

More learnable than thou?

Testing metrical phonology representations with child-directed speech

LISA PEARL, TIMOTHY HO, & ZEPHYR DETRANO
*University of California, Irvine**

1 Introduction

One way to describe a language’s grammar is as a compact system that encodes the regularities of the language. This system allows someone to immediately comprehend and generate novel linguistic items that follow those encoded regularities, and so grammars are often viewed as generative systems. Notably, because languages vary with respect to the specific regularities they have, the generative system can be instantiated in various ways, based on language-specific input (e.g., as a specific set of parameter values in a parametric system or a specific ordering of constraints in a constraint-ranking system). The variables that can be used in a language’s grammar (e.g., the specific parameters or constraints) are defined by the knowledge representation (**KR**) and so a KR defines the set of possible grammars underlying human languages, based on those variables.

The utility of KRs for language acquisition then becomes apparent: if the child already has access to the KR, the hypothesis space of grammars that could encode the regularities of the language is already defined. So, the child already knows which variables in the linguistic environment matter, and can focus her attention on simply selecting the appropriate instantiation of the KR (i.e., the language-specific grammar), based on those relevant variables. The language acquisition task is about choosing the correct grammar for the language from those defined by the KR.

These two aspects of KRs lead to two natural criteria for any KR. The first criterion is the cross-linguistic variation criterion, which states that the right KR should be able to explain the constrained cross-linguistic variation we observe in the world’s languages. The cognitive premise of this kind of argument is that it is surprising to see such limited variation if there is no common underlying KR that humans are drawing their language-specific grammars from. KR theorizing then focuses on identifying the most compact representation than can account for the observed, limited variation. In this vein, Hayes (1995:55) notes, for example, that a successful representation of stress knowledge is one that is “capable of describing all the stress systems of the world’s languages” and is “maximally restrictive”.

The second criterion is the learnability criterion, and states that if children have access to

*We are very grateful to Joanna Lee for her help in preparing the child-directed speech corpus. We have additionally benefited from comments and suggestions from Pranav Anand, Matt Wagers, Sandy Chung, Armin Mester, Jim McCloskey, Adrian Brasoveanu, and the audiences at the Institute for Mathematical Behavioral Sciences 2013 colloquium at UC Irvine, the Logic & Philosophy of Science 2013 colloquium at UC Irvine, the Linguistics 2014 colloquium at UC Santa Cruz, and the Berkeley Linguistics Society 2014 meeting. All errors are of course our own and not at all their fault.

the right KR, they should be able to learn any language’s grammar from the language input they encounter. This again relates to the idea that the right KR helpfully circumscribes the child’s hypothesis space. However, previous work investigating a parametric KR (Pearl 2009, 2011) has suggested that acquisition is not so straightforward when the learner is given realistic child-directed data to learn from. That is, even with access to the KR, selecting the correct language-specific grammar when learning from realistic language data is non-trivial.

Traditionally, KRs have been proposed based on the first criterion of accounting for constrained cross-linguistic variation (e.g., Halle and Vergnaud 1987; Hayes 1995; Hammond 1999; Pater 2000; Tesar and Smolensky 2000; Prince and Smolensky 2002). This has led to several KR options in different linguistic domains (e.g., metrical phonology: parameters (Halle and Vergnaud 1987; Hayes 1995), violable constraints (Tesar and Smolensky 2000; Prince and Smolensky 2002)). While these KRs often overlap with respect to the linguistic aspects that comprise their variables, they do not rely on the exact same set of variables and so are not obviously notational variants of each other.

Notably, an often tacit assumption has been that a KR that satisfies the cross-linguistic variation criterion will easily satisfy the learnability criterion, precisely because the right KR highlights the relevant variables for the child (e.g., Chomsky 1981; Dresher 1999; Crain and Pietroski 2002). To choose among different KRs, which are viable with respect to the cross-linguistic variation criterion, it seems reasonable to examine them with respect to the learnability criterion. That is, since they can account for cross-linguistic variation, are the grammars they define also learnable?

More specifically, how learnable is the language-specific grammar within a KR, given realistic data from that language? To answer this, we need to define what it means to be learnable, empirically grounding a learnability assessment with available acquisition data and quantifying how learnable a grammar defined by a KR is. We then need to concretely test the KR with a specific language and the data children are using to learn that language’s grammar. Another often tacit assumption about KRs is that having the appropriate KR makes successful acquisition possible even in difficult acquisition scenarios – that is, the reason acquisition occurs so quickly for these “hard” cases is because the child has access to the KR, which tremendously simplifies the acquisition task. Therefore, an informative test for any KR is its ability to handle the hard acquisition cases, and so we should select a language whose grammar is likely to be non-trivial to learn when assessing learnability.

In the remainder of this paper, we first establish formal metrics for comparing KRs, based on a learnability analysis that is empirically grounded in child-directed speech data. We then demonstrate how to use this approach on a case study in metrical phonology, where we compare three KRs. We briefly review the KRs, and then discuss how we will evaluate them with respect to their ability to learn English, which is a notoriously difficult case for acquisition due to known irregularity in the data. We discover, perhaps surprisingly, that all three KRs have learnability issues when applied to learning English, and so the most straightforward argument from acquisition cannot be made for any of them. Nonetheless, we discuss various ways to solve the learnability issues for each KR, which include incorporating additional knowledge that can be derived during acquisition, equipping the learner with helpful learning biases that guide learning, and adjusting the definition of what the English grammar is within the KR. Thus, even though all three KRs have learnability issues for English at first glance, this does not mean we must immediately abandon those KRs –

instead, we can adjust different aspects of the acquisition problem to see if and how each KR’s English grammar can become learnable. In this way, we can identify KRs that satisfy both the cross-linguistic variation criterion and the learnability criterion, and so are likely to be more accurate descriptions of the mental representations of this linguistic knowledge.

2 Learnability metrics

2.1 The learnability approach

Many different approaches to assessing learnability exist (e.g., Pearl 2011; Clark and Lappin 2012; Legate and Yang 2013; Fulop and Chater 2013), and here we propose one that is similar to those used by Pearl (2011) and Legate and Yang (2013). We will assess learnability by (i) using realistic input and (ii) assessing it at the computational level (in the sense of Marr 1982). By evaluating learnability with realistic input, we can more concretely link learnability to the language acquisition task children actually face. By evaluating learnability at the computational level, we can focus on the utility of the hypothesis space defined by the KR: does this view of the relevant grammar variables easily lead the learner to that specific language’s grammar, given the available language data? Notably, this type of analysis focuses on the choices that a rational learner would make, given the current hypothesis space and learning preferences (Goldwater et al. 2009; Pearl et al. 2011; Perfors et al. 2011; Feldman et al. 2013; Dillon et al. 2013). It abstracts away from how that choice is actually made, given the cognitive resources available to children. A computational-level analysis can thus highlight if learnability issues *already* exist given a particular hypothesis space and learning assumptions, even before cognitive constraints come into play.

A rational learner will select what it perceives to be the best grammar, and we posit that the best grammar is the grammar able to account for the most data in the input perceived as relevant. This relates to the utility of grammars: a grammar is useful because it allows the learner to compactly represent the regularities in the language data, and so language data captured by the grammar do not need to be stored in detail. Instead, the relevant aspects of these data can be described by the compact representation provided by the grammar. So, the more data accounted for by the grammar, the more useful the grammar is because there are fewer data that must be dealt with separately (e.g., stored explicitly). Because of this, from a language use standpoint, the best grammar is naturally defined as the one that can account for the most data.

2.2 Specific learnability metrics

Once we define the set of data children are learning from, we can evaluate the grammars defined by any KR on their ability to account for those data. At an individual data point level, a grammar can be *compatible* with the data point. For example, a metrical phonology grammar is compatible with a data point if it can generate the observed stress contour for that data point. The proportion of data points a grammar is compatible with is its *raw compatibility* with that data set. For example, a grammar compatible with 70% of the data set has a raw compatibility of 0.70. When comparing grammars within a KR, a higher raw

compatibility is better since this indicates the grammar is more useful for accounting for the available data. Thus, the best grammar will have the highest raw compatibility, and be the most useful.

