Making an argument from acquisition: Testing theories of knowledge representation by how learnable they are from realistic input

Lisa Pearl, Timothy Ho, & Zephyr Detrano
*University of California, Irvine**

Abstract

One (often implicit) motivation for a theory of knowledge representation (KR) comes from an *argument from acquisition*, with the idea that language acquisition is straightforward if children's hypothesis space is defined by the correct KR. Acquisition is then the process of selecting the correct grammar from that hypothesis space, based on language input. To compare KR theories, we establish quantitative acquisition-based metrics that assess learnability from realistic input data. We conduct a learnability analysis for three KR theories proposed for metrical phonology and evaluate them on English, a language that is notoriously irregular with respect to metrical phonology and therefore non-trivial to learn. We find that all three KRs have similar learnability potential, but the proposed English grammars within each theory are not the grammars able to account for the most English child input data, even if the learner has some knowledge of the interaction between English metrical phonology and morphology. This suggests learnability issues exist for the proposed English grammar in all three theories if a learner is attempting to learn a grammar that accounts for as much input data as possible. We discuss ways a learner may still be able to learn the English grammar from English input by incorporating (i) additional useful linguistic knowledge about English metrical phonology interactions and (ii) biases to selectively learn from the input. We additionally discuss which aspects of the proposed English grammars are hurting learnability, observing that small changes in parameter values or constraint-rankings lead to significantly better learnability results.

1 Introduction

One way to describe a language's grammar is as a compact system in the human mind 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 should be able to 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

*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.

system). The variables that comprise 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. This means the child already knows which variables in the linguistic environment matter, and so can focus her attention on simply selecting the appropriate instantiation of the KR (i.e., the language-specific grammar), based on those relevant variables. So, 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*: 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 "maximally restrictive" and "capable of describing all the stress systems of the world's languages".

The second criterion is the *learnability criterion*: if children have access to 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 hypothesis space and limits children's attention to the relevant variables in the input. 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 not accomplished so easily.

Traditionally, KR theories have been proposed based on the cross-linguistic variation criterion (e.g., Halle and Vergnaud 1987; Hayes 1995; Hammond 1999; Pater 2000; Tesar and Smolensky 2000; Prince and Smolensky 2002). 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).

Because KR theorizing has been primarily driven by a KR's ability to account for constrained cross-linguistic variation, this has led to several KR options in different linguistic domains (e.g., metrical phonology: parameters whose values must be set (Halle and Vergnaud 1987; Hayes 1995), violable constraints that must be ordered with respect to importance (Tesar and Smolensky 2000; Prince and Smolensky 2002)). While these KR theories 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. To choose among these different KR theories, which are all 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 of a given KR from realistic language data? To answer this, we need to first 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 theory 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 theory 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 KR theories, 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 KR theories. We briefly review the KR theories, 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 theories 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 theory, 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 theory. Thus, even though all three theories have learnability issues for English at first glance, this does not mean we must immediately abandon those theories – instead, we can alter different aspects of the acquisition problem to see if and how each theory's English grammar can become learnable. In this way, we can identify KR theories 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 linguistic knowledge.

2 Learnability metrics

The essence of learnability is simply how easily children could learn a language's grammar when given realistic data from that language to learn from. When examining the learnability of a particular KR theory, we can assume children are already aware of the KR – and importantly, the variables the KR indicates are relevant for determining the language's grammar.

2.1 The learnability approach

Many different approaches to assessing learnability exist (e.g., Dresher and Kaye 1990; Dresher 1999; Pearl 2011; Clark and Lappin 2012; Legate and Yang 2013; Fulop and Chater 2013), and here we propose one that is similar to those taken by Pearl (2011) and Legate and Yang (2013). In particular, we will assess (i) learnability of realistic input and (ii) learnability 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 theory: 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. Why should the quantity of input data accounted for be important? 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 generated 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 a KR theory on their ability to account for those data. At an individual data point level, a grammar can either be compatible or incompatible 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 theory. 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, 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 theory.

If we want to focus on how easy it would be to learn a grammar with a specific raw compatibility, irrespective of how many grammars can achieve any particular raw compatibility, we might wish to calculate the *relative class compatibility* of the grammar. This is the proportion of raw compatibility scores that the current grammar's score is better than. For example, if there are 362,880 grammars in a hypothesis space, but only 445 distinct raw compatibility scores these grammars achieve, a grammar with a raw compatibility score higher than 350 of these would have a relative class compatibility of $350/445 = 0.787$. Notably, the grammars with the highest relative compatibility would also have the highest relative class compatibility (in the above example, grammars in the best raw compatibility class would have a relative class compatibility of $444/445= 0.998$).

