Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology

Lisa Pearl
University of California, Irvine

Language learning is a tricky business. Knowledge of language consists of multiple complex systems, and there is often a non-transparent relationship between the observable data children have access to and the underlying systems of knowledge speakers use to generate that data. For instance, one system is metrical phonology, which determines which syllables are stressed and which are unstressed. The observable data is the output of the system, a stress contour such as [***stressed*** unstressed ***stressed***] in a word like ***af***er***noon***[1]. Though the output is simple, many pieces of structural knowledge combine to generate it, such as whether syllables are differentiated by weight, which syllables are included in metrical feet, and the size of metrical feet. Because of the interactive nature of the system, it can be difficult to uncover the individual system components by looking only at the output. Yet this is what children do.

The potential range of systems children could choose from is theoretically infinite. A helpful bias children could have is constraints on what systems they consider. These constraints are sometimes theorized as children knowing the parameters of cross-linguistic variation available (metrical phonology: Halle & Vergnaud, 1987; syntax: Chomsky, 1981). Children then set these parameters from exposure to the data in their native language environment.

Yet these constraints do not solve the problem of language learning. Though the range of possible systems is finite rather than infinite, there can still be a large number to choose from. Suppose children are aware of n binary parameters; there are 2^n possible systems to choose from. Even if n is small (say 20), this can lead to a very large number of potential systems ($2^{20} = 1,048,576$) (Clark, 1994, among others). Moreover, data are often ambiguous, and there may not be much informative data for any given parameter value of a system since each output is generated from a combination of interacting parameters (Clark, 1994). For example, figure 1 shows just some of the metrical phonology analyses that can yield the stress contour for ***af***ter***noon***. It is not so simple to determine which components are active from the stress contour alone.

Another way children might be constrained is the data they learn from.

1. Stressed syllables will be in ***bold italics*** in the remainder of the paper.
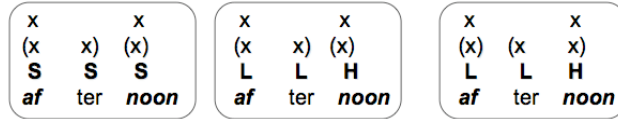
Figure 1. The ambiguity of the data: multiple analyses for *af*ter*noon*.

Specifically, children could learn only from a small subset of the data perceived as maximally informative: *unambiguous* data (Pearl & Weinberg, 2007; Fodor, 1998). In effect, children would implement a filter on the data intake that would cause them to ignore information in ambiguous data. The unambiguous status of a data point depends on the child's current knowledge state; data that are unambiguous early on in the learning process may not be informative later, and vice versa. Thus, the data perceived as unambiguous will be gauged subjectively by the child, and so will change throughout the learning process.

A reasonable concern is the viability of the parametric system and data intake filtering given a realistic data set. A parametric system may be useful for describing cross-linguistic variation, but how learnable is a realistic complex parametric system? If a system is not learnable, then it is not a very good model of the knowledge children use to constrain their hypotheses. Data intake filtering may provide highly informative data, but can unambiguous data be found for a complex system? If the data are too sparse to learn from, then this is not a good bias for children to have.

Here, we will investigate the feasibility and sufficiency of learning a realistic complex parametric system from a realistic data set by using data intake filtering. The child must learn an instantiation of metrical phonology with 9 interacting parameters (adapted from Dresher (1999)), and converge on the adult English system. Given highly ambiguous and exception-filled English child-directed speech (CHILDES, MacWhinney (2000)), a simulated child using an unambiguous data filter can nonetheless converge on the correct parameter values for English. This supports the viability of both the parametric system as a knowledge implementation and the unambiguous data filter as a learning strategy. Moreover, this study highlights the utility of empirically-grounded modeling as a tool for investigating the language learning mechanism.