From a learnability perspective however, what matters more than raw compatibility is how a grammar compares to other grammars defined by the KR. This is captured by *relative compatibility*, which is how a grammar’s raw compatibility compares to the raw compatibilities of other grammars in the hypothesis space. We define a grammar’s relative compatibility as the proportion of grammars in the hypothesis space that this grammar is better than, with respect to raw compatibility. The best grammar will have a relative compatibility of approximately 1.00, since it will be better than all other grammars. For example, if there are 768 grammars, then the best grammar is better than 767, which gives a relative compatibility of $767/768 = 0.999$. Importantly, no matter what the raw compatibility of the best grammar is, it is the one a rational learner would choose because it is the best of all the grammars defined by the KR.

We can also evaluate the *learnability potential* of a KR, which is simply the raw compatibility of the best grammar (with relative compatibility ≈ 1.00) defined by the KR. In effect, the learnability potential indicates how good the grammar variables defined by the KR are at accounting for the available data in the learner’s input.

2.3 Evaluating the language-specific grammar

To satisfy the learnability criterion in the most straightforward way, the language-specific grammar should be the grammar that is learned most easily from the language’s child-directed input data. This can be empirically tested using the metrics above. If the language-specific grammar is the most easily learned grammar, it should have the highest raw compatibility, which will cause it to have a relative compatibility of approximately 1.00. This, in turn, would cause this grammar’s raw compatibility to be equivalent to the learnability potential of the KR that defines it, since it would be the grammar defined by that KR that is the best at accounting for the language’s child-directed input data.

3 Knowledge representations in metrical phonology

For metrical phonology, the observable data is the stress contour associated with a word. For example, *octopus* has stress on the first syllable, but not on the second and third syllables. We can represent this as *óctopus* (/óktəpʊs/) having the stress contour 100.¹ All the KRs we examine define grammars that assume a word has been divided into syllables and those syllables are classified according to their syllable rimes, so that syllable onsets are ignored (e.g., *strong* (/stɹɑŋ/) is equivalent to /tɹɑŋ/, /ɹɑŋ/, and /ɑŋ/). All grammars then form metrical feet comprised of one or more of those syllables, which we will indicate with parentheses, as in (1). Metrical feet are used for determining which syllables to stress, with a single syllable within a metrical foot being stressed.

¹Here we are only concerned with the distinction between stressed and unstressed syllables, rather than the additional consideration of primary vs. secondary stress among stressed syllables.

(1)	Sample metrical structure for <i>octopus</i> (/aktəpʊs/)			
stress	1	0	0	
metrical feet	(VC	V)	VC	
syllable rimes	VC	V	VC	
syllable IPA	ak	tə	pʊs	

A grammar defined by a KR will be associated with an underlying metrical structure, as shown in (1), whose observable form is the stress contour for the word. Importantly for our empirical purposes, each KR has defined a grammar that is meant to account for English, and so that is the grammar we will be particularly interested in evaluating against English child-directed input data. We now briefly review the three KRs we will compare, which include both parametric and constraint-ranking representations.

3.1 Parametric knowledge representations

3.1.1 The HV parametric representation

The first parametric KR is adapted from Halle and Vergnaud (1987) (**HV**), and its learnability has been previously investigated by Pearl (2007, 2011). The HV representation involves five main parameters with three sub-parameters, yielding 156 grammars in the hypothesis space. For a more detailed description of each of the parameters and their interactions with each other, see Dresher and Kaye (1990), Dresher (1999), and Pearl (2007).

Quantity Sensitivity. Quantity sensitivity determines whether syllables are treated identically or instead differentiated by syllable rime weight for the purposes of stress assignment. A language could be quantity sensitive (**QS**), so that syllables are differentiated into (**H**)eavy and (**L**)ight syllables. Long vowel syllables with or without codas (VV(C)) are Heavy, short vowel syllables (V) are Light, and short vowel syllables with codas (VC) can be either Light (**QS-VC-L**) or Heavy (**QS-VC-H**), yielding three syllable type distinctions (long, short, and closed).² In contrast, if the language is quantity insensitive (**QI**), all syllables are treated identically (represented below as *S*). Both kinds of analyses are shown in (2) for *beautiful*.

(2)	QS and QI analysis of <i>beautiful</i> (/bjutəfʊl/)							
	QS analysis	H	L	L/H	QI analysis	S	S	S
	syllable rime	VV	V	VC	syllable rime	VV	V	VC
	syllable IPA	bju	tə	fʊl	syllable IPA	bju	tə	fʊl

Extrametricality. Extrametricality determines whether all syllables of the word are contained in metrical feet. In languages allowing extrametricality, either the leftmost syllable (**Em-Left**) or the rightmost syllable (**Em-Rt**) is excluded (indicated by angled brackets ⟨...⟩). In contrast, languages without extrametricality (**Em-None**) have all syllables included in metrical feet. Example (3a) shows extrametricality applied to *giraffe* and *octopus*, while (3b) shows Em-None applied to *afternoon*.

²Vowel length in English typically corresponds to the tense/lax distinction, such that tense vowels (including diphthongs) are long, while lax vowels are short.

- (3) a. Extrametricality, with QS, QS-VC-H
- | | Em-Left | | | Em-Rt | |
|----------------|---------|-----|--|-------|--------|
| syllable class | ⟨L⟩ | H | | H | L ⟨H⟩ |
| syllable rime | V | VC | | VC | V VC |
| syllable IPA | dʒə | ˌæf | | ak | tə pʊs |
- b. No extrametricality (Em-None), with QS, QS-VC-L
- | | | | |
|----------------|----|-----|-----|
| syllable class | L | L | H |
| syllable rime | VC | VC | VVC |
| syllable IPA | æf | təɪ | nʌn |

Foot Directionality. Once the syllables to be included in metrical feet are known, metrical feet can be constructed. Feet can be constructed beginning at the left (**Ft-Dir-Left**), as in (4a), or the right (**Ft-Dir-Rt**), as in (4b).

- (4) a. Ft-Dir-Left, starting metrical foot construction from the left: (L L H
 b. Ft-Dir-Rt, starting metrical foot construction from the right: L L H)

Boundedness. The boundedness parameter determines the size of metrical feet. An unbounded (**Unb**) language has no arbitrary limit on foot size; a metrical foot is only closed upon encountering a Heavy syllable or the edge of the word. If there are no Heavy syllables or the syllables are undifferentiated (S) because the language is quantity insensitive, then the metrical foot encompasses all the non-extrametrical syllables in the word. Some example unbounded foot constructions are shown in (5).

- (5) Unbounded metrical foot construction
- a. Em-None, Ft-Dir-Left for L L L H L
- | | | | | | |
|----------------------------|----|---|----|----|----|
| begin | (L | L | L | H | L |
| heavy syllable encountered | (L | L | L) | (H | L |
| end | (L | L | L) | (H | L) |
- b. Em-None, Ft-Dir-Rt for S S S S S
- | | | | | | |
|-------|----|---|---|---|----|
| begin | S | S | S | S | S) |
| end | (S | S | S | S | S) |

The alternative is for metrical feet to be Bounded (**B**), and so to be no larger than a specific size. A metrical foot can be either two units (**B-2**) or three units (**B-3**); units are either syllables (**B-Syl**) or sub-syllabic units called moras (**B-Mor**) that are determined by the syllable's weight (Heavy syllables are two moras while Light syllables are one). Only if the word edge is reached can metrical feet deviate from this size (by being smaller than this size). Example (6) demonstrates different bounded foot constructions, with various combinations of these parameter values.

- (6) Sample bounded analyses of five-syllable sequences
- a. B-2, B-Syl with QS, Em-None, Ft-Dir-Left: (H L) (L L) (L)
- b. B-3, B-Syl with QI, Em-None, Ft-Dir-Left: (S S S) (S S)

- c. B-2, B-Mor with QS, Em-None, Ft-Dir-Left:
 mora analysis $\mu\mu$ $\mu \mu$ $\mu \mu$
 syllable classification (H) (L L) (L L)

Foot Headedness. Once the metrical feet are formed, the foot headedness parameter determines which syllable within a foot is stressed. Feet headed on the left have the leftmost syllable of the foot stressed (**Ft-Hd-Left**), shown in (7a), while feet headed on the right have the rightmost syllable of the foot stressed (**Ft-Hd-Rt**), shown in (7b).

