed internal VV syllable

Ex: VV VV VV

(b) QS: 2 syllable word with 2 stresses, or 3 syllable word with 2 adjacent stresses if extrametricality is known

Ex: (1) VV VV (2) Em-Right: VV VV VV

(c) QS-VC-Light: unstressed internal VC syllable

Ex: VV VC VV

(d) QS-VC-Heavy: 2 syllable word with 2 stresses, with at least one syllable VC (or 3 syllable word with 2 adjacent stresses and at least one syllable VC if extrametricality is known)

Ex: (1) VV VC (2) Em-Right: VV VC VV

Note that if a default-marked system was preferred, the QI and QS-VC-Light values would function as the default values, with cues existing for QS and QS-VC-Heavy. I offer some speculation as to why the QI and QS-VC-Light values might be the default. One could argue that a QI system, because it treats all the syllables as the same, is a simpler method than dividing syllables into Light and Heavy. One could also argue that a QS-VC-Light system is simpler than a QS-VC-Heavy system. In particular, if a division between Light and Heavy syllables must be made, and Heavy syllables are marked in some way, having only VV syllables be Heavy is simpler than having other syllables such as VC also be Heavy.

5.4.2 Extrametricality

In the cue set proposed by Drescher, having an extrametrical syllable is the default state. This may be a difficult default to defend, however, since one might view extrametricality (i.e. ignoring certain edge syllables) as a marked feature of the metrical structure that the learner would need evidence for. Nonetheless, in the Drescher (1999) system, cues rule out extrametricality for each side (Em-Left and Em-Right). To rule out extrametricality for a given side, the edge syllable (leftmost for Em-Left and rightmost for Em-Right) must have stress.

(17) Drescher cues for extrametricality

- (a) Em-None: Both leftmost and rightmost syllables have stress
Ex: VV VC
- (b) Em-Some (Left or Right): default

An alternate cue set has cues for Em-Some (both Em-Left and Em-Right) as well as for Em-None. The cue for no extrametricality (Em-None) is similar to the Drescher-style cue: both edge syllables are stressed (18a). The cue for Em-Some is that a Heavy syllable at either edge of the word is unstressed (18b); the cue for Em-Left is that the leftmost syllable is Heavy and unstressed (18c) while the cue for Em-Right is the rightmost syllable is Heavy and unstressed (18d).

(18) Alternate cue set for extrametricality

- (a) Em-None: Both leftmost and rightmost syllables have stress
Ex: VV VC
- (b) Em-Some: Either edge syllable is Heavy and unstressed
Ex: (1) H L H (2) H L H
- (c) Em-Left: Leftmost syllable is Heavy and unstressed
Ex: H L H
- (d) Em-Right: Rightmost syllable is Heavy and unstressed
Ex: H L H

Note again that the alternate cue set could also be set up as a default-marked system. In the alternate cue set, having no extrametricality (Em-None) could be argued as the default under the assumption that all syllables should be included for metrical feet groupings until the learner is forced by evidence to do otherwise.

5.4.3 Feet Directionality

The cue set proposed by Drescher requires the feet directionality cues to be combined with the feet headedness cues, and so I will examine these cues together in section 5.4.5. An alternate cue set has cues for feet directionality separate from cues for feet headedness.

In the alternate set, the cue for Feet Directionality Left is dependent on the quantity sensitivity value. If the system is quantity insensitive (QI), the cue is 2 stressed adjacent syllables at the right edge of the word (19a1); if the system is quantity sensitive (QS), the cue is 2 stressed adjacent syllables with the first syllable Heavy and the second Light at the right edge of the word (19a2). In addition, if the system is known to have extrametricality on the rightmost syllable, then the cue is shifted to the previous two syllables. The cue for Feet Directionality Right is exactly the same, except that the 2 stressed adjacent syllables are at the left edge of the word (19b).

(19) Alternate cue set for feet directionality

(a) Feet Directionality Left

(1) If QI: 2 stressed adjacent syllables at the right edge of the word (if extrametricality exists for the rightmost syllable, the 2 stressed adjacent syllables are shifted over one position)

Ex: (1) S S S S (2) S S S S <S>

(2) If QS: 2 stressed adjacent syllables at the right edge of the word, with the first as H and the second as L (if extrametricality exists for the rightmost syllable, the 2 stressed adjacent syllables are shifted over one position)

Ex: (1) L L H L (2) L L H L <L>

(b) Feet Directionality Right

(1) If QI: 2 stressed adjacent syllables at the left edge of the word (if extrametricality exists for the leftmost syllable, the 2 stressed adjacent syllables are shifted over one position)

Ex: (1) S S S S (2) <S> S S S S

(2) If QS: 2 stressed adjacent syllables at the left edge of the word, with the first as L and the second as H (if extrametricality exists for the leftmost syllable, the 2 stressed adjacent syllables are shifted over one position)

Ex: (1) L H L L (2) <L> L H L L

5.4.4 Boundedness

In the cue set proposed by Drescher, the Unbounded value is the default and cues signal that the system is bounded. The cue for boundedness is the presence of an internal stressed Light syllable.

(20) Drescher cues for boundedness

(a) Unbounded: default

(b) Bounded: an internal stressed Light syllable

Ex: L L L L

An alternate cue set has cues for both Unbounded and Bounded, as well as for the subparameters of Bounded (B-2 vs. B-3, B-Syl vs. B-Mor). The cue for an unbounded system depends on the system's quantity sensitivity. If the system is QI, the cue is three or more unstressed syllables in a row (21a1); if the system is QS, the cue is three or more unstressed Light syllables in a row (21a2).⁴⁸

The cue for a bounded system is really the union of the cues for B-2 and B-3, which are again dependent on the quantity sensitivity of the system. If the system is QI, the B-2 cue is three or more syllables in a row with every other syllable stressed

⁴⁸ Note that this cue can interact with extrametricality. If the learner knows the system is extrametrical (either left or right), that syllable would be excluded from the three (or more) unstressed syllables necessary to be an Unbounded cue.

(21c1); if the system is QS, the cue is three or more Light syllables in a row with every other syllable stressed (21c2). The B-3 cue is nearly identical, except that there must be four or more (Light) syllables in a row with every third one stressed (21d).

This leaves the cues for a system that counts syllables (B-Syl) vs. a system that counts moras (B-Mor). The B-Syl cue also depends on the quantity sensitivity of the system. If the system is QI, then the cue is identical to the B-2 and B-3 cues (3+ syllables in a row with every other one stressed or 4+ syllables in a row with every third one stressed) (21e1). If the system is QS and B-2, the cue is 2 adjacent syllables with the pattern ‘H L’ or ‘L H’ (21e2); if the system is QS and Bounded-3, the cue is 3 adjacent syllables with the pattern ‘H L L’ or ‘L L H’ (21e3).

The B-Mor cue is far simpler: a 2 syllable word with both syllables stressed, and both syllables are Heavy (21f). If extrametricality is known to apply, then the cue is the same except that it applies to the 2 adjacent syllables that aren’t extrametrical.

(21) Alternate cue set for boundedness

(a) Unbounded: 3+ unstressed (Light) syllables in a row

Ex: (1) QI: S S S S (2) QS: L L L L

(b) Bounded: union of B-2 and B-3 cues

(c) B-2: 3+ (Light) syllables in a row, every other one stressed

Ex: (1) QI: S S S S (2) QS: L L L L

(d) B-3: 4+ (Light) syllables in a row, every third one stressed

Ex: (1) QI: S S S S (2) QS: L L L L

(e) B-Syl:

(1) QI: is union of B-2 and B-3 cues for QI

(2) QS, B-2: 2 adjacent syllables with pattern ‘H L’ or ‘L H’

Ex: (1) L L H L (2) L H L L

(3) QS, B-3: 3 adjacent syllables with pattern ‘H L L’ or ‘L L H’

Ex: (1) H L L L L (2) L L L L H

(f) B-Mor: 2 syllable word with both syllables stressed and Heavy

Ex: (1) H H (2) (Em-Left) < L > H H

The complexity of some of the cues for boundedness suggests that a default-marked system might be quite attractive here. In this case, I speculate that Unbounded would be the default, since it is an assumption that there is no arbitrary metrical foot size.⁴⁹ Also, counting by syllables (B-Syl) as opposed to moras (B-Mor) could be argued as the default, since words are already divided into syllables for many of the other parameters.

5.4.5 Feet Headedness

The set of cues proposed by Dresher has a single “cue” for feet directionality and feet headedness. In fact, this cue is really very much like a find-all-parses analysis using the restricted parameter set $F = \{\text{Feet Directionality, Feet}$

⁴⁹ Also, the Unbounded value as default falls out from the metrical phonology system implemented by Idsardi (1992).