1. The Parametric System of Metrical Phonology

The complex parametric system considered is an instantiation of a metrical phonology system with 9 interactive parameters (5 main and 4 sub-parameters), adapted from Dresher (1999). A sample structural analysis for 'emphasis' is shown in figure 2. The word is divided into syllables ('em', 'pha', 'sis'), which are classified as either Light or Heavy. The rightmost syllable ('sis') is

extrametrical, and so not included in a metrical foot. [2] The metrical foot spans two syllables ('em', 'pha'), and the leftmost syllable within the foot ('em') is stressed. This leads to the observable stress contour for 'emphasis': *em*phasis.
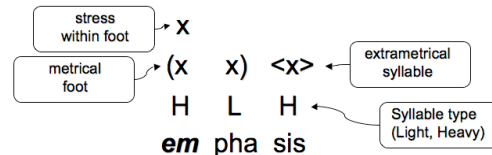

Figure 2. A sample metrical phonology structural analysis for "emphasis".

Many structural components combine to produce the seemingly simple observable stress contour. We will now briefly step through the various parameters involved to give a more detailed sense of how they interact.

1.1 Quantity Sensitivity

Quantity sensitivity refers to whether syllables are differentiated by rime weight[3]. If a system is quantity insensitive (QI), syllables are not differentiated. Long vowel syllables (VV), short vowel syllables without codas (V), and short vowel syllables with codas (VC) are all treated the same, as in (1).

(1)   syllable class   S   S   S
         VV/V/VC   VV   V   VC
         syllables   *lu*   di   crous

If a system is quantity sensitive (QS), syllables are differentiated into (H)eavy and (L)ight syllables. Long vowel syllables (VV) are H, short vowel syllables without codas (V) are L, and short vowel syllables with codas (VC) are either L (QS-VC-L) or H (QS-VC-H).

(2)   syllable class   H   L   H/L
         VV/V/VC   VV   V   VC
         syllables   *lu*   di   crous

1.2 Extrametricality

Stress assignment relies on both syllable weight and metrical foot formation. For example, if a syllable is H, it is stressed. However, syllables not in metrical feet (*extrametrical* syllables) cannot be stressed; so, even if an extrametrical syllable is H, it cannot receive stress.

In systems without extrametricality (Em-None), all syllables are included in metrical feet (3). In extrametrical systems (Em-Some), either the leftmost (4,

---

2. Metrical feet are indicated by parentheses ( ). Extrametrical syllables are indicated by angle brackets < >.

3. Syllable onset does not matter in this system: 'en' is like 'ten', 'sten', and 'stren'.

Em-Left) or rightmost syllable (5, Em-Right) is not included in a metrical foot. Note in (5) that the rightmost syllable, while H, is not stressed because of the parametric interaction with extrametricality.

(3)   syllable class   (L   L)   (H)
         VV/V/VC   VC VC   VV
         syllables   **af**   ter   ***noon***

(4)   syllable class   <L> (H   L)
         VV/V/VC   V   VC   V
         syllables   a   ***gen***   da

(5)   syllable class   (H   L) <H>
         VV/V/VC   VV   V   VC
         syllables   ***lu***   di   crous

1.3 Feet Directionality & Metrical Feet Bounds

Once the syllables to be included in metrical feet are known, metrical feet can be constructed. However, there is variation on which edge of the word metrical foot construction begins at. It can begin from either the left side (6, Ft Dir Left) or the right side (7, Ft Dir Right).

(6) Start from the left: (L   L   H

(7) Start from the right: L   L   H)

Then, the size of the metrical feet must be determined. A system could have no arbitrary limit on foot size. In an unbounded (Unb) system, a metrical foot is only closed upon encountering an H syllable or the word's end. Thus, there is an interaction with quantity sensitivity, which determines which syllables are H. Also, if a word has H syllables in it, the size of the metrical feet can be altered by feet directionality; starting metrical foot construction from the left (8) can yield different metrical feet than starting it from the right (9).

(8) Unb, Ft Dir Left: (L   L   L) (H   L)

(9) Unb, Ft Dir Right: (L   L   L   H) (L)

Another option is for metrical feet to be a specific, arbitrary size; these are Bounded systems. A metrical foot can be either 2 (10, B-2) or 3 units (11, B-3); units are either syllables (12, B-Syl) or moras (13, B-Mor). Only if the word edge is reached can metrical feet deviate from this size (e.g. in (10), the final foot contains only one syllable). If the counting units are syllables, there is no interaction with quantity sensitivity. It does not matter how the syllables are differentiated, or even if they are differentiated. However, if the counting units are moras, the weight of the syllable matters. H syllables count as 2 units while

L syllables are 1. This can lead to different metrical feet than counting by syllables would (e.g., compare 12b to 13).