- (7) Analyses for (L L) (L), which uses QS, Em-None, Ft-Dir-Left, B-2, B-Syl

- a. Ft-Hd-Left: $\begin{matrix} 1 & 0 & 1 \\ (L & L) & (L) \end{matrix}$
 b. Ft-Hd-Rt: $\begin{matrix} 0 & 1 & 1 \\ (L & L) & (L) \end{matrix}$

The English HV grammar. The English grammar for the HV representation differentiates syllables into Heavy and Light, treating VC syllables as Heavy (QS, QS-VC-H). The rightmost syllable of a word is extrametrical (Em-Rt), and metrical feet are built from the right side (Ft-Dir-Rt). A metrical foot spans two syllables (B, B-2, B-Syl), and the leftmost syllable within a foot is stressed (Ft-Hd-Left). A sample analysis using the English grammar is shown for *octopus* in (8). The generated stress contour (100) matches the observed stress contour (*óctopus*).

- (8) English grammar analysis for *octopus* (/aktəpʊs/):
 QS, QS-VC-H, Em-Rt, Ft-Dir-Rt, B, B-2, B-Syl, Ft-Hd-Left

stress 1 0 0
 analysis (H L) ⟨H⟩
 syllable IPA ak tə pʊs

3.1.2 The Hayes parametric representation

The second parametric system is adapted from Hayes (1995) (henceforth **Hayes**), and includes eight parameters that concern the basic distinction between stressed and unstressed syllables. These eight parameters yield 768 grammars in the hypothesis space.

Syllable Weight. Syllables are characterized as (H)heavy or (L)ight, similar to the QS option in the HV representation. Syllables with long vowels (VV) in their rimes are always Heavy, and syllables with short vowels only in their rimes (V) are always Light. Similar to the HV representation, closed syllables with a short vowel and one or more consonants (VC+) can be treated as either Heavy (**VC-H**) or Light (**VC-L**).

Extrametricality. Extrametricality is also similar to extrametricality in the HV system. In addition to no extrametricality (**Em-None**) and syllable extrametricality on the rightmost

(**Em-Right**) or leftmost (**Em-Left**) syllable, this representation also permits extrametricality on the rightmost consonant (**Em-RtCons**), where the rightmost consonant of a word is removed from metrical consideration. Notably, Em-RtCons can interact with syllable weight, as shown in (9). Because Em-RtCons can change the syllable type (e.g., turning a VC syllable into a V syllable), four syllabic distinctions are required in the Hayes representation: short (V), potentially short (VC), closed (VCC+), and long (VV or VVC+).

- (9) Sample syllable weight representations interacting with extrametricality, given VC-H

	Em-None		Em-RtCons	
syllable class	H	H	H	L
extrametricality	VV	VC	VV	V
syllable rime	VV	VC	VV	VC
syllable IPA	pe	pəɪ	pe	pəɪ
syllables	pa	per	pa	per

Foot Directionality. Similar to the HV representation, metrical foot construction can begin from the left edge (**Ft-Dir-Left**) or the right edge (**Ft-Dir-Rt**).

Parsing Locality. The parsing locality parameter indicates whether metrical feet are built as adjacently as possible. Strong local parsing (**LP-Strong**) requires that after a foot is constructed, the next foot should begin with the next syllable (10a). Weak local parsing (**LP-Weak**) requires that one Light syllable be skipped between feet (10b). Note that Heavy syllables are never skipped, even with weak local parsing.

- (10) Sample parsing locality foot construction, with feet comprised of exactly two syllables

- a. Em-None, Ft-Dir-Left, LP-Strong
- | | | | | | |
|-----------------|----|----|----|----|---|
| begin | L | H | L | L | L |
| start next foot | (L | H) | (L | L | L |
| end | (L | H) | (L | L) | L |
- b. Em-None, Ft-Dir-Left, LP-Weak
- | | | | | | |
|-----------------|----|----|---|----|----|
| begin | L | H | L | L | L |
| skip L syllable | (L | H) | L | (L | L |
| end | (L | H) | L | (L | L) |

Foot Inventory. When constructing metrical feet, there are three options: Syllabic Trochees (**Tro-Syl**), Moraic Trochees (**Tro-Mor**), and Iambs (**Iamb**). A Tro-Syl foot can take two forms: (i) two syllables of any weight with stress on the leftmost syllable (\acute{S} S), or (ii) a single stressed Heavy syllable at the end of metrical foot construction (\acute{H}). A Tro-Mor foot can also take two forms, based on the idea that each foot has two moras (L syllables = μ , H syllables = $\mu \mu$): (i) two Light syllables with stress on the leftmost syllable (\acute{L} L), or (ii) a single stressed Heavy syllable (\acute{H}). An Iamb foot can also take two forms: (i) a Light syllable followed by a syllable of any weight, with stress on the rightmost syllable (L \acute{S}), or (ii) a single stressed Heavy syllable (\acute{H}). Example (11) demonstrates foot construction for a word of form H L L H H with each of the different foot types.

- (11) Tro-Syl, Tro-Mor, and Iamb metrical feet built for H L L H H, given Em-None, Ft-Dir-Left, and LP-Strong

Tro-Syl						Tro-Mor					
Tro-Syl foot 1	(\acute{H}	L)	L	H	H	Tro-Mor foot 1	(\acute{H})	L	L	H	H
Tro-Syl foot 2	(\acute{H}	L)	(\acute{L}	H)	H	Tro-Mor foot 2	(\acute{H})	(\acute{L}	L)	H	H
Tro-Syl foot 3	(\acute{H}	L)	(\acute{L}	H)	(\acute{H})	Tro-Mor foot 3	(\acute{H})	(\acute{L}	L)	(\acute{H})	H
						Tro-Mor foot 4	(\acute{H})	(\acute{L}	L)	(\acute{H})	(\acute{H})
Iamb											
Iamb foot 1	(\acute{H})	L	L	H	H						
Iamb foot 2	(\acute{H})	(L	\acute{L})	H	H						
Iamb foot 3	(\acute{H})	(L	\acute{L})	(\acute{H})	H						
Iamb foot 4	(\acute{H})	(L	\acute{L})	(\acute{H})	(\acute{H})						

Degenerate Feet. After constructing feet, edge syllables may remain unfooted. If a language has a strong prohibition against degenerate feet (**DF-Strong**) and an edge syllable is unfooted, a degenerate foot is not allowed to form and the analysis fails (12, lefthand side). If a language instead has a weak prohibition against degenerate feet (**DF-Weak**), a degenerate foot may form if the remaining syllable is Light (12, righthand side).

- (12) Analyses of L H with DF-Strong and DF-Weak, given Em-Right and Tro-Mor

DF-Strong				DF-Weak			
H extrametrical	L	<H>		H extrametrical	L	<H>	
L too small for Tro-Mor foot.	L	<H>		L too small for Tro-Mor foot.	L	<H>	
L \neq degenerate foot.	L	<H>		L = degenerate foot.	(\acute{L})	<H>	
Analysis crashes.							

Word Layer End Rule. The Word Layer End Rule (WLER) can interact with degenerate feet and the analysis direction (see next section) to alter the observable stress contour. If degenerate feet are formed (due to DF-Weak), the WLER determines whether the stress on the degenerate foot survives. WLER can be set to either Left (**WLER-L**) or Right (**WLER-R**) and will allow the stress of any degenerate foot to survive if it is closer to the corresponding edge of the word than any other foot. For example, in a WLER-R language with a degenerate foot on the right edge of the word, the degenerate foot's stress will survive (13a). In contrast, if the degenerate foot is on the left edge of the word and there are additional feet closer to the right edge, the degenerate foot's stress will not survive (13b).