Headedness}. The learner parses the known set of words with all combinations of feet headedness and feet directionality ((1) Ft Dir Left/Ft Hd Left, (2) Ft Dir Left/Ft Hd Right, (3) Ft Dir Right/Ft Hd Left, (4) Ft Dir Right/Ft Hd Right). For a given combination, if all the known words can be parsed such that all Light syllables that aren't the head of a metrical foot are unstressed, then this situation is the "cue" for this combination of feet directionality and feet headedness.

However, as I proposed an alternate set of cues for feet directionality by itself, I propose an alternate set for feet headedness by itself. The cue for Feet Headedness Left is that the leftmost syllable of the leftmost foot is stressed (22a); the cue for Feet Headedness Right is that the rightmost syllable of the rightmost foot is stressed (22b).

(22) Alternate cue set for feet headedness⁵⁰

(a) Feet Hd Left: the leftmost syllable of the leftmost foot is stressed

(1) VV VC V

(2) (Em-Left) < VV > VC V V

(b) Feet Hd Right: the rightmost syllable of the rightmost foot is stressed

(1) VV VC V

(2) (Em-Right) VC V VC < VV >

5.4.6 Summary: Cues

I have now stepped through cues for each of the relevant parameters in the metrical phonology domain. Recall that one of the strengths of cues is that the learner can easily identify unambiguous data, since it will match the cue the learner knows. However, the cues proposed here are heuristic in nature and may cause the learner to perceive false positives and false negatives, which could in turn lead the learner astray.

5.5 The Parsing Method for Finding Unambiguous Data

5.5.1 The find-all-parses analysis

The parsing method differs from the cues method in that it is more resource-intensive to identify unambiguous data, but far less likely to identify false positives and false negatives. A learner using the parsing method will parse the given data point with all *available* values of all the parameters in the relevant parameter set. Note that a parameter value ceases to be available when the learner decides the other value is correct for the language. For instance, if the learner has decided that the system is QS, no parses will be generated that use the value QI.

I have termed this procedure the find-all-parses analysis. While there are other implementations of the parsing method that are not as resource-taxing as the find-all-parses analysis, the find-all-parses analysis is the most inclusive version. I want to give the parsing method the best chance for successful identification of

⁵⁰ The feet headedness cues can actually apply to the same word if it has stress on both the initial and final syllable. However, the learner effectively learns nothing from such a word since neither parameter value gets the advantage over the other from this word. Alternatively, the learner might choose to explicitly ignore such a word as inconsistent, since it displays cues for mutually exclusive parameter values.

unambiguous data to see how it compares to the cues method. If this version of the parsing method is superior to the cues method, then we can see if weaker versions using less resources are also superior. First, however, I investigate whether the find-all-parses implementation can get the job done.

After the learner has conducted a find-all-parses analysis on the data point, the learner then sees if only one parameter value of a parameter leads to a successful parse of the data point. If so, the data point is considered unambiguous for that value. The results of an example find-all-parses analysis for a data point are shown in (23).

(23) The results of a find-all-parses analysis of ‘af ter noon’: sets of parameter values that yield a matching stress contour

- (a) (QI, **Em-None**, Ft Dir Left, B, B-2, B-Syl, Ft Hd Left)
- (b) (QI, **Em-None**, Ft Dir Rt, B, B-2, B-Syl, Ft Hd Rt)
- (c) (QS, QS-VCL, **Em-None**, Ft Dir Left, Unb, Ft Hd Left)
- (d) (QS, QS-VCL, **Em-None**, Ft Dir Left, B, B-2, B-Syl, Ft Hd Left)
- (e) (QS, QS-VCL, **Em-None**, Ft Dir Rt, B, B-2, B-Syl, Ft Hd Rt)

Since all successful parses share the parameter value Em-None, this data point would be considered unambiguous for Em-None.

Note that if the relevant parameter set is restricted, the find-all-parses analysis returns fewer parameter value sets and the same data point may then be considered unambiguous for other parameter values. Thus, a data point may be viewed as unambiguous for *different* parameter values at different points in time. This emphasizes how the definition of “unambiguous” is relative to the learner’s current knowledge state. This shift in perceived “unambiguity” is demonstrated in (24). In this example, suppose the learner has determined that the system is QI. The find-all-parses analysis will disregard any parses that include QS values.

(24) The results of a find-all-parses analysis of ‘af ter noon’, with the restriction that the system is QI: sets of parameter values that yield a matching stress contour

- (a) (QI, **Em-None**, Ft Dir Left, **B, B-2, B-Syl**, Ft Hd Left)
- (b) (QI, **Em-None**, Ft Dir Rt, **B, B-2, B-Syl**, Ft Hd Rt)

Given these results from the find-all-parses analysis, the learner using the parsing method would view this data point as unambiguous for Em-None, Bounded, Bounded-2, and Bounded-Syl.

5.5.2 Summary: Parsing

I have now described how the parsing method identifies unambiguous data in the input. Recall that one of the strengths of parsing is that it is not heuristic in nature and therefore will not perceive false positives or false negatives. The simplicity of the method is also appealing, since the learner needs only to use a procedure already available for language comprehension.

However, the find-all-parses method proposed here has its own pitfalls. I reiterate that it may be quite resource-intensive for a learner to implement. Moreover,

a fundamental problem with the parsing method is that it cannot use data it cannot parse. In a language learning situation in which there aren't many exceptions to the adult parameter values, this is not too damaging. But the English data set, as I have mentioned before, is fraught with such exceptions. Once the learner has set some of the parameter values correctly, it will be unable to parse the non-trivial portion of the data that are exceptions. In this sense, the parsing method may not be "flexible" enough to cope with noisy data. We will see how this inflexibility impacts the learning path the learner must take to converge on the correct set of parameter values for English. First, however, we will verify the performance of these two methods on an easier language learning case where the available data set is not exception-filled.

5.6 Cues and Parsing in a Clean Language Environment: An Easy Case

A "clean" language environment is one in which there are *no* conflicting unambiguous data in the learner's input. A clean language environment makes convergence on the adult parameter values very straightforward. Given sufficient time, the learner will be exposed to enough unambiguous data points to converge on the adult parameter values. There are no "garden paths" provided by unambiguous data for the incorrect parameter values (in contrast to noisy data sets such as English). As long as these methods allow the learner to perceive *some* data as unambiguous, the learner will eventually converge on the correct parameter values. I sketch how these methods would work for a language like Maranungku, which has stress on every odd syllable counting from the left (examples in (25) from Drescher (1999) and Kager (1995)).

(25) Maranungku Stress Contour Examples (stressed syllables underlined)

- (a) lang ka ra ta ti - 'prawn'
- (b) we le pe le man ta - 'kind of duck'
- (c) ya ngar ma ta - 'the Pleiades'
- (d) me re pet - 'beard'
- (e) ti ralk - 'saliva'

The parameter values for Maranungku are in (26).

(26) Maranungku metrical phonology parameter values (QI, Em-None, Ft Dir Left, B, B-2, B-Syl, Feet Hd Left)

Each of the words in (25) can be analyzed with either the cues or parsing method to identify if any are unambiguous for any of the parameter values (see table 5.2). As we can see, the two methods do not always agree on how many values a given data point is unambiguous for. Nonetheless, all the data are always unambiguous for the correct adult system parameter values, due to the clean language environment.

	Unambiguous for Cues	Unambiguous for Parsing
--	----------------------	-------------------------

<u>l</u> ang ka <u>r</u> a te <u>t</u> i	Bounded-2, Em-None	Bounded-2, Em-None
<u>w</u> e le <u>p</u> e le <u>m</u> an ta	Bounded-2, Ft Hd Left	Bounded-2
<u>y</u> a ngar <u>m</u> a ta	Ft Hd Left	Bounded-2
<u>m</u> e re <u>p</u> et	Em-None	Em-None
<u>t</u> i ralk	Ft Hd Left	<i>Nothing</i>

Table 5.2. The results of using the cues and parsing methods to classify five Maranungku words as unambiguous for available parameter values.

Because there are no conflicting unambiguous data, neither the cues nor parsing method would classify the data as unambiguous for the incorrect parameter value. The learner should thus converge on the adult Maranungku system no matter which method is used, given exposure to sufficient data. Moreover, there are no order constraints on the learning path: the learner should converge on the correct adult values, no matter what order the learner sets the parameters in.

5.7 Learning English: A Harder Case

English, however, poses a more difficult challenge since it *does* have conflicting unambiguous data points, as perceived by the learner. I turn now to how I tested each of these methods for learning the English metrical phonology system.