(10) B-2, Ft Dir Left: (x   x) (x   x) (x)

(11) B-3, Ft Dir Left: (x   x   x) (x   x)

(12) B-2, B-Syl, Ft Dir Left

  (a) (L   H) (L   L)   (b) (H   H) (L   L)   (c) (S   S) (S   S)

(13) B-2, B-Mor, Ft Dir Left
    mora analysis   x x x x   x   x
    syllable classification   (H) (H) (L   L)

1.4 Stress Within a Metrical Foot

Once metrical feet are formed, one syllable per metrical foot is stressed. It is either the leftmost (Ft Hd Left, 14a) or rightmost syllable (Ft Hd Right, 14b).

(14)   (a) Ft Hd Left: (**H** L) (**L**)   (b) Ft Hd Right: (H **L**) (**L**)

1.5 Parametric Metrical Phonology Summary

We have now seen how metrical phonology parameters interact to generate the observable stress contour children encounter. Because of the interaction, it is difficult to tease apart which parameter value is responsible for generating a given stress contour. This is one difficulty of noisy data: ambiguity. Another difficulty is *exceptional* data, which will be discussed in the next section.

2 The Data Set and Target System: English

English was chosen as the target system because the English data children use as input are particularly noisy. Not only are data ambiguous, but there are numerous exceptions: nearly 27% of maximally-informative data indicate non-English values of the metrical phonology system (estimates from child-directed speech in the Bernstein–Ratner and Brent corpora[4] from CHILDES: MacWhinney (2000)). For example, the English system is extrametrical (Em-Some), but there are words with stress on both the first and last syllable (e.g. *af*ter***noon***) which are incompatible with either edge syllable being extrametrical.

The English system values are QS, QS-VC-H, Em-Some, Em-Right, Ft Dir Rt, Bounded, B-2, B-Syl, and Ft Hd Left. The exceptional data indicate QI, QS-VC-L, Em-None, Em-Left, Ft Dir Left, Unb, B-3, B-Mor, or Ft Hd Right as the correct target value. Converging on even one of these parameter values will

---

4. Children's age in these corpora ranged from 6 months to 2 years, which was an estimate for when components of the metrical phonology system might be learned.

lead the learner to a system that is not English (though it may be very similar parametrically). The child must therefore overcome the misleading data in order to converge on the correct target system.

Using corpora of child-directed speech, we can estimate the data distribution available to English. The Bernstein-Ratner and the Brent corpora of CHILDES (MacWhinney, 2000) yield 540505 words. Each word was then divided into syllables and assigned likely stress contours, referencing two psycholinguistic databases of syllabification and pronunciation (CALLHOME, Canavan et al. (1997); MRC, Wilson (1988)).

3 Identifying Unambiguous Data: Cues and Parsing

Given the noisy data of English, it may be difficult to identify unambiguous data for any single parameter value from only the stress contour. But, this is precisely what we propose children could do. There are two existing proposals for identifying unambiguous data for parameters in a complex linguistic system – *cues* (Dresher, 1999; Lightfoot, 1999) and *parsing* (Fodor, 1998, Sakas & Fodor, 2001). We examine these two proposals, and their instantiation for the parametric system of metrical phonology under consideration.

A cue is a "specific configuration in the input" associated with a parameter value (Dresher, 1999). It matches the observable form of a data point – here, the combination of syllable structure and stress. It may only match a portion of the word, instead of the entire word. Cues for each value of the metrical phonology system are given in Table 1, with an example of each cue in parentheses after the description of the cue. Note that some cues depend on the current state of the child's knowledge about the system (e.g. QS, Ft Dir Left, B-Syl, Ft Hd Left). To identify unambiguous data with cues, a child simply matches the cue to the observable data. This makes identification simple. However, the child must already have knowledge of what the cue is in order to learn this way.