- (13) Sample analyses of word form L L H L L, showing the interaction of Ft-Dir-Left and Ft-Dir-Rt with WLER-R

- a. Em-None, **Ft-Dir-Left**, Tro-Syl, LP-Strong, DF-Weak, **WLER-R**

Tro-Mor foot 1	(\acute{L}	L)	H	L	L
Tro-Mor foot 2	(\acute{L}	L)	(\acute{H}	L)	L
Degenerate foot	(\acute{L}	L)	(\acute{H}	L)	(\acute{L})
Degenerate foot stress survives	(\acute{L}	L)	(\acute{H}	L)	(\acute{L})

- b. Em-None, **Ft-Dir-Rt**, Tro-Syl, LP-Strong, DF-Weak, **WLER-R**
- | | | | | | |
|---|-----------------|---------------|----|---------------|----|
| Tro-Mor foot 1 | L | L | H | (\acute{L} | L) |
| Tro-Mor foot 2 | L | (\acute{L} | H) | (\acute{L} | L) |
| Degenerate foot | (\acute{L}) | (\acute{L} | H) | (\acute{L} | L) |
| Degenerate foot stress does not survive | (L) | (\acute{L} | H) | (\acute{L} | L) |

Stress Analysis Direction. This parameter determines whether metrical stress analysis begins with creating feet and then determining word-level stress via WLER (**Bot-Up**) or begins with word-level analysis using the WLER and subsequently creates feet (**Top-Down**). Notably, in Top-Down languages, the WLER decides whether the initial (WLER-L) or final (WLER-R) syllable should be stressed, regardless of weight. Parsing of syllables into feet is then constrained by the stress assigned by the WLER at word level. All previous analyses presented have used the Bot-Up value. We demonstrate in (14) how stress analysis direction can interact with the WLER.

- (14) Sample analyses of word form L H using Bot-Up versus Top-Down, with Em-None, Ft-Dir-Right, Iamb, LP-Strong, DF-Weak, WLER-L
- | Bot-Up | Top-Down |
|--------------------------------------|--|
| Iamb foot (L \acute{H}) | WLER stresses leftmost syllable \acute{L} H |
| No degenerate feet. (L \acute{H}) | Cannot create (L \acute{H}) Iamb foot \acute{L} (\acute{H}) |
| Word-level stress remains as is. | due to \acute{L} , so create (H) Iamb foot |
| | DF-Weak creates degenerate foot (\acute{L}) (\acute{H}) |

The Hayes English grammar. The English grammar for the Hayes representation treats VC syllables as Heavy (VC-H) and views the rightmost consonant as extrametrical (Em-RtCons). Metrical feet are built from the right (FtDir-Rt) as adjacently as possible (LP-Strong), and are two moras in size with the leftmost syllable stressed (Tro-Mor). Degenerate feet are not allowed (DF-Strong), so although stress on a degenerate foot would be allowed to survive if it was the rightmost syllable (WLER-R), this aspect does not matter for this layer of metrical stress in English (though WLER-R does matter for distinguishing between primary and secondary stress one layer above). In addition, metrical feet are created before word-level stress is assigned (Bot-Up). A sample analysis using the English grammar is shown for *octopus* in (15). Note that the English grammar generates the incorrect stress contour for this word (110 instead of the observed 100).

- (15) English grammar analysis for *octopus* (/aktəpʊs/):
- | |
|--|
| VC-H, Em-RtCons, FtDir-Rt, LP-Strong, Tro-Mor, DF-Strong, WLER-R, Bot-Up |
| stress 1 1 0 |
| analysis (\acute{H}) (\acute{L}) L) |
| syllables ak tə pʊ<s> |

3.2 Constraint-based representations


Optimality Theory (**OT**) (Tesar and Smolensky 2000; Prince and Smolensky 2002) characterizes linguistic knowledge as a universal set of constraints whose interaction determines

the form of observable linguistic data, and a language's grammar is a ranking of these constraints. Given n constraints, there are $n!$ rankings. In our instantiation of OT (Hammond 1999; Pater 2000), there are nine phonological constraints, defining a hypothesis space of $9! = 362,880$ grammars. Additionally, there is one inviolable principle called **ROOTING**, which requires all words to have some stress on them and so entails that their analyses contain at least one metrical foot.

3.2.1 Constraints


Non-Finality (NonFin). The final syllable is unfooted. In (16), the first candidate form for *little* (/lɪrl/) is preferred since the final syllable is not included in a metrical foot.

(16)

Input: /lɪrl/	NONFIN
a.  (lɪrl)	
b. lɪ (rɪl)	*!


Trochaic (Tro). Feet are headed on the left. In (17), the first candidate for *mommy* (/mami/) is preferred since its sole foot has stress on the leftmost syllable.

(17)

Input: /mɑ mi/	TRO
a.  (má mi)	
b. (mɑ mí)	*!


Weight-to-Stress Principle VV (WSP-VV). Syllables with long vowels should be stressed. The first candidate in (18) for *canoe* (/kənu/) is preferred since its second syllable has a VV rime and is stressed.

(18)

Input: /kə nu/	WSP-VV
a.  (kə nú)	
b. (kʌ nu)	*!

Weight-to-Stress Principle VC (WSP-VC). Syllables closed by consonants should be stressed. The first candidate in (19) for *little* (/lɪrl/) is preferred since its second syllable has a VC rime and is stressed.

(19)

Input: /lɪrl/	WSP-VC
a.  (lɪ rɪl)	
b. (lɪ rl)	*!

Foot Binarity (FtBin). Feet are binary (contain two units) at some level of analysis (e.g., syllables or moras). The first candidate for *little* in (20) is preferred since the sole metrical foot contains two syllables.

(20)

	Input: /lɪ rɪl/	FTBIN
a.	(lɪ rɪl)	
b.	(lɪ) rɪl	*!

Align Right (Align-R). Align the right edge of a foot to the right edge of the prosodic word. This constraint prefers metrical feet to have their right edge as close as possible to the right edge of the word, and so the third candidate for *horizon* (həɾɔjzən) in (21) is preferred.

(21)

	Input: /hə ɾɔj zən/	ALIGN-R
a.	(hɹ) ɾɔj zən	*!*
b.	hə (ɾɔj) zən	*!
c.	(hə ɾɔj zən)	

Align Left (Align-L). Align the left edge of a foot to the left edge of the prosodic word. This constraint prefers metrical feet to have their left edge as close as possible to the left edge of the word, and so the first candidate for *horizon* in (22) is preferred.

(22)

	Input: /hə ɾɔj zən/	ALIGN-L
a.	(hə ɾɔj zən)	
b.	hə (ɾɔj) zən	*!
c.	hə ɾɔj (zən)	*!*


Parse-Syllable (Parse- σ). Syllables must belong to feet. Extrametrical syllables violate this constraint and so the first candidate for *mommy* in (23) is preferred.

(23)

	Input: /mɑ mi/	PARSE- σ
a.	(mɑ mi)	
b.	(mɑ) mi	*!

***Sonorant Nucleus (*SonNuc).** Syllables should avoid having sonorant nuclei. The first candidate for *little* in (24) is preferred since none of its syllables have sonorant nuclei.

(24)

Input: /lɪ rɪl/	*SONNUC
a.  (lɪ rɪl)	
b. (lɪ rɪl)	*!

3.2.2 Syllabic distinctions

These constraints require eight syllabic distinctions, which divide syllables generally into short, closed, long, and super-long variants. The short variants are these: (i) short vowel open (V), as in the first syllable of *kitty* (/kɪ ri/), and (ii) sonorant nucleus (R), as in the second syllable of *actor* (/æk tr/). The closed variants are these: (i) short vowel closed (VC), as in *took* (/tʊk/), (ii) short vowel closed by a sonorant consonant (VR), as in *them* (/ðɛm/), (iii) short vowel closed by a sonorant consonant and another consonant (VRC), as in *tent* (/tɛnt/), and (iv) sonorant nucleus closed by another consonant (RC), as in *heard* (/hɜrd/). The long variant is a long vowel (VV), as in the second syllable of *kitty* (/kɪ ri/), and the super-long variant is a long vowel closed with a consonant (VVC), as in *boot* (/but/).

3.2.3 The OT English grammar

The OT “grammar” for a language is often a partial ordering of constraints, and so corresponds to multiple grammars that are explicit rankings of all nine constraints. In this vein, the English grammar described by Hammond (1999) and Pater (2000) obeys ten constraint ranking relationships, which correspond to 26 grammars that explicitly rank all nine constraints. This partial ordering is shown in Figure 1, where each arrow represents a constraint ordering that is true of the English grammar.