5.7.1 Estimating the Composition of the Input to the Learner

I compiled caretaker speech to children between the ages of 6 months and 2 years from the CHILDES database (MacWhinney, 2000), for a total of 540505 words. Each of these words were then divided into syllables and marked with stress, using the CALLHOME database of telephone conversation (Canavan et al., 1997) and the MRC psycholinguistics database (Wilson, 1988) as references for likely syllabic divisions and stress contours. I assumed that this was a reasonable estimation of the composition of the data English learners would be exposed to.

5.7.2 The Logical Problem of Learning English Metrical Phonology

The correct parameter values for English are listed in (27).

- (27) English metrical phonology parameter values
(QS, QS-VC-Heavy, Em-Right, Ft Dir Right, Ft Hd Left, B, B-2, B-Syl)

Converging on the correct parameter values for English adults is non-trivial, given realistic distributions of input to English children. We must ask what parameter-setting orders (if any) will lead the learner to converge on the adult parameter values. Importantly, every time learners set one parameter, they may then view all subsequent data differently. So, the setting of one parameter in one way could bias the learner to set another parameter in another way later on. Thus, the order of parameter-setting can have a significant effect on the final set of parameter values the learner converges on. A viable parameter-setting order will lead the

learner to converge on the correct set of parameter values for the language.

The viable orders are derived via an exhaustive walk through all possible parameter-setting orders; hence, this is exploring the logical problem of learning, in that we are interested in whether the target state is achievable at all using these learning methods. To conduct the exhaustive walk for a given learning method (cues or parsing), we must try out every single parameter-setting order with the input.

In the worst case, no order will suffice – the target set of parameter values is unreachable, given the input and this learning method. Learning with an unambiguous data filter produces insufficient behavior.

A better scenario is that learning with an unambiguous data filter *does* produce sufficient behavior. A slightly better case is that there *is* a set of orders that will allow the learner to reach the target set, but these orders are completely unrelated to each other. There is no way to make the knowledge necessary for acquisition success concise; the learner must somehow be aware of the viable orders explicitly. In an even better case, there is a set of orders that will work, and they can be captured by a small number of *order constraints*, though these order constraints may need to be stipulated. A still better case is that a set of viable orders exists that can be captured by *principled* order constraints that are independently derivable. In the best case, all parameter-setting orders will be viable so there is no need to worry about the order of parameter-setting at all.⁵¹ In this last case, since there are no constraints on the order of parameter-setting, there is no need to explain how the learner knows them or why the learner follows them.

5.3 Conducting an Exhaustive Walk Through All Possible Orders

5.3.1 The Algorithm for Identifying All Viable Parameter-Setting Orders

Here, I describe the method for conducting an exhaustive walk through all possible parameter-setting orders to determine which, if any, will lead the learner to converge on the adult set of parameter values.

(28) Algorithm for identifying all viable orders of parameter-setting for a given learning method

- (a) For all currently unset parameters, determine the unambiguous data distribution in the corpus (i.e. how much unambiguous data there is for each value of each unset parameter).
- (b) Choose a currently unset parameter to set. The value chosen for this parameter is the value that has a higher probability in the data the learner perceives as unambiguous. This logic behind this is that, given enough data points (i.e. a sufficiently long learning period), this parameter value will eventually accrue enough probability to become the winning parameter value.
- (c) Repeat steps (a-b) until all parameters are set.
- (d) Compare final set of values to target set of values. If they match, this is a viable parameter-setting order.

⁵¹ This is the case of the clean language environment described in the previous section.

(e) Repeat (a-d) for all parameter-setting orders.

The process of determining the distribution of unambiguous data in this corpus after each parameter is set (28a) is meant to reflect how the learner perceives the incoming data at different points in the parameter-setting process. I want to use all the data available in the sample corpus to estimate what the input distributions are for the learner at any given point in time. Thus, after each parameter is set, this algorithm gauges how the learner would then view the available input in the linguistic environment by recalculating the unambiguous data distributions in the corpus.

In (28b), the learner chooses the parameter value with a higher probability in the unambiguous data. There are two ways to measure unambiguous data probability, depending on what the learner is relativizing the probability against. One way, which I will refer to as the *relativize-against-all* approach, relativizes the unambiguous data for that parameter value against the entire input set. The second way, which I will refer to as the *relativize-against-potential* approach, is for the learner to relativize the unambiguous data for that parameter value against the set of *potential* unambiguous data points. The set of potential unambiguous data points is smaller than the entire input set and may vary across parameters, since not every data point satisfies the preconditions necessary to be an unambiguous data point. Moreover, the preconditions will vary depending on whether the learner uses cues or parsing to identify unambiguous data. I will describe in detail below why this occurs. Meanwhile, it is unclear a priori which relativization approach should be preferred by the learner, so I will examine the effects of both separately.

5.7.3.2 Relativization of Unambiguous Data Probability

The relativize-against-all approach can intuitively be characterized by the question, “How likely is it that a random data point chosen from the entire input set will be an unambiguous data point for the parameter value of interest?” It does not matter for this approach whether unambiguous data are identified via cues or via parsing because the relativizing set (the input set size) is constant across both cues and parsing.

As a concrete example, suppose the data set provides 11213 data points perceived by the learner as unambiguous for Quantity Sensitive (QS) and 2140 perceived as unambiguous for Quantity Insensitive (QI). The total data set size is 540505 words, so the relativized probability for an unambiguous QS data point is $11213/540505 = 0.0207$ and the relativized probability for an unambiguous QI data point is $2140/540505 = .00396$. The learner will choose QS (.0207) over QI (.00396).

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	540505	540505
Relativized Probability	0.00396	0.0207

Table 5.3. Relativize-against-all approach, for both the cues and parsing method. The learner will choose QS.

The relativize-against-potential approach can intuitively be characterized by the question, “How likely is it that a random data point chosen from the set of data points satisfying the preconditions to be unambiguous will actually be an unambiguous data point for the parameter value of interest?” The relativizing set (the set of potential unambiguous data points) will vary in size, depending on whether cues or parsing is used to identify unambiguous data points. Specifically, if the learner uses cues, the relativizing set will vary *across parameter values*. In contrast, if the learner uses parsing, the relativizing set will remain constant across parameter values.

If the learner uses cues to identify unambiguous data, the learner is looking for a combination of structure and stress within a word (e.g. words of 2 syllables that are both stressed for QS). Words that do not match the structural requirement of the cue (e.g. word of 4 syllables for the QS cue) cannot possibly have the correct structure and stress combination to be a cue, since they already lack the correct structure. Thus, these data points are excluded from the set of potential unambiguous data points since they do not obey the necessary structural preconditions. Because of the different structural requirements of the cues for different parameter values, the relativizing set size will vary from cue to cue.

As a concrete example, suppose the data set provides 11213 data points perceived by the learner as unambiguous for QS and 2140 data points perceived by the learner as unambiguous for QI. Suppose also that the potential set of QS cues (words having 2 syllables, etc.) is 85268 while the set of potential QI cues (words of at least 3 syllables, etc.) is 2755. The relativized probability for an unambiguous QS data point is $11213/85268 = 0.132$ and the relativized probability for an unambiguous QI data point is $2140/2755 = .777$. The learner using the cues method will choose QI (.777) over QS (.132).

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	2755	85268
Relativized Probability	0.777	0.132

Table 5.4. Relativize-against-potential approach, for the cues method. The learner will choose QI.

If the learner uses parsing to identify unambiguous data, the set of potential cues consists of all parseable words. The number of parseable words will depend on the currently set parameter values, since some words may not be able to be parsed once certain parameter values are set (e.g. words with syllable-final stress will not be parseable if Em-Right (extrametricality on the rightmost syllable) is set). However, in contrast with the cues method, the size of the relativizing set (the parseable words) will *not* vary from parameter value to parameter value. Thus, all unambiguous data point counts are normalized against the same value, just as in the relativize-against-all approach (though the actual value is less than the entire input set).

As a concrete example, suppose the data set provides 11213 data points perceived by the learner as unambiguous for QS and 2140 data points perceived by the learner as unambiguous for QI. Suppose also that there are p parseable words,

given the current parameter settings. The relativized probability for an unambiguous QS data point is $11213/p$, which will be larger than the relativized probability for an unambiguous QI data point, $2140/p$. The learner using the parsing method will choose QS ($11213/p$) over QI ($2140/p$).

	QI	QS
Unambiguous Data Points	2140	11213
Relativizing Set	p	p
Relativized Probability	$2140/p$	$11213/p$

Table 5.5. Relativize-against-potential approach, for the parsing method. The learner will choose QS.