The parsing method involves the learner using the structure-assigning ability, parsing, used during language comprehension (Sakas & Fodor, 2001). No additional knowledge beyond the parameters under consideration is necessary. The parsing instantiation presented here tries to analyze a data point with "all possible parameter value combinations" in the relevant parameter set, conducting an "exhaustive search of *all* parametric possibilities" (Fodor, 1998). A successful parameter value combination will generate a stress contour that matches the observed stress contour - this is then a successful parse of the data point. If all successful parses use only one of the available parameter values for a parameter, that data point is viewed as unambiguous for that parameter value.

For example, suppose the learner encounters '*af*ter***noon***', and successfully recognizes the syllables 'af' (VC), 'ter' (VC), and 'noon' (VV), and the associated stress contour (***VC*** VC ***VV***). A parsing child would try all available parameter value combinations and come up with 5 that are successful (15). These all share Em-None, meaning that Em-None was required for a successful parse. The child then views this data point as unambiguous for Em-None.

Table 1. Cues for metrical phonology parameter values.

Param	Cue
QI	Unstressed internal VV syllable (…*VV*…)
QS	Em-None or Em unknown: 2 syllable word with 2 stresses (*VV VC*) Em-Some: 3 syllable word, with 2 adjacent syllables stressed (VC *VV VC*)
QS-VC-L	Unstressed internal VC syllable (…*VC*…)
QS-VC-H	Em-None or Em unknown: 2 syllable word with 2 stresses, one or more are VC syllables (*VV VC*) Em-Some: 3 syllable word, with 2 adjacent syllables stressed, one or more are VC syllables (VC *VV VC*)
Em-None	Both edge syllables are stressed (*V*…*VC*)
Em-Some	One edge syllable is Heavy and unstressed (H…)
Em-Left	Leftmost syllable is Heavy and unstressed (H…)
Em-Right	Rightmost syllable is Heavy and unstressed (…H)
Ft Dir Left	QI or Q-unknown, Em-None/Left or Em unknown: 2 stressed adjacent syllables at right edge (…*VC V*) QI or Q-unknown, Em-Right: 2 stressed adjacent syllables followed by unstressed syllable at right edge (…*VC V* VV) QS, Em-None/Left or Em unknown: stressed H syllable followed by stressed L syllable at right edge (…*H L*) QS, Em-Right: stressed H syllable followed by stressed L syllable followed by unstressed syllable at right edge (…*H L* H)
Ft Dir Rt	QI or Q-unknown, Em-None/Right or Em unknown: 2 stressed adjacent syllables at left edge (*VC V*…) QI or Q-unknown, Em-Left: unstressed syllable followed by 2 stressed adjacent syllables at left edge (VC *V VV*…) QS, Em-None/Right or Em unknown: stressed L syllable followed by stressed H syllable at eft edge (*L H*…) QS, Em-Left: unstressed syllable followed by stressed L syllable followed by stressed H at left edge (H *L H*…)
Unb	QI: 3+ unstressed syllables in a row (…VC VV VC…) QS: 3+ unstressed Light syllables in a row (…L L L)
Bounded	Union of B-2 and B-3 cues
B-2	QI: 3+ syllables in a row, every other one stressed (… *VC* VV *VC*…) QS: 3+ Light syllables in a row, every other one stressed (…*L* L *L*…)
B-3	QI: 4+ syllables in a row, every third one stressed (…*V* VC VV *V*…) QS: 4+ Light syllables in a row, every third one stressed (…*L* L L *L*…)
B-Syl	QI: Union of QI B-2 and QI B-3 cues QS, B-2: 2 adjacent syllables, one stressed Heavy and one unstressed Light (…*H* L…) QS, B-3: 3 adjacent syllables, 2 unstressed Light preceding a stressed Heavy or following a stressed H (…*H* L L…), (…L L *H*…)
B-Mor	2 syllable word with both syllables Heavy and stressed (*H H*)
Ft Hd Left	Em-None or unknown: Leftmost syllable is stressed (*VC*…) Em-Left or unknown: 2nd from leftmost syllable is stressed (VV *VC*…)
Ft Hd Rt	Em-None: Rightmost syllable is stressed (…*VC*) Em-Right: 2nd from rightmost syllable is stressed (…*VC* VV)

Recall that the informativity of a data point changes over time. When all parameters are available (because none have been set yet), 5 successful parses

exist (15). However, suppose the child has some knowledge of the system, e.g. that English is Bounded. Then, the previously successful parse using the Unbounded parameter value (the last one in (15)) will no longer be tried, since it uses the wrong parameter value (Unb). The remaining parses have more in common than just Em-None; so, this very same data point will now be viewed as unambiguous for Em-None, Bounded, B-2, and B-Syl (16).