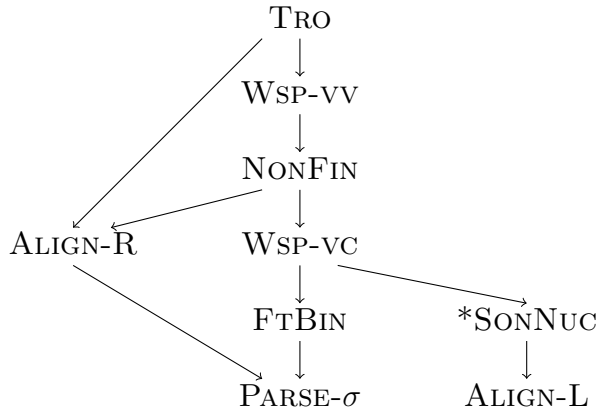


Figure 1: Partial ordering of constraints defining the English grammar.

The tableau below is an evaluation of a grammar satisfying the English constraint rankings on *little* (/lɪrlɪl/). Because the final /l/ could be the nucleus of the second syllable, eight candidates are generated. The optimal candidate for the grammar has a stress contour that matches the observed stress contour of *little* (lɪttle).

Input: /lɪ rɪl/	TRO	WSP-VV	NONFIN	WSP-VC	ALIGN-R	FTBIN	PARSE- σ	*SONNUC	ALIGN-L
(lí rɪl)			*!	*	*				
(lɪ rɪl)	*!		*						*
(lí) rɪl				*!	*	*	*		
lɪ (rɪl)			*!			*	*		*
(lí)(rɪl)			*!			**			
(lí) rɪl			*!		*			*	
(lɪ) rɪl	*!		*					*	*
☞ (lí) rɪl					*	*	*	*	
lɪ (rɪl)			*!			*	*	*	*
(lí)(rɪl)			*!			**		*	

4 English metrical phonology

English metrical phonology is an excellent test case for metrical phonology KRs, since the data are irregular and make acquisition of the target knowledge difficult. So, if a learner using a particular KR can succeed at acquiring the English grammar from realistic English input data, this provides very good support for the utility of this KR for acquisition.

But what makes English data so difficult with respect to acquisition? The first issue is that many data are ambiguous for which parameter value or constraint-ranking they implicate, due to parameter or constraint interaction. For example, consider two grammars defined by the HV parametric KR that *cucumber* (/kʲukʌmbəɪ/) is compatible with, shown in (25). Quite strikingly, these two grammars have no parameter values whatsoever in common, yet are able to generate the same stress contour (contour 110).

- (25) Two grammars *cucumber* is compatible with
- QI, Em-None, Ft-Dir-Rt, B, B-2, B-Syl, Ft-Hd-Left
 Analysis (Ś) (Ś S)
 Syllables kju kʌm bəɪ
 - QS, QS-VC-H, Em-Some, Em-Right, Ft-Dir-Left, Unb, Ft-Hd-Rt
 Analysis (Ĥ) (Ĥ) (H)
 Syllables kju kʌm bəɪ

Data ambiguity is a common problem for language acquisition – in fact, the infamous poverty of the stimulus concerns exactly this issue (e.g., Chomsky 1980; Baker and McCarthy 1981; Hornstein and Lightfoot 1981; Crain 1991; Pinker 2004; Pearl and Mis 2011). Clearly English metrical phonology is no exception. We can easily see that the observable stress contour data can be compatible with multiple hypotheses about the underlying structure from (25) above.

English metrical phonology data have another less common problem, however: There are numerous exceptions to the underlying system representing the target grammar, no matter which grammar is selected as the target grammar (Pearl 2011). How could this be? First, there are known interactions with both morphology (Chomsky and Halle 1968; Kiparsky

1979; Hayes 1982) and grammatical category (Chomsky and Halle 1968; Hayes 1982; Kelly 1988; Kelly and Bock 1988; Hammond 1999; Cassidy and Kelly 2001). For example, in *prétty/préttier/préttiest* and *sensátion/sensátional/sensátionally*, adding inflectional and derivational morphology does not shift the stress, despite adding syllables to the word. This would be unexpected in the purely phonological systems described by the KRs above, since additional syllables typically alter which syllables are stressed in a word. For grammatical categories, there are examples like *cónduct/condúct* and *désert/desért*, where the grammatical category influences the stress pattern (i.e., nouns are stress-initial while verbs are stress-final). This is again unexpected in the purely phonological systems described above, since they would generate/select a single stress pattern for a syllabic word form (i.e., abstracted to syllable rime, so *conduct* is VC VC), irrespective of grammatical category.

Notably, these irregularities in the data can cause multiple stress contours to appear for a single syllabic word form, as we saw in the grammatical category examples above. This is problematic because, as mentioned above, a grammar can only generate/select a single stress contour per syllabic word form. This means there is no way for a single grammar – no matter which grammar it is – to account for all the English data in a learner’s input.

But how often does a syllabic word form have multiple stress contours associated with it in realistic acquisition data? We examined the Brent corpus of the American English subsection of CHILDES (MacWhinney 2000), which contains speech directed at children between the ages of six and twelve months (99968 multisyllabic word tokens, 4780 multisyllabic word types). If we examine only non-monosyllabic words, under the assumption that at least two syllables are required to have a stress contour, we find this issue occurs quite often: 51% (HV), 58% (Hayes), or 37% (OT) of all syllabic word forms have multiple stress contours. This underscores why no single grammar can be compatible with all the input data, and thus why acquisition of the target grammar for English may be difficult, given realistic English acquisition input. In particular, it will be impossible for the English grammar in any of these KRs to account for all the input data, due to the numerous irregularities in the input data.

Clearly, the interactions between metrical phonology, morphology, and grammatical category that lead to some of these irregularities are part of the complete target knowledge for English. However, children may not hypothesize these interactions when they first begin learning grammars for metrical phonology (which Kehoe 1998 suggests occurs before the age of two). Thus, in the initial stages of English metrical phonology acquisition, children may assume the metrical phonology system is autonomous and only look within the phonological KRs to select the grammar that best accounts for the input, perhaps noting that there are irregularities that must be accounted for later on.

5 Learnability comparison for English input data

5.1 Learnability potential

Given how many syllabic word forms have multiple stress contours, it is reasonable to wonder how well any one grammar within these KRs could possibly do. In particular, what is the largest quantity of data that any single grammar can account for? This represents the learnability potential of the KR. It turns out that all three KRs have a grammar that is

able to account for about $\frac{2}{3}$ of the word types (0.657-0.683) and about $\frac{3}{4}$ of the word tokens (0.729-0.750), as shown in Table 1. This suggests that the best grammar in each knowledge representation is quite useful to have, since it can account for a large portion of the input (even if not all the input can be accounted for). Therefore, each KR is capable of defining a grammar that would be useful for the child to acquire.

5.2 English grammar compatibility

Since it is possible to learn a useful grammar from these data, the next reasonable question is whether the English grammar is the most useful one to learn. This is indicated by the English grammar’s compatibility with the English input data. Table 1 shows that the English grammar in all three KRs (or the best instantiation of the English grammar, in the case of the OT representation) is *not* compatible with as much data as the best grammar (types: 0.485-0.593, tokens: 0.531-0.716). The (best) English grammar is clearly not the most compatible grammar, and so a rational learner looking for the grammar capable of accounting for the most input data would not select it.

But recall that raw compatibility does not matter as much as relative compatibility, since a learner is selecting a grammar from a circumscribed hypothesis space. Though the (best) English grammar accounts for fewer data than the best grammar, how does it compare to the rest of the grammars that are available? It could be that the (best) English grammar, while having a significantly lower raw compatibility than the best grammar, is the next best grammar overall for accounting for the English input data. If that were true, children might have a better chance of selecting the English grammar, especially if they are not perfectly rational learners. That is, if the relative compatibility of the (best) English grammar is very high, children may still be able to learn it fairly easily from English input.