5.7.3.3 An Example of Testing a Parameter-Setting Order

In (29), I demonstrate steps (28a-b) for a learner using the parsing method and the relativize-against-all approach, testing a parameter-setting order that begins by setting the quantity sensitivity parameter to QS. Note how the distribution of unambiguous data for a given parameter (such as Extrametricality below) can shift drastically, depending on what parameters are currently set.

(29) Testing an order for the parsing method with the relativize-against-all approach that begins by setting quantity sensitivity

(a) Currently unset parameters: Quantity Sensitivity, Extrametricality, Feet Directionality, Boundedness, Feet Headedness

Quantity Sensitivity		Extrametricality	
QI: 0.00398	QS: 0.0205	Em-None: 0.0284	Em-Some: 0.0000259
Feet Directionality		Boundedness	
Ft Dir Left: 0.000	Ft Dir Rt: 0.00000925	Unb: 0.00000370	Bounded: 0.00435
Feet Headedness			
Ft Hd Left: 0.00148	Ft Hd Rt: 0.000		

Table 5.6. Unambiguous data distribution from corpus: probability of finding unambiguous data point in input data set, using parsing method and relativize-against-all (probabilities calculated out of 540505 words)

(b) Choose quantity sensitivity to set. QS has a higher probability of finding an unambiguous data point (QS probability is 0.0205, which is greater than QI's probability of 0.00398). Set Quantity Sensitivity to QS.

(c) Currently unset parameters: QS-VC-Light/QS-VC-Heavy, Extrametricality, Feet Directionality, Boundedness, Feet Headedness

QS-VC-Light/QS-VC-Heavy	Extrametricality
--------------------------------	-------------------------

VC-Light: 0.00265	VC-Heavy: 0.00309	Em-None: 0.0240	Em-Some: 0.0485
Feet Directionality		Boundedness	
Ft Dir Left: 0.000	Ft Dir Rt: 0.00000555	Unb: 0.00000370	Bounded: 0.00125
Feet Headedness			
Ft Hd Left: 0.000588	Ft Hd Rt: 0.0000204		

Table 5.7. Unambiguous data distribution from corpus: probability of finding unambiguous data point in input data set, using parsing method and relativize-against-all (probabilities calculated out of 540505 words)

This process then continues for the remaining unset parameters in the system until all parameters are set.

5.8 English Learning Results

5.8.1 Order Constraints as a Metric

If both learning methods yield a set of parameter-setting orders that lead to the correct target values for English, then both solve the logical problem of language learning for the English metrical phonology system. That is, both have at least one parameter-setting order that leads the learner to the target state. If there is more than one viable order, we can then compare the two methods by how well-formed the sets of viable parameter-setting orders are.

First, we can determine if the set for each method can be captured by order constraints at all, whether stipulated or principled. If so, then the set is at least well-formed enough to be described in a more compact representation than explicitly listing all the viable orders in the set. After that, we can then consider the nature of the order constraints that capture each set. A method with a set that can be described by principled constraints will be considered superior to a method with a set that can only be described by constraints that must be explicitly stipulated.

5.8.2 Parameter-setting Orders that Lead to English Target Values

As it turns out, both methods yield a set of parameter-setting orders that will cause a learner to converge on the English values when the relativize-against-all approach is used to calculate the relativized probability of unambiguous data. Both methods thus pass the first hurdle of solving the logical problem of language learning for the English metrical phonology system. Again, this is no mean feat given the interactive nature of the 9 parameters that produce stress contours and the noisiness of the data to the learner.

However, only the parsing method succeeds when the relativize-against-potential approach is used. Because the relativizing set for parsing is constant across parameter values for both relativization approaches, the set of viable orders for parsing is the same for each approach. In contrast, the relativizing set for cues varies

across parameter values in the relativize-against-potential approach, and in fact leads to *no* orders being viable to reach the target state for English. Table 5.8 summarizes the behavior of the cues and parsing methods when combined with different approaches to relativizing the probability of the unambiguous data.

	Cues	Parsing
relativize-against-all	Successful	Successful
relativize-against-potential	<i>Unsuccessful</i>	Successful

Table 5.8. Comparison of success of different methods of identifying unambiguous data with different approaches to relativizing probability of unambiguous data.

Given that the parsing method always has a viable set of orders, one might believe that parsing is therefore the superior method for identifying unambiguous data. It succeeds no matter what the probability is relativized against because the relativizing set is constant across all parameter values. However, recall that the characterization of the viable set of orders is also important. A set characterized by constraints that are principled is more desirable than a set characterized by constraints that must be stipulated. I shall therefore examine the viable set of orders for both methods and see how they compare.

In (30a) below, I list a sample of the parameter-setting orders for the cues method that allowed the learner to converge on the English values. In (30b), I list a sample of the parameter-setting orders that failed to work. For a complete listing of the orders that were successful, see the Appendix.

(30a) Sample of Cues Method Parameter-Setting Orders that Succeeded

- (a) QS, QS-VC-Heavy, B, B-2, Feet Hd Left, Feet Dir Right, Em-Right, B-Syl
- (b) QS, B, B-2, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, Em-Right, B-Syl
- (c) B, B-2, Feet Dir Right, QS, Feet Hd Left, QS-VC-Heavy, Em-Some, Em-Right, B-Syl
- (d) Feet Hd Left, Feet Dir Right, B, B-2, QS, QS-VC-Heavy, Em-Some, Em-Right, B-Syl
- (e) Feet Dir Right, QS, Feet Hd Left, B, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl

(30b) Sample of Cues Method Parameter-Setting Orders that Failed

- (a) Em-Some, Em-Right, Feet Dir Right, QS, Feet Hd Left, B, QS-VC-Heavy, B-2, B-Syl
- (b) QS, B, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, Em-Some, Em-Right, B-Syl, Bounded-2
- (c) Feet Hd Left, Feet Dir Right, B, B-Syl, B-2, QS, QS-VC-Heavy, Em-Some, Em-Right

In addition to a viable set of parameter-setting orders existing for the cues method, this viable set can also be described more succinctly by the order constraints in (31).

(31) Cues Method: Order Constraints

- (a) QS-VC-Heavy set before Em-Right
- (b) Em-Right set before B-Syl
- (c) B-2 set before B-Syl

The rest of the parameters are freely ordered with respect to each other.

In (32a) below, I list a sample of the parameter-setting orders for the parsing method that allowed the learner to converge on the English values. In (32b), I list a sample of the parameter-setting orders that failed to work. For a complete listing of the orders that were successful, see the Appendix.

(32a) Sample of Parsing Method Parameter-Setting Orders that Succeeded

- (a) QS, B, Feet Hd Left, QS-VC-Heavy, Feet Dir Right, B-Syl, B-2, Em-Some, Em-Right
- (b) B, QS, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, B-Syl, Em-Some, Em-Right, B-2
- (c) Feet Hd Left, QS, QS-VC-Heavy, B, Feet Dir Right, En-Some, Em-Right, B-Syl, B-2

(32b) Sample of Parsing Method Parameter-Setting Orders that Failed

- (a) Feet Dir Right, QS, Feet Hd Left, B, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl
- (b) Em-Some, Em-Right, QS, B, Feet Hd Left, Feet Dir Right, QS-VC-Heavy, B-Syl, B-2
- (c) QS, QS-VC-Heavy, Feet Hd Left, Feet Dir Right, B, B-Syl, B-2, Em-Some, Em-Right
- (d) QS, Feet Dir Right, QS-VC-Heavy, Feet Hd Left, B, B-Syl, B-2, Em-Some, Em-Right
- (e) B, Feet Dir Right, QS, QS-VC-Heavy, Feet Hd Left, B-Syl, B-2, Em-Some, Em-Right

In addition to a viable set of parameter-setting orders existing for the parsing method, the viable set can also be described more succinctly by dividing the parameters into three groups. The parameters within each group are freely ordered with respect to each other (33).

(33) Parsing Method: Order Constraints as Freely-Ordered Groups

- (a) Group 1: QS, Ft Hd Left, B
- (b) Group 2: Ft Dir Right, QS-VC-Heavy
- (c) Group 3: Em-Some, Em-Right, B-2, B-Syl

At first glance, the order set for the cues method appears to be less constrained than the order set for the parsing method. However, the true criterion of merit is to compare how easily each of the constraints can be derived from other properties of the learning system.

5.8.3 Deriving Constraints

There are several ways I could think of to derive constraints from properties of the learning system: data saliency, data quantity, and default values.⁵² I describe each of these in turn.