It would of course be good if the best grammar also had a high raw compatibility, since this would mean the best grammar was able to compactly represent a large proportion of the available data. Put simply, it would be very useful for the learner to select this grammar. However, this is not required – the best grammar simply has to account for more data than any other grammar. No matter how few data points a grammar accounts for, if it accounts for more than any other grammar does, a rational learner will choose it as the best grammar to explain the language data in the learner's input. Thus, while raw compatibility is helpful to know from a grammar utility perspective, relative compatibility and relative class compatibility are more direct measures of learnability for a grammar.

We can also evaluate the *learnability potential* of a KR, which is simply the raw compatibility of the best grammar (with relative compatibility $\approx$1.00) defined by the KR. This indicates the utility of the KR, as instantiated by the best grammar it defines. 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

Language-specific grammars have often been derived with the goal of accounting for the language data adults know (e.g., Halle and Vergnaud 1987; Hayes 1995; Hammond 1999; Pater 2000). For example, a particular parameter value or constraint ordering may be based on the existence of a certain multisyllabic word in the adult lexicon. From this perspective, the main criterion for defining a language-specific grammar within a KR theory is that the language-specific grammar should account for the most adult data. Nonetheless, learnability is crucial: the language-specific grammar defined by the KR theory should be learnable from realistic child input data, since this is a main motivation for having a KR in the first place. 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 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 and relative class 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 *octopus* (/ɑktəpʊs/) having the stress contour 100.[1] All the KR theories 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.

(1) Sample metrical structure for *octopus* (/ɑktəpʊs/)

stress	1	0	0
metrical feet	(VC	V)	VC
syllable rimes	VC	V	VC
syllables	ɑk	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 KR theories we will compare, which include both parametric and constraint-ranking representations.

3.1 Parametric theories of knowledge representation

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, 2009, 2011). The HV representation involves five main parameters with three sub-parameters, yielding 156 grammars in the hypothesis space.[2] For a more detailed description of each of the parameters and their interactions with each other, see 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

[1]Here 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.

[2]Note that this is less than the full combinatoric possibilities of 180, as some parameter value combinations are incompatible, such as B-Mor (which requires syllables to be differentiated by weight) with QI (which does not differentiate syllables by weight).

closed).[3] 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 analyses of *beautiful* (/bjutəfʊl/)

QS	H	L	L/H	**QI**	S	S	S
syllable rime	VV	V	VC	syllable rime	VV	V	VC
syllables	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* (/dʒəɹæf/) and *octopus* (ɑktəpʊs), while (3b) shows Em-None applied to *afternoon* (æftəɹnun).

(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	ɹæf	ɑk	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	təɹ	nun

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) **Unb**ounded metrical foot construction

a. Em-None, Ft-Dir-Left for L L L H L

begin	(L	L	L	H	L
H syllable encountered	(L	L	L)	(H	L
end	(L	L	L)	(H	L)

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

b. Em-None, Ft-Dir-Rt for L L L H L

begin	L	L	L	H	L)
H syllable encountered	L	L	L	H)	(L)
end	(L	L	L	H)	(L)

c. 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 **B**ounded 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	$\mu\mu$	μ μ	μ μ
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**: (Ĺ L) (Ĺ)

b. **Ft-Hd-Rt**: (L Ĺ) (Ĺ)

The HV English 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* (/ɑktəpʊs/):
QS, QS-VC-H, Em-Rt, Ft-Dir-Rt, B, B-2, B-Syl, Ft-Hd-Left

stress	1	0	0
analysis	(H́	L)	⟨H⟩
syllable IPA	ɑk	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)eavy 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**).

(9) VC-H and VC-L analyses of *paper* (/pepəɹ/):

	VC-H		**VC-L**	
syllable class	H	H	H	L
syllable rime	VV	VC	VV	VC
syllables	pe	pəɹ	pe	pəɹ

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 (10). 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 (VVC+).

(10) 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
syllables	pe	pəɹ	pe	pəɹ

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 (11a). Weak local parsing (**LP-Weak**) requires that one Light syllable be skipped between feet (11b). Note that Heavy syllables are never skipped, even with weak local parsing.

(11) Sample parsing locality feet 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 (Ś S), or (ii) a single stressed Heavy syllable at the end of metrical foot construction (H́). A Tro-Mor foot can also take two forms, based on the idea that each foot has two moras (L = μ, H = μ μ): (i) two Light syllables with stress on the leftmost syllable (Ĺ L), or (ii) a single stressed Heavy syllable (H́). 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 Ś), or (ii) a single stressed Heavy syllable (H́). Example (12) demonstrates foot construction for a word of form H L L H H with each of the different foot types.