Unfortunately, this turns out not to be true for any of the knowledge representations. As Table 1 shows, the parametric English grammars are better than about $\frac{2}{3}$ of the grammars in the hypothesis space (types: 0.673-0.676, tokens: 0.673-0.685) and the best constraint-based grammar is better than about $\frac{4}{5}$ of the grammars in the hypothesis space (types: 0.817, tokens: 0.785). This indicates that the English grammars are better than many other grammars – but there are a large number of grammars that are better than the English grammars. For the parametric KRs, tens or hundreds of grammars are better able to account for the English input (types: HV=51, Hayes=249; tokens: HV=51, Hayes=242) while for the constraint-based KR, tens of thousands of grammars are better (types: OT=66,407, tokens: OT=78,019). In short, the target English grammar is unlikely to be easily learnable from this hypothesis space of grammars.

5.3 Learnability summary

For all three KRs, there are learnability issues. Using realistic English child-directed input, it is difficult to learn the English grammar in each KR. More specifically, a rational learner looking for the grammar best able to account for the observable English input would not select the target English grammar in any of these KRs.

Table 1: Learnability analyses for the three knowledge representations: HV, Hayes, and OT. The three metrics shown are learnability potential of the knowledge representation (KR:Pot), raw compatibility of the (best) English grammar (Eng:Raw), and relative compatibility of the (best) English grammar (Eng:Rel), which are computed over word types and word tokens in English child-directed speech. The word token score is shown in parentheses.

	KR:Pot	Eng:Raw	Eng:Rel
HV	0.668 (0.739)	0.593 (0.716)	0.673 (0.673)
Hayes	0.683 (0.750)	0.485 (0.531)	0.676 (0.685)
OT	0.657 (0.729)	0.573 (0.574)	0.817 (0.785)

6 Addressing the learnability issues

The learnability problem can effectively be summarized as the learner not being able to learn the target grammar, given the initial knowledge state provided by the KR and realistic English input. Below we present three potential ways around this apparent problem so that the three KRs could satisfy learnability for English.

6.1 Intermediate knowledge states

Experimental data suggest that there may be several intermediate knowledge states when learning English metrical phonology. First, there is a trochaic metrical template used at age two (Gerken 1994, 1996), which could implement a preference for a quantity insensitive metrical foot spanning two syllables, with stress on the leftmost syllable. Then, by age three, children have recognized that the metrical system is quantity sensitive, but not that the rightmost syllable is typically extrametrical (Kehoe 1998). It therefore seems possible that there are additional transitory states before the final knowledge state is achieved at age four or five (e.g., Pettinato and Verhoeven 2008; Arciuli et al. 2010).

One reason that there might be intermediate knowledge states is that children may perceive the input differently as they gain more linguistic knowledge. For example, when learning the metrical phonology system, gaining knowledge about the interaction between metrical phonology and morphology would allow children to perceive and analyze their input data differently. This new analysis could then cause them to abandon an intermediate non-target grammar and instead learn the target English grammar, because the target English grammar would then become the one able to account for the most input data.

One useful piece of knowledge to acquire is that productive affixes in English tend to be stressless or bear weak stress if there are multiple stressed syllables in a word (Hayes 1995). For example, in *sensationally*, the derivational affixes *-al* and *-ly* are not stressed, and in *prettiest*, the inflectional affix *-est* is not stressed. But when do children acquire knowledge of productive English affixes, and is it early enough that they’re likely to use this knowledge when acquiring the English metrical phonology grammar? While knowledge of derivational morphology appears to develop fairly late (well into primary school, where it may be explicitly instructed (Tyler and Nagy 1989; Stotko 1994; McBride-Chang et al. 2005; Jarmulowicz et al. 2008)), children develop knowledge of inflectional morphology much earlier, often using it

productively in their own utterances by age three (Brown 1973). Given their own usage of inflectional morphology, it is possible that children have noticed by this age that inflectional morphology is not stressed and rarely alters the stress on a word. They could then apply this acquired knowledge when learning the target English metrical phonology grammar. In particular, they could ignore inflectional morphology when attempting to determine the analysis underlying an observable stress contour. Thus, *préttiest* (/p.ɹiɪst/) would be viewed as *prétti* (/p.ɹi/) for the purposes of learning the metrical phonology grammar.

To investigate the impact of this kind of acquired knowledge, we re-analyzed the English input data for their compatibility with the various grammars after removing inflectional morphology (plural *-s*, possessive *'s*, past tense *-ed*, progressive *-ing*, past participle *-en*, comparative *-er*, and superlative *-est*).³ This simulates the learner ignoring inflectional morphology in the input when learning the English metrical phonology grammar.

How does the (best) English grammar compare to the rest of the grammars in the hypothesis space, once the learner has this acquired knowledge about inflectional morphology? Does this now make the (best) English grammar have a high relative compatibility? In fact, it does not: HV = 0.712, Hayes = 0.704, and OT = 0.786 by types. This again indicates that the English grammars are better than many other grammars (between 70 and 80% of the hypothesis space) – but there are still a large number of grammars that are better than the English grammars. Once again, the target English grammar is unlikely to be easily learnable from this hypothesis space of grammars, even with this acquired knowledge.

One pervasive issue that occurs is that the English grammars in all three KRs typically want syllables with a long vowel nucleus (e.g., *sweet*) to be stressed. This can be problematic for realistic child-directed speech since many words (and often very frequent words) have unstressed long vowel syllables, such as *báby* (2158 tokens) and *kítty* (1261 tokens).

One way to deal with these problematic data is to acquire additional knowledge that will allow the learner to view them as compatible with the English grammar. For example, the perception that the diminutive affix /i/ is a kind of inflectional morphology could occur because it communicates affection and attaches to the root form of a noun as a suffix (e.g., *dog* becomes *doggie*), occasionally altering the root form in the process (e.g., *cat* becomes *kitty*). It is unclear when children acquire knowledge of the diminutive in English, but if they are able to use it productively around the time when they productively use other inflectional morphology, then it is likely they acquire it while they are learning the metrical phonology grammar of English. They could then use this knowledge to perceive these diminutive data differently, ignoring the diminutive affix for purposes of metrical phonology. The diminutives then become compatible with the English grammars in all three KRs. This general strategy of ignoring certain aspects of the input data could be very helpful, since so many stress data are irregular in English. In the next section, we discuss other ways children might utilize this type of selective learning strategy on English input.

³Note that if a word was previously multisyllabic but became monosyllabic after ignoring inflectional morphology (e.g., *sweetest* becoming *sweet*), it was analyzed as being compatible with all grammars.

6.2 More sophisticated learning

Perhaps the learning process is more sophisticated, with the learner having useful prior knowledge that guides learning. Thus, the initial knowledge state also includes helpful learning biases for navigating the hypothesis space defined by the KR. Both of the learning biases we discuss below are predicated on the same basic idea: instead of trying to account for all of the input data with a grammar, the learner tries to account for only a subset of the input data that are perceived as relevant for determining the correct grammar. In essence, the learner’s data intake (Fodor 1998; Pearl 2007; Gagliardi et al. 2012; Gagliardi 2013) is a subset of the available input, due to these learning biases.

The first learning bias of this kind is to learn only from data perceived as unambiguous by the learner (Fodor 1998; Pearl 2008). This might result from a more general bias to prefer highly informative data, where unambiguous data would be viewed as maximally informative. Pearl (2008, 2011) demonstrated how this kind of bias could be used to learn the HV English grammar from realistic English child-directed data. Data points were viewed as potentially unambiguous with respect to a particular parameter value in the HV KR, e.g., a bisyllabic word with stress on the leftmost syllable like *báby* would be viewed as unambiguous for metrical feet headed on the left (Ft-Hd-Left). This allowed the learner to identify a very small subset of useful data points (never more than 5% of the input for any parameter value). When coupled with some additional knowledge about the order in which parameters must be learned, this unambiguous data bias allowed the learner to successfully navigate the HV hypothesis space of grammars. Thus, this more sophisticated learner did not encounter the learnability problem we discovered here for the HV KR with an unbiased learner. It is therefore possible that the other KRs would also surmount their learnability issues if the learner were equipped with these helpful learning biases.