I begin with the saliency of the data. Data that are better signals might be noticed and used more easily by the learner than data that aren't. This is true no matter what the domain. In the domain of metrical phonology, it has been suggested that the unexpected presence of stress is more informative than the unexpected absence of stress (Bill Idsardi, *personal communication*). The presence of stress is a stronger logical signal since there are many factors that could cause the absence of stress if the stress system is unknown, e.g. stress deletion under clash, conflation of secondary stresses, and segmental rules such as vowel devoicing (Halle & Idsardi, 1995; among others). The presence of stress, however can pinpoint a parametric cause (or a lexically pre-existing stress that has to be stored explicitly in the system anyway (Halle & Idsardi, 1995)). Moreover, the presence of stress may be a stronger acoustic signal, since a stressed syllable is more prominent than an unstressed syllable. Stressed syllables might therefore be more readily attended to by the learner.

There is also morphological evidence that the presence of stress is psychologically more salient. Morphological rules exist that restrict affix attachment to words with stress on the edge syllable (-al for final stress words: remove + al = removal), but I am currently unaware of any morphological rules that exist that restrict affix attachment to words *without* stress on the appropriate syllable. This suggests some psychological priority for paying attention to stressed syllables over their unstressed counterparts. Given the informational asymmetry between the presence and absence of stress, we might expect parameters that rely on the learner noticing the absence of stress to be deprioritized. Extrametricality (Em-Some, Em-Right, Em-Left) is just such a parameter; thus, we might expect it to be set later than other parameters.

Secondly, the quantity of data available to the learner could also affect parameter-setting order. Again, this will be true irrespective of the domain. Parameters with more unambiguous data available are likely to be set before parameters with less, simply because there is more data for the learner to use for updating.

Thirdly, if the learner is using a default value, we can dispense with constraints for that value if it is the correct adult value since it is already set by default. Again, this will be true irrespective of the domain. The logic behind this is that a constraint of the form "Parameter value A1 must be set before parameter value B1" results from either (a) A1 not being able to be set correctly if B1 is set first (i.e. the unambiguous data distribution favors A2 after B1 is set) or (b) B1 not being able to be set correctly until A1 is set (i.e. the unambiguous data distribution favors B2 until A1 is set). Depending on which it is, this problem can be side-stepped if either

⁵² There may in fact be more as well. These three come to mind as being fairly general properties of the learning system.

(a) A1 is the default for A or (b) B1 is the default for B, respectively, since the correct parameter value is already set. The constraint “Parameter value A1 must be set before parameter value B1” is then unnecessary.

I note that using defaults only applies to the cues method since the instantiation of the parsing method used here must use all available values to conduct a find-all-parses analysis. One might argue that the parsing method could in fact be instantiated with a default system under a different implementation. However, this has an inherent problem. Specifically, the only values available to the parser would be the default values. Thus, only parses using the default values would be considered by the learner initially. This is fine if the adult values for the system are the default values. But, suppose they are not. How will the learner recognize unambiguous data for the non-default values, a problem noted by Valian (1990)? The parser, by definition, can only use data it can parse. The non-default values are not in its set of available values, and so it will not be able to parse data that can only be parsed with those values. In short, the parsing method cannot comprehend data that are unambiguous for the non-default values since it cannot parse such data with the default values. This is in sharp contrast to the cues method, which can still recognize unambiguous data for the marked values even while the default values are in place.

I will now examine which constraints for the cues and parsing methods can be accounted for using these three explanations: data saliency, data quantity, and default values.

5.8.3.1 Cues Method with Relative-Against-All: Accounting for Constraints

The first constraint (31a) was that QS-VC-Heavy must be set before Em-Right. We can derive this via data saliency, and argue that noticing the absence of stress for extrametricality is more difficult than noticing the presence of stress in the pattern for QS-VC-Heavy.

The second constraint (31b) was that Em-Right must be set before Bounded-Syl. (This is due to Bounded-Mor being favored until Em-Right is set.) When we examine that unambiguous data distribution, it turns out that Em-Right has at least 20 times as much data as Bounded-Syl (and so, the learner is 20 times more likely to find an Em-Right cue) at any given point in time. Thus, this constraint could be derived from data quantity. Also, I noted in section 5.4.4 that the cues learner could use Bounded-Syl as a default value once the more general Bounded value is set. If this is the case, then Bounded-Syl will already be set and this constraint disappears from the use of default values.

The third constraint (31c) was that Bounded-2 must be set before Bounded-Syl. (Bounded-Mor is favored until Bounded-2 is set.) Unfortunately, the unambiguous data distribution favors Bounded-Syl over Bounded-2 initially so we cannot directly derive this constraint from data quantity. However, there is a partial ordering with Em-Right which can be useful. Specifically, once Em-Right is set, a Bounded-2 cue is at least 4 times as likely to be found as a Bounded-Syl cue at any given point and would then be set first. So, once Em-Right is set, this constraint can be derived from data quantity. However, this requires Em-Right to be set before Bounded-2.

Fortunately, an Em-Right cue is about 270 times more probable than a Bounded-2 cue, so Em-Right could easily be set first. Thus, this constraint could be derived from data quantity: set Em-Right, then Bounded-2, and then Bounded-Syl. Also, we could rely on default values again to cause this constraint to disappear: Bounded-Syl is the default value once the more general Bounded value is set.

What is striking here is that *all* of the cues method order constraints are derivable from other properties of the learning system (either the learner's learning preferences or the available data). They do not need to be explicitly stated or available to the learner as pre-specified knowledge. This makes these constraints highly attractive.

5.8.3.2 Parsing Method with Relative-Against-All/Potential: Accounting for Constraints

The parsing method's constraints, however, are not so easily derived. Recall that the parsing method learner must constrain parameter-setting to three parameter groups that are ordered with respect to each other (33) – all the ones in the first group (QS, Feet Hd Left, Bounded) must be set before all the ones in the second group (Ft Dir Right, QS-VC-Heavy), and all the ones in the second group must be set before all the ones in the third group (Em-Some, Em-Right, Bounded-2, Bounded-Syl). Since the parsing method learner in this model cannot use default values, the constraints can be derived only from the properties of data saliency and data quantity.

I note that even supposing the parsing method *could* somehow use default values, these constraints still cannot all be derived. The only constraint that default values could account for is Bounded-Syl in the third grouping: Bounded-Syl is the default value, and so would already be set. There is no need for it to be set after other parameters. No other constraints could be accounted for by default values since the adult values are the non-default values (QS, QS-VC-Heavy, Bounded, Em-Some).

Still we can ask how much can be accounted for by the remaining two properties. Data saliency will explain why Em-Some and Em-Right are in the last group: noticing the absence of stress puts these parameters later in the learning path (group 3). This leaves data quantity to account for all the rest. Unfortunately, data quantity will not separate the remaining parameters into the three necessary groups. A parsing method learner would need to have these groups explicitly built in as prior knowledge, which makes these constraints less attractive than their cues method counterparts. The ability to derive all of the relevant order constraints thus seems to favor the cues method, when used with the relativize-against-all approach.

The success of the cues and parsing methods are compared below in Table 5.9. As we saw previously, the parsing method seems more flexible because it succeeds no matter what relativization approach is used. The cues method, however, has a set of order constraints that can be derived from properties of the learning system when this method does actually succeed.

	Reaches Target State	All Order Constraints Derivable
Cues + Relative-Against-All	Yes	Yes
Parsing + Relativize-Against-All	Yes	No
Parsing + Relativize-Against-Potential	Yes	No
Cues + Relativize-Against-Potential	No	<i>N/A</i>

Table 5.9. Comparing the performance of the cues and parsing methods, when used with different relativization approaches. The optimal combination for this case seems to be the cues method with the relative-against-all approach, since it both reaches the target state and has derivable order constraints.

5.9 Discussion

5.9.1 Cues: Why Better Constraints on Parameter-Setting Order?

As we saw in the previous section, the cues method results in a set of parameter-setting orders that can be captured by constraints that are independently derivable and few in number. This is not true for the order constraints that capture the parsing method's set: that set is far more restricted, and requires a larger number of constraints, most of which must be stipulated. I speculate that this has to do with the nature of the data that a cues method learner uses.

Cues themselves are small pieces of highly informative surface structure, such as 2 syllable words with 2 stresses (QS, QS-VC-Heavy, Bounded-Mor) or the leftmost syllables in a word with stress in a certain pattern (Ft Dir Rt, Ft Hd Left). Crucially, the learner doesn't have to understand the entire data point to identify a cue in the data point. In fact, the data point can be in conflict with values that are already set but *still* contain cues for currently unset values.

For example, a 2 syllable word with 2 stresses is in conflict with Em-Right since it has stress on the rightmost syllable, but is still useful as a cue for QS. This gives a cues method learner more flexibility than a parsing method learner has, since the cues learner can make use of the non-problematic portions of data points instead of having to disregard these portions along with the entire data point.