(15) Successful parameter value combinations for '**af**ter**noon**'

(QI, **Em-None**, Ft Dir Left, Bounded, B-2, B-Syl, Ft Hd Left)
(QI, **Em-None**, Ft Dir Rt, Bounded, B-2, B-Syl, Ft Hd Rt)
(QS, QS-VCL, **Em-None**, Ft Dir Left, Bounded, B-2, B-Syl, Ft Hd Left)
(QS, QS-VCL, **Em-None**, Ft Dir Rt, Bounded, B-2, B-Syl, Ft Hd Rt)
(QS, QS-VCL, **Em-None**, Ft Dir Left, Unb, Ft Hd Left)

(16) Successful parameter value combinations for '**af**ter**noon**': Bounded known

(QI, **Em-None**, Ft Dir Left, **Bounded, B-2, B-Syl**, Ft Hd Left)
(QI, **Em-None**, Ft Dir Rt, **Bounded, B-2, B-Syl**, Ft Hd Rt)
(QS, QS-VCL, **Em-None**, Ft Dir Left, **Bounded, B-2, B-Syl**, Ft Hd Left)
(QS, QS-VCL, **Em-None**, Ft Dir Rt, **Bounded, B-2, B-Syl**, Ft Hd Rt)

A strength of parsing is that, within the relevant parameter set, it will identify only truly unambiguous data. However, identification of unambiguous data is non-trivial since exhaustive-search parsing is resource-intensive.

Still, despite their weaknesses, cues and parsing both share an important strength: both are incremental, meaning that they extract information from a data point as it comes in, rather than requiring the child to hold a large collection of data in memory to conduct analyses on at some later point. This lends both cues and parsing some psychological plausibility as procedures real children could use to identify unambiguous data for learning.

4 Learning English

The outline of the learning procedure itself is as follows. The child encounters a data point and decides if it is unambiguous for any parameter values. If so, the probability of the corresponding parameter values are updated (either increased or decreased as appropriate). Now, if the child is trying to set a given parameter, the parameter value that has more unambiguous data in the input will eventually win – regardless of what particular probabilistic learning procedure is used[5]. So, the parameter value whose unambiguous data have a higher probability in the intake set will be the one the child chooses.

---

5. For instance, it could be a Linear Reward Penalty scheme (Yang, 2002), or some instantiation of Bayesian learning.

Recall that current knowledge of the system influences the data perceived as unambiguous. So, the parameters that are set influence the data the learner subsequently perceives as unambiguous for the unset parameters. The order in which parameters are set may thus determine if they are set correctly (Dresher, 1999). As an example, consider Tables 2 and 3, which show the probability of encountering unambiguous data for each available parameter value at the different points during learning. Table 2 shows the probabilities before any parameters are set. Table 3 shows the probabilities after QS is set - note that a sub-parameter has become available under QS: QS-VC-L vs. QS-VC-H. Also, note that the probabilities overall are quite small, since much of the input is ambiguous. Moreover, the probabilities can shift quite dramatically, depending on the child's current knowledge: before QS is set, Em-Some is much less probable than Em-None (Table 2); after QS is set, Em-Some is twice as probable as Em-None (Table 3).

Table 2. Initial probabilities of unambiguous data.

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 3. Probabilities of unambiguous data after QS is set.

QS VC Syllables		Extrametricality	
VC-L: 0.00265	VC-H: 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		

This explicitly shows how knowledge of the system changes the child's perception of unambiguous data; the parameter-setting order can be crucial. We can now examine the learnability of English metrical phonology from realistic distributions of English child-directed speech. In the worst case, no parameter-setting order will allow the child to converge on the English values. Learning from unambiguous data is thus insufficient. In a better case, there *are* some viable parameter-setting orders, even if there are some that fail. Learning from unambiguous data is sufficient. In the best case, all parameter-setting orders are viable, and learning English using only unambiguous data is a good strategy.