In a similar vein, another potentially helpful learning bias is to learn only from data viewed as regular (rather than irregular), with the idea that a regular data point will have a productive rule associated with it (Legate and Yang 2013). Each productive rule is then something the learner is interested in capturing with the grammar for the language. One way children might implement this bias when learning metrical phonology is to assume that for every syllabic word form that has multiple stress contours (e.g., V VV: *kítty*, *awáy*, *úh óh*), one stress contour may be the regular, productive stress contour while the others are exceptions. A formal way to identify if there is a productive rule for a set of items is the Tolerance Principle (Yang 2005; Legate and Yang 2013), which is used to estimate how many exceptions a rule can tolerate before it’s no longer useful for the learner to have the rule at all. Effectively, if there are too many exceptions, it is better to simply deal with the exceptions on an individual basis rather than bothering to learn a rule that is often violated. For N items, the total exceptions a rule can tolerate is $\frac{N}{\ln N}$. If there are more exceptions than this, then the rule is not productive.

The metrical phonology learner would apply the Tolerance Principle when considering any syllabic word form with multiple stress contours. At any point during acquisition, there are two possible outcomes. One option is that one contour may be the regular contour according to the Tolerance Principle, and so the learner would attempt to account for only the data with that stress contour (e.g., *kítty*), ignoring the other data for that syllable word form (e.g., *awáy*, *úh óh*) when trying to learn the grammar. The other option is that no

contour is regular according to the Tolerance Principle, and so all the data for that syllable word form are ignored when trying to learn the grammar. Similar to a learner using the unambiguous data bias, it may be that this bias to learn only from regular data helps the learner perceive the input in a way that causes the English grammar in each KR to be compatible with far more data, and so surmount the apparent learnability problem.

6.3 Different target states for English

A third way to deal with the learnability problem is to simply change what the target grammar for English is. One idea is to look at the grammars within each KR that are more compatible with the English child-directed data, and examine what about these high compatibility grammars makes them more compatible. We examine each KR in turn, looking at the grammars that have ≥ 0.70 token compatibility, since that seemed to be a good threshold that separated the high-compatibility grammars from the lower-compatibility grammars.

For the HV KR, it turns out that the grammar currently defined for English uses all the parameter values that the majority of the high-compatibility grammars do (e.g., 53 of 58 high-compatibility grammars use the Em-Rt value, which makes the rightmost syllable extrametrical, and this is the same parameter value the current English grammar uses). So, there is no one value that is obviously “wrong” for English child-directed input – it’s simply that when all these parameter values are combined, they’re sub-optimal. Therefore, there is no obvious change to the target grammar definition that will improve learnability for the HV KR. However, this does suggest that a learner attempting to learn these parameter values individually may succeed, and that may be why the unambiguous data bias used in Pearl (2008) worked so well – it forced the learner to learn the values separately, since data are unambiguous only with respect to a single parameter value.

However, it turns out that the Hayes KR can benefit significantly by altering some of the values in the current definition of the English grammar. One change is to alter the form of the metrical foot from moraic trochees (Tro-Mor) to syllabic trochees (Tro-Syl). This would allow the grammar to account for bisyllabic words with an unstressed Heavy syllable at the end, such as *báby* and *kítty*. Making this update boosts the relative compatibility to 0.910 by types. Thus, this new English grammar with the Tro-Syl foot inventory value now is far more learnable compared to other grammars in the Hayes hypothesis space, as it is better at accounting for the English child-directed data than 9 out of 10 of them.

An alternative update to the definition of the Hayes English grammar also yields promising results. This involves changing the extrametricality parameter from Em-RtCons (the rightmost consonant is extrametrical) to Em-Rt (the rightmost syllable is extrametrical) and changing the degenerate foot parameter from DF-Strong (don’t allow degenerate feet) to DF-Weak (allow degenerate feet for single Light syllables at word edge). This allows the grammar to account for words like *prétty*, as the rightmost syllable is extrametrical (and therefore unstressed), while the single Light syllable *pre* can form a degenerate foot and be stressed. Making this update boosts the relative compatibility to 0.923 by types. Thus, this new English grammar with the Em-Rt and DF-Weak values now is also far more learnable compared to the other grammars in the Hayes hypothesis space, as it is also better at accounting for the English child-directed data than 9 out of 10 of them.

Turning to the OT KR, we find that there is a single ordering constraint update that

is very beneficial. All of the high-compatibility grammars rank NONFIN higher than WSP-vv, which means that it is more important to make the rightmost syllable extrametrical (NONFIN) than it is to stress long vowel syllables (WSP-vv). The current definition of the English grammar has the opposite ranking (WSP-vv > NONFIN), preferring to stress all long vowel syllables no matter where they are in the word. This makes the current English grammar unable to account for words like *báby*, which have an unstressed long vowel syllable as the rightmost syllable. Altering this ranking so that NONFIN is more important (NONFIN > WSP-vv) boosts the relative compatibility to 0.988. Thus, even moreso than the Hayes English grammar, updating the OT English grammar allows this grammar to be far more learnable than the other grammars in the OT hypothesis space, since it is better at accounting for the English child-directed data than nearly 99 out of 100 of them.

6.4 Summary

The basic issue we encountered is that all three KRs appear to have learnability issues when it comes to learning the English grammar they define from realistic English child-directed input. So, while these KRs satisfy the criterion of accounting for constrained cross-linguistic variation, they all seem to fail the learnability criterion when it comes to English. However, there are ways that they may be able to satisfy the learnability criterion after all.

First, it may be that children do not reach the target English grammar immediately, but instead pass through one or more transitory grammars. As they acquire useful knowledge about English metrical phonology, they may perceive the input differently and so update their non-target grammars to the target English grammar. We investigated the addition of one type of useful knowledge about the interaction of English metrical phonology with morphology that is likely to be acquired early enough to be used by children. However, this knowledge was not sufficient on its own, and other knowledge may be required for unbiased learners to learn the target English grammar in each KR. Experimental work may be able to determine what useful knowledge children acquire early enough to use when learning their metrical phonology grammar, as well as any transitory grammars they may converge on during acquisition.

A second option is that children are not unbiased learners, as our basic learnability analysis assumed, and they have useful learning biases that help them navigate the hypothesis space of grammars defined by each KR. Two potentially useful biases involve learning a grammar that accounts for a subset of the available input data (as defined by each bias), rather than all of it. The HV KR has been shown to benefit from exactly this type of bias, when the learner also has some prior knowledge about the order in which to learn parameters (Pearl 2008).

A third option is to update the definition of the target grammar for English to something that is more learnable from realistic input data. For the Hayes and OT KRs, there were potential minor updates that significantly improved learnability. Future experimental investigations can determine if adult knowledge corresponds better to these updated English grammars or if these grammars are perhaps intermediate knowledge states for children during acquisition. Future computational investigations on adult-directed English data may also help determine if the current target grammars are the most compatible with the data adults typically encounter. If not, this suggests that updating the definition of the target

English grammar is warranted – not only would the updated grammar be more learnable from child-directed speech, but it would be more compatible with adult knowledge.

For all three KRs, there are ways to potentially make the English grammars in them learnable. If it turns out that some are still not learnable, then this is support in favor of the ones that are in fact learnable. If it instead turns out that all are learnable, then it matters what each KR needs to satisfy learnability for English. If transitory knowledge states are assumed, we must find evidence that children pass through those transitory states. If prior knowledge is required, we must find evidence that children have that prior knowledge. If the adult knowledge is assumed to be different, we must find evidence that adult knowledge is indeed that way. Thus, computational investigations about learnability can lead to targeted experimental and further computational investigations that indicate which theoretical representations are accurate.

7 Conclusion

In this paper, we have established a methodology for quantitatively evaluating different knowledge representations (KRs), based on the learnability of their language-specific grammars from realistic acquisition input. We have used this approach to investigate KRs in metrical phonology, evaluating them on their ability to learn the target English grammar from realistic English data. English is an excellent test case for metrical phonology learnability, since it contains many irregularities and therefore represents a difficult acquisition scenario. So, if a KR allows a learner to successfully acquire the English grammar, that KR truly is useful for acquisition.