For the parsing method learner, if a data point can't be parsed (because the learner doesn't understand the entire data point or the data point is in conflict with currently set values), the data point can't be used at all. Note that this problem persists even when using other less resource-intensive parsing strategies (Sakas & Fodor, 2001) since those strategies consider cases where multiple parses can describe the data point, but not cases where *no* parses describe the complete data point. Unless the parsing method can retrieve information from only a subpart of the data point, the problem that plagues the parsing method here will persist. The noisiness of the English metrical phonology data set greatly penalizes the parsing method learner, which is reflected in the greater quantity of order constraints required to capture the more restricted set of viable parameter-setting orders.

However, the flexibility of cues is not without its drawbacks – a cues-learner can be led irrecoverably astray in some cases as we saw previously. When the cues method is combined with the relativize-against-potential approach, certain values that

are not in the English target state persist no matter what other values are set. For instance, because the relativizing set of QI unambiguous data is significantly smaller than that of the QS unambiguous data (QI: 2755, QS: 85268), a cues learner using the relativize-against-potential approach consistently awards a higher probability to the QI unambiguous data. No other parameter settings will influence the potential QI set because the QI cue does not interact with any other parameter value (e.g. the way QS cues do with Extrametricality), and so it will *always* be significantly smaller than the potential QS set. Because no other parameter settings affect the cue for the QI value, the relativizing QI set can never be altered. Unfortunately for a learner of English, having such a small relativizing QI set will cause the learner to favor the QI unambiguous data over the QS unambiguous data. Since QS is the correct value for English, no viable parameter-setting orders exist for cues when using the relativize-against-potential approach. Thus, we see that the flexibility the cues method has can be both a strength and a weakness, depending on what other learning strategies the learner adopts. Nonetheless, it is this flexibility which yields a more concise representation of knowledge necessary for acquisition success (the order constraints) when the method does, in fact, succeed.

5.9.2 Relativization

I examined two different approaches a learner might adopt for relativizing the probability of an unambiguous data point for a given parameter value: relativize-against-all and relativize-against-potential. While we had no a priori reason for assuming one approach was superior to the other, we may wish to use the results obtained here to support the relativize-against-all approach. Specifically, in order to reach the target state and have a set of viable orders that can be described by a small set of principled order constraints, a learner must use the cues method coupled with the relative-against-all approach. Thus, the learning procedure relativizes the probability of an unambiguous data point against the entire set of input seen so far. This is in contrast to a learning procedure that keeps track of the quantity of potential unambiguous data points, and relativizes the probability of an unambiguous data point against that set (which will vary across parameter values for cues). Because the learner does not need to keep track of the set of potential unambiguous data points for each parameter value, the relativize-against-all approach is likely less resource-intensive to implement as well. This is a desirable quality for a psychologically plausible learning strategy.

5.9.3 Cues and Parsing: A Viable Combination?

Cues and parsing have a complementary array of strengths and weaknesses as methods for identifying unambiguous data. From the case study examined here, we have seen an additional strength and weakness for both cues and parsing. Cues give us a principled set of order constraints, but aren't robust across different strategies of relativization. The opposite is true for parsing: we find robustness across different relativization approaches, but a set of order constraints that must be mostly stipulated. We also examined additional weaknesses in section 5.2 for both methods. Cues are

knowledge the learner must have already available; parsing can only use the entire data point, rather than just a subpart.

A very interesting question is if there is a way to combine these two methods to capitalize on their complementary strengths and mitigate their complementary weaknesses. I speculate now on how this might be accomplished. Cues themselves are small pieces of highly informative surface structure that are usually smaller than the entire data point. Given this, perhaps a learner might derive cues from a limited kind of parsing (perhaps limited by time and mental resources available). Such a limited parsing method could be biased to use subparts of a data point rather than trying to assign full parses to the entire data point.

For example, suppose a learner with no values set hears a sequence of syllables in the speech stream and realizes that two consecutive syllables are the beginning of a new word.⁵³ The learner then tries to analyze these two syllables alone. Suppose the first of these two syllables is stressed and contains a long vowel (VV) while the second is unstressed and contains a short vowel with a coda (VC).

(34) speech stream, with two syllables of new word (signaled by #): ...# VV VC...

The learner then tries to parse these two syllables with any parameter values that can be applied, given that only the beginning of the word is known. (It is possible that these two syllable comprise the entire word, but the learner is unaware of this.) The learner then tries to parse this sequence of syllables with all *applicable* parameter values – i.e., values that can apply to the front subpart of a word alone. The set of applicable values would be Quantity-Insensitive, Quantity-Sensitive, Extrametricality-Some [Left], Unbounded, Bounded, Bounded-Syllabic, Bounded-Moraic, Feet Headed Left, Feet Headed Right, and Feet Directionality Left. Em-None and Em-Right are not applicable since the right edge of the word is unknown, so nothing can be observed about the final syllable. Feet Dir Right is also not applicable for similar reasons: the learner cannot construct feet starting from the right edge since the right edge is unknown.

⁵³ Note that there may be some interleaving of learning the metrical phonology system and learning to segment words successfully. Learners early on have a sense of the basic rhythmic properties of their language (Mehler et al. 1988, Nazzi et al., 2000) – for instance, trochaic (first syllable stressed) or iambic (second syllable stressed) as stereotypical (Jusczyk et al. 1993). They may then use this highly probable rhythmic pattern to segment syllables in the speech stream into words (Jusczyk et al, 1999; Houston et al., 2000; Houston et al., 2004). Sometimes, this will result in mis-segmentation: *ba na na* becomes segmented as simply *na na* in English. This could then lead to misanalysis in the more complex metrical phonology domain, since the “word” being analyzed isn’t actually the word in the target language (analysis of “nana” instead of “bana”). If the more elaborate metrical phonology system examined here is learned early enough that correct word segmentation isn’t regularly successful, this could be another factor that determines learners’ success. In effect, they are applying an additional filter to the available input and only perceiving words that match the basic rhythmic bias they have acquired already. Thanks to the CUNY Supper Club for very useful discussion of this point.

(35) Viable Partial Parses for the syllable sequence #VV VC...

- (a) (QI, **Ft Dir Left**, Unb, **Ft Hd Left**)
- (b) (QI, **Ft Dir Left**, B, B-Syl, B-2, **Ft Hd Left**)
- (c) (QI, **Ft Dir Left**, B, B-Syl, B-3, **Ft Hd Left**)
- (d) (QS, QS-VC-Light, **Ft Dir Left**, Unb, **Ft Hd Left**)
- (e) (QS, QS-VC-Light, **Ft Dir Left**, B, B-Syl, B-2, **Ft Hd Left**)
- (f) (QS, QS-VC-Light, **Ft Dir Left**, B, B-Syl, B-3, **Ft Hd Left**)
- (g) (QS, QS-VC-Light, **Ft Dir Left**, B, B-Mor, B-3, **Ft Hd Left**)

Of all the available values, only Feet Headed Left and Feet Directionality Left are used by all parses of this two syllable sequence. Because Feet Directionality Right was not applicable, the learner will not conclude anything about Feet Directionality. Similar reasons preclude the learner from using this data point to signal Extrametricality – the full range of values for that parameter was not applicable: even though the learner would perhaps be able to rule *out* Extrametricality-Left, there is no definitive distinction between Em-None, Em-Some, or Em-Right. However, all the values for Feet Headedness *were* applicable: both Feet Headed Left and Feet Headed Right. Since Feet Headed Left was required for all parses, the learner would perceive this two syllable sequence as unambiguous for Feet Headed Left.

In this way, the learner would be deriving cues from a limited form of parsing that operates over subparts of data points. Note that if the learner derives cues from the implementation of parsing used here, the learner loses the ability to use default values since default values are not compatible with this instantiation of parsing. A learner cannot unlearn default values if the learner only ever uses default values to parse data; data indicating the non-default values are unparseable and therefore cannot be learned from (Valian, 1990). However, it may be possible to sidestep this problem with a probabilistic parser that favors default values and probabilistically uses them for parsing. Then, the learner would still be able to parse (a portion of) the unambiguous data encountered for the non-default value, if the adult system used the non-default value.

Still, we also lose the benefit from parsing that allows probability relativization to be constant across parameter values. The relativize-against-potential approach would have a relativizing set consisting only of the data which that value could possibly have parsed. The example we described above would be included in the relativizing set for Feet Headed Right (since Feet Headed Right was applicable), but not in the relativizing set for Feet Directionality Right (since Feet Directionality Right wasn't applicable).