To determine which (if any) parameter-setting orders lead to English, we can follow the procedure in (17).

(17) Procedure for discovering viable parameter-setting orders
    (a) Calculate the probabilities of encountering unambiguous data for each parameter value.

(b) Choose one parameter value to set. The one chosen will have the higher probability in the data set, and this is the one a probabilistic learner will eventually converge on (e.g. QS over QI in Table 2).
(c) Repeat steps (a)-(b) until all parameters are set.
(d) If the final parameter values chosen are all the English values, this is a viable parameter-setting order.
(e) Repeat for all possible parameter-setting orders.

Can a child learning only from unambiguous data (identified by either cues or parsing) converge on the English system of metrical phonology?

5 Results: An Unambiguously Good Idea

It turns out that there are in fact some viable parameter-setting orders that will lead a child using unambiguous data to the English system (Table 4). [6] Though there are some orders that do not work (Table 5), learning from unambiguous data can still produce sufficient learning behavior. Given the complex parametric system and the noisy data set, this is no small feat.

Table 4. Examples of viable parameter-setting orders.

Cues
(1) (QS, QS-VC-H, Bounded, B-2, Ft Hd Left, Ft Dir Rt, Em-Some, Em-Right, B-Syl)
(2) (Bounded, B-2, Ft Hd Left, Ft Dir Rt, QS, QS-VC-H, Em-Some, Em-Right, B-Syl)
(3) (Ft Hd Left. Ft Dir Rt, QS, QS-VC-H, Bounded, Em-Some, Em-Rt, B-2, B-Syl)
Parsing
(1) (Bounded, QS, Ft Hd Left, Ft Dir Rt , QS-VC-H, B-Syl, Em-Some, Em-Right, B-2)
(2) (Ft Hd Left, QS, QS-VC-H, Bounded, Ft Dir Rt, Em-Some, Em-Right, B-Syl, B-2)
(3) (QS, Bounded, Ft Hd Left, QS-VC-H, Ft Dir Rt, B-Syl, Em-Some, Em-Rt, B-2)

Table 5. Examples of non-viable parameter-setting orders.

Cues
(1) (QS, QS-VC-H, Bounded, B-2, *B-Mor, …*)
(2) (Bounded, B-2, Ft Hd Left, *B-Mor, …*)
(3) (*Em-None, …*)
(4) (Ft Hd Left, *Em-None, …*)
Parsing
(1) (QS, QS-VC-H, Bounded, B-Syl, B-2, Em-Some, Em-Right, *Ft Hd Rt…*)
(2) (Bounded, B-Syl, B-2, *Em-None, …*)
(3) (*Em-None, …*)
(4) (Ft Hd Left, *Ft Dir Left, …*)

These parameter-setting orders represent knowledge the English child needs for acquisition success. That is, if the child happens to set the parameters in one of the viable orders, the child will converge on English. As it stands, this knowledge is an explicit listing of the viable orders. Yet it seems unsatisfactory for the child to have these orders explicitly known *a priori*.

---

6. Order read from left to right. Ex: (…B-2, B-Syl) means B-2 is set before B-Syl.

Fortunately, it turns out that the viable orders for both methods can be captured compactly by a small set of *order constraints* (Table 6). For cues, there are three constraints such that a parameter value *p* must be set before some other parameter value *q*. For parsing, there are three groups such that group one is set before group two, which is set before group three. If the child follows these order constraints, then the viable parameter-setting orders can be derived. Thus, to reach English, the child needs only to know this much smaller set of knowledge. Moreover, some of these order constraints can be derived from properties of the learning system such as data salience, data quantity, and default values (Pearl, 2007). In addition, the child may bring helpful biases from knowing other rhythmic properties of the language before word segmentation is reliable (Turk, Jusczyk, & Gerken, 1995; Jusczyk, Cutler, & Redanz, 1993).