While we found that all three KRs have apparent learnability issues, we also were able to discover what causes the failure and what could be done about it. This led us to propose possible changes to the way acquisition must proceed for a given KR and possible changes to the definition of the target grammars for English within existing KRs. Thus, this computational approach allows us to suggest useful alterations to both the theories about how learning proceeds in this domain and the theories about how knowledge in this domain is represented.

8 References

- Joanne Arciuli, Padraic Monaghan, and Nada Seva. Learning to assign lexical stress during reading aloud: Corpus, behavioral, and computational investigations. *Journal of Memory and Language*, 63(2):180–196, 2010.
- C. L. Baker and John McCarthy. *The logical problem of language acquisition*. MIT Press, Cambridge, MA, 1981.
- Roger Brown. *A first language: The early stages*. Harvard University Press, Cambridge, MA, 1973.
- Kimberly Cassidy and Michael Kelly. Children’s use of phonology to infer grammatical class in vocabulary learning. *Psychonomic Bulletin & Review*, 8(3):519–523, 2001.

- Noam Chomsky. Rules and representations. *Behavioral and Brain Sciences*, 3:1–61, 1980.
- Noam Chomsky. *Lectures on Government and Binding*. Foris, Dordrecht, 1981.
- Noam Chomsky and Morris Halle. *The Sound Pattern of English*. Harper and Row, New York, NY, 1968.
- Alexander Clark and Shalom Lappin. Computational learning theory and language acquisition. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Philosophy of Linguistics*, pages 445–475. Elsevier, 2012.
- Stephen Crain. Language acquisition in the absence of experience. *Behavioral and Brain Sciences*, 14:597–612, 1991.
- Stephen Crain and Paul Pietroski. Why language acquisition is a snap. *The Linguistic Review*, 19:163–183, 2002.
- Brian Dillon, Ewan Dunbar, and William Idsardi. A single-stage approach to learning phonological categories: Insights from Inuktitut. *Cognitive Science*, 37:344–377, 2013.
- B. Elan Dresher. Charting the learning path: Cues to parameter setting. *Linguistic Inquiry*, 30(1):27–67, 1999.
- B. Elan Dresher and Jonathan Kaye. A computational learning model for metrical phonology. *Cognition*, 34(2):137–195, 1990.
- Naomi Feldman, Thomas Griffiths, Sharon Goldwater, and James Morgan. A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, 120(4):751–778, 2013.
- Janet D. Fodor. Unambiguous Triggers. *Linguistic Inquiry*, 29:1–36, 1998.
- Sean Fulop and Nick Chater. Learnability theory. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(3):299–306, 2013. doi: 10.1002/wcs.1228.
- Annie Gagliardi. All input isn’t equal: How the nature of the learner shapes language acquisition. *Studia Linguistica*, 67(1):68–81, 2013.
- Annie Gagliardi, Naomi Feldman, and Jeffrey Lidz. When suboptimal behavior is optimal and why: Modeling the acquisition of noun classes in Tsez. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, pages 360–365, Sapporo, Japan, 2012. Cognitive Science Society.
- LouAnn Gerken. Young children’s representation of prosodic phonology: Evidence from english-speakers’ weak syllable productions. *Journal of Memory and Language*, 33(1):19–38, 1994.
- LouAnn Gerken. Prosodic structure in young children’s language production. *Language*, 72(4):683–712, 1996.

- Sharon Goldwater, Thomas Griffiths, and Mark Johnson. A Bayesian Framework for Word Segmentation: Exploring the Effects of Context. *Cognition*, 112(1):21–54, 2009.
- Morris Halle and Jean-Roger Vergnaud. *An essay on stress*. Mit Press, Cambridge, MA, 1987.
- Michael Hammond. *The Phonology of English: A Prosodic Optimality-Theoretic Approach*. Oxford University Press, Oxford, UK, 1999.
- Bruce Hayes. Extrametricality and English stress. *Linguistic Inquiry*, 13:215–225, 1982.
- Bruce Hayes. *Metrical stress theory: Principles and case studies*. University of Chicago Press, Chicago, IL, 1995.
- Norbert Hornstein and David Lightfoot. Introduction. In Norbert Hornstein, editor, *Explanation in Linguistics: The Logical Problem of Language Acquisition*, pages 9–31. Longman, London, 1981.
- Linda Jarmulowicz, Valentina Taran, and Sarah Hay. Lexical frequency and third-graders’ stress accuracy in derived English word production. *Applied Psycholinguistics*, 29(2):213–235, 2008.
- Margaret Kehoe. Support for metrical stress theory in stress acquisition. *Clinical linguistics & phonetics*, 12(1):1–23, 1998.
- Michael Kelly. Rhythmic alternation and lexical stress differences in English. *Cognition*, 30: 107–137, 1988.
- Michael Kelly and Kathryn Bock. Stress in time. *Journal of Experimental Psychology*, 14: 389–403, 1988.
- Paul Kiparsky. Metrical structure assignment is cyclical. *Linguistic Inquiry*, 10(4):421–441, 1979.
- Julie Legate and Charles Yang. Assessing Child and Adult Grammar. In Robert Berwick and Massimo Piatelli-Palmarini, editors, *Rich Languages from Poor Inputs*, pages 168–182. Oxford University Press, Oxford, UK, 2013.
- Brian MacWhinney. *The CHILDES Project: Tools for Analyzing Talk*. Lawrence Erlbaum Associates, Mahwah, NJ, 2000.
- David Marr. *Vision*. W.H. Freeman, San Francisco, CA, 1982.
- Catherine McBride-Chang, Richard Wagner, Andrea Muse, Bonnie Chow, and Hua Shu. The role of morphological awareness in children’s vocabulary acquisition in English. *Applied Psycholinguistics*, 26(3):415–435, 2005.
- Joe Pater. Non-uniformity in English secondary stress: The role of ranked and lexically specific constraints. *Phonology*, 17(2):237–274, 2000.

- Lisa Pearl. *Necessary Bias in Natural Language Learning*. PhD thesis, University of Maryland, College Park, College Park, MD, 2007.
- Lisa Pearl. Putting the Emphasis on Unambiguous: The Feasibility of Data Filtering for Learning English Metrical Phonology. In Harvey Chan, Heather Jacob, and Enkeleida Kapia, editors, *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Child Language Development*, pages 390–401. Cascadilla Press, Somerville, MA, 2008.
- Lisa Pearl. Learning English Metrical Phonology: When Probability Distributions Are Not Enough. In Jean Crawford, Koichi Otaki, and Masahiko Takahashi, editors, *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition, North America (GALANA 2008)*, pages 200–211. Cascadilla Press, Somerville, MA, 2009.
- Lisa Pearl. When unbiased probabilistic learning is not enough: Acquiring a parametric system of metrical phonology. *Language Acquisition*, 18(2):87–120, 2011.
- Lisa Pearl and Benjamin Mis. How Far Can Indirect Evidence Take Us? Anaphoric One Revisited. In L. Carlson, C. Höschler, and T. Shipley, editors, *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 879–884, Austin, TX, 2011. Cognitive Science Society.
- Lisa Pearl, Sharon Goldwater, and Mark Steyvers. Online Learning Mechanisms for Bayesian Models of Word Segmentation. *Research on Language and Computation*, 8(2):107–132, 2011.
- Amy Perfors, Joshua Tenenbaum, and Terry Regier. The learnability of abstract syntactic principles. *Cognition*, 118:306–338, 2011.
- Michèle Pettinato and Jo Verhoeven. Production and perception of word stress in children and adolescents with Down syndrome. *Down Syndrome Research & Practice*, 13:48–61, 2008.
- Steven Pinker. Clarifying the logical problem of language acquisition. *Journal of Child Language*, 31:949–953, 2004.
- Alan Prince and Paul Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar*. ROA, New Brunswick, NJ, 2002.
- Elaine Stotko. Investigating children’s use of English derivational morphology. *First Language*, 14(42–43):348–349, 1994.
- Bruce Tesar and Paul Smolensky. *Learnability and optimality theory*. The MIT Press, Boston, MA, 2000.
- Andrea Tyler and William Nagy. The acquisition of english derivational morphology. *Journal of Memory and Language*, 28(6):649–667, 1989.
- Charles Yang. On productivity. *Yearbook of Language Variation*, 5:333–370, 2005.