However, it is possible that using limited parsing to derive cues gains some of benefits associated with using cues in the first place. In particular, operating over subparts of a data point is what I believed allowed the cues method to have fewer order constraints. It's possible that using cues derived from limited parsing would also produce a set of viable orders that can be characterized by fewer constraints. So, I posit that a learner using cues derived from limited parsing would potentially have the desired behavior combining the strengths of parsing and cues: less necessarily innate knowledge and fewer order constraints. This prediction, of course, remains to be explored.

Also, a limited parsing method would likely result in partial analyses that are more heuristic than exact, possibly at the expense of more false positives and false negatives. Though this may seem to be undesirable, such behavior may be good from the perspective of language change since certain language changes require *imperfect* learning, as we saw in the previous chapter. If data the learner considers unambiguous are keyed more to the surface form and are less well-connected to the abstract grammatical parameters, then it is easier for slippage to occur over time.

As a specific example, recall from the previous chapter that the change in Old English from Object-Verb order to Verb-Object order has been argued to be the result of imperfect learning in just this way (Lightfoot, 1991). Learners use cues (or parsing over a limited set of parameters) to find data they perceive as unambiguous, though this data may actually be ambiguous if parsed more fully. This allows the learners to converge on a slightly different probability distribution than the adults of the population have. Specifically for Old English, the system is a probability distribution between Object-Verb and Verb-Object order. The learners end up with a final probability that is marginally different from the probability of the rest of the population. Over time, these small “slips” lead to language change in the population. Importantly for communication purposes, the slips are, as mentioned, *small*. Cues derived from limited parsing would potentially allow learning to be successful enough to achieve the desired target state in most cases, but not so successful that small changes are impossible.

5.9.4 Future Directions

There are several immediate ways to build upon the findings concerning (a) other instantiations of the unambiguous data filter, (b) the sufficiency of the unambiguous data filter for other languages, (c) the necessity of the unambiguous data filter for learning metrical phonology, (d) experimentally testable predictions made by the unambiguous data filter, and (e) distinguishing systematic exceptions from noise in order to form irregular sub-systems given the available data.

The previous section described a potential combination of the methods for identifying unambiguous data that would retain the strengths of both the methods examined, cues and parsing. This combination strategy’s ability to actually converge on the English system should be examined, as well as any constraints required for its success. When I examined the cues and parsing methods separately, each required different constraints for acquisition success on the English dataset: cues required a particular assumption about how the learner relativizes the probability of unambiguous data, while parsing required order constraints that would need to already be available to the learner. The combination strategy might require constraints of both kinds (probability relativization and prior knowledge of parameter-setting order), one kind, or neither kind.

From the perspective of the logical problem of language learning, future work could also test the cues, parsing, and limited parsing methods on other languages for which we have sufficient corpora of child-directed speech. These methods can also be investigated in other domains besides metrical phonology.

The necessity of the unambiguous data filter also can be examined for this case study. As in the previous chapter's future directions, there are various ways to relax the unambiguous data filter and have the learner use ambiguous data. For instance, the learner could weight ambiguous data points such that they're not as influential as unambiguous data (again, as done in chapter 3 for learning anaphoric *one*). For the parsing method, the learner might adopt a probabilistic weighting of ambiguous data based on the percentage of successful parses that share a certain value. As an example, suppose 4 of 5 successful parses for a data point require Extrametricality-None, while 1 requires Extrametricality-Some. The learner might then give 80% credit to Extrametricality-None and 20% credit to Extrametricality-Some.

The learner might also adopt an ambiguous data strategy that probabilistically chooses one parameter value for each parameter to parse the data point. Successful parses reward all the parameter values used while unsuccessful parses punish all the parameter values used, as instantiated in the Naïve Parameter Learning model of Yang (2002). As an example, suppose the learner encounters an ambiguous data point and only has Bounded-2 vs. Bounded-3 and Extrametricality-Right vs. Extrametricality-Left remaining to be set. Suppose also that Bounded-2 is favored over Bounded-3, with associated probabilities of .8 (B-2) and .2 (B-3), and Extrametricality-Right is similarly favored over Extrametricality-Left, .8 (Em-Right) to .2 (Em-Left). Then, the learner chooses one of the four combinations of parameter values to parse the data point with, based on their associated combined probability: (a) B-2, Em-Right ($.8 * .8 = .64$), (b) B-3, Em-Right ($.2 * .8 = .16$), (c) B-2, Em-Left ($.8 * .2 = .16$), (d) B-3, Em-Left ($.2 * .2 = .04$). If the combination of values yields a successful parse of the data point, all values are rewarded; if the parse fails, all values are punished. This learning strategy is implicitly driven by the unambiguous data in the input since unambiguous data for one parameter value (e.g. Em-Right) will be unparseable by the opposing value (e.g. Em-Left), and so punish the opposing value (e.g. Em-Left). However, this strategy does not explicitly seek unambiguous data nor does it restrict the learner to use only unambiguous data, allowing it to avoid the sparse data problem that could potentially plague an unambiguous data learner.

In addition, the unambiguous data filter explored here makes testable predictions about which parameters should be set first in a given language, based on the order constraints required for acquisition success for either method of identifying unambiguous data. For instance, both cues and parsing would predict that a learner should set Quantity Sensitivity before Extrametricality. These predictions can be tested with both modeling (by using realistic estimates of the quantity of data children are exposed to) and standard experimental techniques for infants such as head-turn preference (Jusczyk & Aslin (1995)).

For the modeling extension, we can also investigate whether the necessary order constraints leading to the correct English target state (e.g. a cues learner setting Extrametricality before Bounded-Syllable) can emerge with a high probability simply from the distributions of data available to children or if instead data saliency explanations and/or default values are required. If default values are required, this suggests a prior probability distribution that strongly favors the default value. Moreover, the situation where the learner has a strong bias for one value over another

may be analogized to second-language learning: the adult has a very strong initial bias for the native language values. Exploring the behaviors produced from strong initial biases as well as ways to recover from these strong initial biases can have implications for second language learning.

Finally, the current learning model can be extended to search for systematic exceptions in the data in order to form irregular sub-classes. Exceptions (and errors) would be recognized once the learner has some of the system known. For instance, if the learner has determined the English system is Quantity Sensitive, an exceptional data point would be unambiguous for Quantity Insensitive. So, the learner can start recognizing exceptions even before the entire regular metrical system is learned. The learner might then be able to invoke a rule competition model, similar to Yang (2002)'s implementation for forming irregular past tense classes, in order to group irregular metrical phonology data points together into sub-classes. Systematic exceptions would be recognized as distinct from noise (or singular exceptions that should be memorized) based on the frequency of the words – and importantly, the different words – that are exceptional in that way. This again draws from Yang's (2002) implementation of forming irregular classes for the English past tense.

As an example, suppose the learner has decided the main system is Quantity Sensitive. However, the learner then keeps encountering data points that are incompatible with that parameter value: *ponytail*, *ladybug*, *jellybean*, etc. If these examples are frequent enough, the learner might hypothesize that there is an irregular class of words where the second syllable with the long vowel /i/ ('ny', /ni/; 'dy', /di/; 'ly', /li/) is destressed (even though /i/ is a long vowel and should receive stress given the regular system). If the learner is at a stage in learning where meaning is associated with words, then the irregular class might (additionally) be defined over something like compound words.

Importantly, the learner would need to recognize the exceptional data points as distinct from the main system being learned, but regular enough to warrant positing systematicity for them. To recognize the exceptional data points, the learner must already have some of the parameters for the main system set. The learner would thus benefit from the "preset" parameters of the main system in order to recognize and extract systematicity in any irregular sub-systems that might exist.

5.10 Summary

In this chapter, I have investigated the feasibility of using an unambiguous data filter on the learner's intake for metrical phonology, a complex system with multiple interacting parameters. I have shown that an unambiguous data filter can indeed allow a learner to converge on the correct set of adult values for English metrical phonology, which is a noisy system containing unambiguous data for the incorrect values as well as for the correct values. The learner is successful whether the unambiguous data filter is implemented by using the domain-specific representation of cues (Dresher, 1999; Lightfoot, 1999) or the domain-specific learning procedure of parsing (Fodor, 1998b, 1998c; Sakas & Fodor, 2001).

Nonetheless, there are differences between the two methods in terms of what must be explicitly stipulated and what can be derived from the learning system. In

addition, the two methods differ on their flexibility across different approaches of relativizing probability.

The parsing method does not need to stipulate additional information to identify unambiguous data, since the domain-specific procedure of assigning structure to a data point is already employed for language comprehension. Moreover, the parsing method succeeds no matter which probability relativization approach is used to analyze the data. Yet, the inability of the parsing method to use default values and make use of subparts of a data point force it to have a more restricted set of viable parameter-setting orders. This in turn leads to order constraints that must be stipulated in the case examined here.