Table 6. Order constraints for viable parameter-setting orders

Cues: Follow these constraints, other parameters freely ordered
(1) QS-VC-H before Em-Right
(2) Em-Right before B-Syl
(3) B-2 before B-Syl
Parsing: Group 1 before Group 2, Group 2 before Group 3
Group 1: QS, Ft Hd Left, Bounded
Group 2: Ft Dir Rt, QS-VC-H
Group 3: Em-Some, Em-Rt, B-2, B-Syl

6 Discussion & Future Directions

The main result from this study is that the English metrical phonology system is learnable using data perceived as unambiguous, a pleasantly surprising result given the complexity of the system and the noisiness of the data. This supports the viability of the parametric system as knowledge a learner could use to constrain the hypothesis space, since this system can be learned. It also supports the viability of using an unambiguous data filter for learning, since a child using this filter can in fact reach the target system of English.

Moreover, this study has generated predictions for the learning path we expect to see in real children, and can test experimentally. Specifically, if children are learning a parametric system from unambiguous data, we would expect them to follow the parameter-setting order constraints laid out in table 6. For example, whether the child is using cues or parsing, we would expect quantity sensitivity to be known before extrametricality.

Within the modeling realm, there are also several questions we can pursue. For data filtering, how successful is the unambiguous data filter for other languages and other parametric systems? Is there some way to combine the strengths of cues and parsing for identifying unambiguous data? Are there other methods for implementing a data filter, such as learning only from data perceived as systematic (Yang, 2005)? How necessary is data filtering, and how far can probabilistic learning take the child on its own (Yang, 2002; Fodor & Sakas, 2004)? We can also look to the learnability of other knowledge

implementations, such as constraint-satisfaction systems (Tesar & Smolensky, 2000) on noisy data sets. One mark of a system's viability is its learnability from realistic data.

In the modeling study here, we have demonstrated the viability of both the parametric system and the unambiguous data filter on realistic data. Modeling thus provides a way to explore questions of the learning mechanism that may be difficult to do with more standard experimental techniques. In addition, it can generate predictions that *can* be tested with experimental methodologies, thus dovetailing well with experimental research.

References

Canavan, Alexandra, Graff, David, and Zipperlen, George. (1997). *CALLHOME American English Speech*. Linguistic Data Consortium: Philadelphia, PA.

Chomsky, Noam. (1981). *Lectures on Government and Binding,* Foris, Dordrecht.

Clark, Robin. (1994). Kolmogorov complexity and the information content of parameters. *IRCS Report 94-17*. Institute for Research in Cognitive Science, University of Pennsylvania.

Dresher, B. Elan. (1999). Charting the Learning Path: Cues to Parameter Setting. *Linguistic Inquiry, 30*: 27-67.

Fodor, Janet D. (1998). Parsing to Learn. *Journal of Psycholinguistic Research*, 27(3): 339-374.

Halle, Morris. and Jean-Roger. Vergnaud (1987) *An Essay on Stress*. Cambridge: MIT Press.

Jusczyk, Peter, Cutler, Anne, & Redanz, Nancy. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development*, 64:675-687.

Lightfoot, David. (1999). *The development of language: Acquisition, change and evolution*. Blackwell: Oxford.

MacWhinney, Brian. (2000). *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.

Pearl, Lisa. (2007). *Necessary Bias in Natural Language Learning*. Doctoral dissertation, University of Maryland.

Pearl, Lisa & Weinberg, Amy (2007). Input Filtering in Syntactic Acquisition: Answers from Language Change Modeling, *Language Learning and Development, 3(1),* 43-72.

Sakas, William, and Janet D. Fodor. (2001). The Structural Triggers Learner, in S. Bertolo (ed.), *Language Acquisition and Learnability*, Cambridge University Press, 172-233.

Tesar, Bruce & Smolensky, Paul. (2000). *Learnability in Optimality Theory*. The MIT Press, Cambridge, Massachusetts.

Turk, Alice, Jusczyk, Peter & Gerken, LouAnn. (1995). Do English-learning Infants Use Syllable Weight to Determine Stress? *Language and Speech, 38(2)*: 143-158.

Wilson, Michael. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20(1): 6-11.

Yang, Charles. (2002). *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Yang, Charles. (2005). On productivity. *Yearbook of Language Variation*, 5, 333-370.