The cues method, on the other hand, *can* use default values and glean information from data point subparts, which allows the set of viable parameter-setting orders to be far less restricted in the case examined here. However, a cues learner can only succeed when the unambiguous data are relativized against the entire input, rendering this method less flexible than the parsing method. Moreover, the original formulation of cues requires us to stipulate the domain-specific knowledge of cues in order to identify unambiguous data.

I have speculated a way of combining both methods: deriving cues from a limited form of parsing that allows parsing over subparts of a data point. The limited parsing method would thus possess two advantageous properties: (1) minimal knowledge is stipulated to identify unambiguous data and (2) more heuristic identification of unambiguous data that could lead to fewer order constraints. It is uncertain, however, if the limited parsing would succeed across different probability relativizations, since the set of potential unambiguous data would vary across parameter values, as it does for the cues implementation examined here. This remains to be explored.

The results obtained here suggest that an unambiguous data filter can lead to the correct learning results in complex domains. The crucial aspect of such a filter is that data are unambiguous *relative* to the learner's perspective, and the learner has incomplete knowledge of the full adult grammar during the learning process. Thus, data that appear unambiguous at an earlier time point may be viewed as ambiguous later when more information has been obtained, and vice versa. Contrary to severely handicapping the learner, such heuristic, inexact definitions of unambiguous data seem to allow the learner the flexibility to triumph in a noisy system. Given that the linguistic environment is often quite noisy, learners may benefit from treating data that conform to their semi-informed definition of unambiguous data as though they were truly unambiguous data – and therefore, fully informative for learning. In this way, a learner can feasibly implement an unambiguous data filter while avoiding the sparse data problem in realistic language learning cases.

Chapter 6: Learning By Filtering

In the case studies presented in this dissertation, I have explicitly investigated one component of the learning theory mechanism: the definition of the data intake. In each case, filtering the data intake has had enormous effects on the output of learning, separating learning failure from learning success. These case studies suggest that, perhaps contrary to intuition, using all the available data for learning isn't what real human learners do. Instead, young children can succeed by using a select subset of data from which they view as more informative and from which it is in some sense easier for them to extract the correct linguistic systematicity. For anaphoric *one*, learners succeed by heeding only the data that is informative about which N' to choose when there is more than one N' antecedent available. For word order properties such as Object-Verb or Verb-Object order, learners succeed by using degree-0 data that they perceive as unambiguous. For the English metrical phonology system, learners again succeed by using data perceived as unambiguous. Data intake restriction is key: using fewer data points that are cleaner is superior to using many data points that are noisy representations of the underlying linguistic system.

The division of the learning theory into distinct components allows us to combine components of different types together: domain-specific and domain-general, discrete and probabilistic. Moreover, this framework is a tool that can be applied to many learning problems with different hypothesis spaces that combine information across domains. In this dissertation, I have applied it to subset-superset hypotheses in the syntax-semantics interface, probabilistic distributions between hypotheses in syntax, and multiple interacting hypotheses in metrical phonology. In addition, the distinct components of the framework can be investigated separately, as I do here for data intake filtering. For this investigation, computational modeling has been a very valuable tool, since it allows precise control over the learner's data intake in a way that is difficult to achieve with traditional experimental techniques.

In sum, this dissertation represents the first steps towards a theory of the mechanism of language learning. I have answered the more specific questions set out for each case study. Yet, this has opened the way for still more questions. Future work, especially computational modeling work, will hopefully continue to draw on both theoretical and experimental linguistic data to explore how language learning can succeed in the noisy environment that surrounds young learners.

Appendix

This is the list the complete set of parameter-setting orders for each method and relativization approach that allowed the learner to converge on the English metrical phonology parameter values. From these sets, the order constraints described in section 5.7 were derived.

(A1) Viable Parameter-Setting Orders for the Cues Method, Relativize-Against-All

- (QS, QS-VC-Heavy, B, B-2, Ft Hd Left, Ft Dir Rt, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, B-2, Ft Hd Left, Em-Some, Em-Right, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, B-2, Ft Hd Left, Em-Some, Em-Right, B-Syl, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, B-2, Ft Dir Rt, Ft Hd Left, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, B-2, Ft Dir Rt, Em-Some, Em-Right, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, B-2, Ft Dir Rt, Em-Some, Em-Right, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, B-2, Em-Some, Em-Right, B-Syl, Ft Hd Left, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, B-2, Em-Some, Em-Right, B-Syl, Ft Dir Rt, Ft Hd Left)
- (QS, QS-VC-Heavy, B, B-2, Em-Some, Em-Right, Ft Hd Left, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, B-2, Em-Some, Em-Right, Ft Dir Rt, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, B-2, Em-Some, Em-Right, Ft Dir Rt, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Hd Left, B-2, Ft Dir Rt, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Hd Left, B-2, Em-Some, Em-Right, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Hd Left, B-2, Em-Some, Em-Right, B-Syl, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, Ft Hd Left, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Hd Left, Em-Some, Em-Right, B-2, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Hd Left, Em-Some, Em-Right, B-2, B-Syl, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, Ft Hd Left, Em-Some, Em-Right, Ft Dir Rt, B-2, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, B-2, Ft Hd Left, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, B-2, Em-Some, Em-Right, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, Ft Hd Left, B-2, Em-Some, Em-Right, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, Ft Hd Left, Em-Some, Em-Right, B-2, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-2, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, B-Syl, Ft Hd Left, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, B-Syl, Ft Dir Rt, Ft Hd Left)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, Ft Hd Left, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, Ft Hd Left, B-Syl, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, Ft Dir Rt, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, B-2, Ft Dir Rt, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, Ft Hd Left, B-2, Ft Dir Rt, B-Syl)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, Ft Hd Left, B-2, B-Syl, Ft Dir Rt)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, Ft Dir Rt, B-2, Ft Hd Left, B-Syl)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, Ft Dir Rt, B-2, B-Syl, Ft Hd Left)
- (QS, QS-VC-Heavy, B, Em-Some, Em-Right, Ft Dir Rt, Ft Hd Left, B-2, B-Syl)

(Ft Dir Rt, Ft Hd Left, B, B-2, QS, QS-VC-Heavy, Em-Some, Em-Right, B-Syl)

(A2) Viable Parameter-Setting Orders for the Parsing Method, Relativize-Against-All

(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
(QS, B, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-Syl, B-2, Em-Some, Em-Right)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-2, B-Syl, Em-Some, Em-Right)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-Syl, Em-Some, Em-Right, B-2)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-Syl, B-2)
(QS, B, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-2, B-Syl)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, B-Syl, B-2, Em-Some, Em-Right)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, B-2, B-Syl, Em-Some, Em-Right)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, B-Syl, Em-Some, Em-Right, B-2)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-Syl, B-2)
(QS, Ft Hd Left, B, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-2, B-Syl)
(QS, Ft Hd Left, QS-VC-Heavy, B, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
(QS, Ft Hd Left, QS-VC-Heavy, B, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
(QS, Ft Hd Left, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
(QS, Ft Hd Left, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
(QS, Ft Hd Left, B, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
(B, QS, Ft Hd Left, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-Syl, B-2, Em-Some, Em-Right)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-2, B-Syl, Em-Some, Em-Right)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, B-Syl, Em-Some, Em-Right, B-2)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-Syl, B-2)
(B, QS, Ft Hd Left, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-2, B-Syl)
(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)

(B, Ft Hd Left, QS, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, B-Syl, B-2, Em-Some, Em-Right)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, B-2, B-Syl, Em-Some, Em-Right)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, B-2, Em-Some, Em-Right, B-Syl)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, B-Syl, Em-Some, Em-Right, B-2)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-Syl, B-2)
 (B, Ft Hd Left, QS, Ft Dir Rt, QS-VC-Heavy, Em-Some, Em-Right, B-2, B-Syl)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
 (Ft Hd Left, QS, QS-VC-Heavy, B, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, B-Syl, B-2, Em-Some, Em-Right)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, B-2, B-Syl, Em-Some, Em-Right)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, B-2, Em-Some, Em-Right, B-Syl)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, B-Syl, Em-Some, Em-Right, B-2)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
 (Ft Hd Left, QS, B, QS-VC-Heavy, Ft Dir Rt, Em-Some, Em-Right, B-2, B-Syl)

(A3) Viable Parameter-Setting Orders for the Cues Method, Relativize-Against-Potential

No viable orders.

(A4) Viable Parameter-Setting Orders for the Parsing Method, Relativize-Against-Potential

Same set as (A2): parsing method and relativize-against-all approach.