(12) 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-Syl foot 1	(H́	L)	L	H	H
Tro-Syl foot 2	(H́	L)	(Ĺ	H)	H
Tro-Syl foot 3	(H́	L)	(Ĺ	H)	(H́)

Tro-Mor

Tro-Mor foot 1	(H́)	L	L	H	H
Tro-Mor foot 2	(H́)	(Ĺ	L)	H	H
Tro-Mor foot 3	(H́)	(Ĺ	L)	(H́)	H
Tro-Mor foot 4	(H́)	(Ĺ	L)	(H́)	(H́)

Iamb

Iamb foot 1	(H́)	L	L	H	H
Iamb foot 2	(H́)	(L	Ĺ)	H	H
Iamb foot 3	(H́)	(L	Ĺ)	(H́)	H
Iamb foot 4	(H́)	(L	Ĺ)	(H́)	(H́)

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 (13, 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 (13, righthand side).

(13) 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.	(Ĺ)	⟨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 (14a). 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 (14b).

(14) 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	(Ĺ	L)	H	L	L
Tro-Mor foot 2	(Ĺ	L)	(H́	L)	L
Degenerate foot	(Ĺ	L)	(H́	L)	(Ĺ)
Degenerate foot stress survives	(Ĺ	L)	(H́	L)	(Ĺ)

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

Tro-Mor foot 1	L	L	H	(Ĺ	L)
Tro-Mor foot 2	L	(Ĺ	H)	(Ĺ	L)
Degenerate foot	(Ĺ)	(Ĺ	H)	(Ĺ	L)
Degenerate foot stress does not survive	(L)	(Ĺ	H)	(Ĺ	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 (15) how stress analysis direction can interact with the WLER.

(15) 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 Ḿ)	WLER-L stresses leftmost syllable.	Ĺ H
No degenerate feet.	(L H́)	Cannot create (L H́) Iamb foot due to Ĺ, so create (H́) Iamb foot	Ĺ (H́)
Word-level stress remains as is.			
		DF-Weak creates degenerate foot.	(Ĺ) (H́)

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 (16). Note that the English grammar generates the incorrect stress contour for this word (110 instead of the observed 100).

(16) English grammar analysis for *octopus* (/ɑktəpʊs/):
VC-H, Em-RtCons, FtDir-Rt, LP-Strong, Tro-Mor, DF-Strong, WLER-R, Bot-Up

stress	1	1	0
analysis	(H́)	(Ĺ	L)
syllables	ɑk	tə	pʊ⟨s⟩

3.2 Constraint-based theories of knowledge representation

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 $n!$ possible 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. So, only candidate analyses that have at least one metrical foot are considered by the learner.

3.2.1 Constraints

Non-Finality (NonFin). The final syllable is unfooted. In (17), the first candidate form for *little* (/lɪɾɪl/) is preferred since the final syllable is not included in a metrical foot.

(17)

Input: /lɪ ɾɪl/	NonFin
a. ☞ (lɪ́) ɾɪl	
b. lɪ (ɾɪ́l)	∗!

Trochaic (Tro). Feet are headed on the left. In (18), the first candidate for *mommy* (/mami/) is preferred since its sole foot has stress on the leftmost syllable.

(18)

Input: /mɑ mi/	Tro
a. ☞ (mɑ́ mi)	
b. (mɑ mí)	∗!

Weight-to-Stress Principle VV (WSP-VV). Syllables with long vowels should be stressed. The first candidate in (19) for *canoe* (/kənu/) is preferred since its second syllable has a VV rime and is stressed.

(19)

Input: /kə nu/	Wsp-vv
a. ☞ (kə nú)	
b. (kʌ́ nu)	∗!

Weight-to-Stress Principle VC (WSP-VC). Syllables closed by consonants should be stressed. The first candidate in (20) for *little* (/lɪɾɪl/) is preferred since its second syllable has a VC rime and is stressed.

(20)

Input: /lɪ ɾɪl/	Wsp-vc
a. ☞ (lɪ ɾɪ́l)	
b. (lɪ́ ɾɪl)	∗!

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 (21) is preferred since the sole metrical foot contains two syllables.

(21)

Input: /lɪ ɾɪl/	FtBin
a. ☞ (lɪ́ ɾɪl)	
b. (lɪ́) ɾɪl	∗!

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* (hərɑjzən) in (22) is preferred.

(22)

	Input: /hə rɑj zən/	ALIGN-R
a.	(hʌ́) raj zən	∗!∗
b.	hə (rá j) zən	∗!
c. ☞	hə raj (zʌ́n)	

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 (23) is preferred.

(23)

	Input: /hə rɑj zən/	ALIGN-L
a. ☞	(hʌ́) raj zən	
b.	hə (rá j) zən	∗!
c.	hə raj (zʌ́n)	∗!∗

Parse-Syllable (Parse-σ). Syllables must belong to feet. Extrametrical syllables violate this constraint and so the first candidate for *mommy* in (24) is preferred.

(24)

	Input: /mɑ mi/	PARSE-σ
a. ☞	(mɑ́ mi)	
b.	(mɑ́) mi	∗!

***Sonorant Nucleus (*SonNuc).** Syllables should avoid having sonorant nuclei. The first candidate for *little* in (25) is preferred since none of its syllables have sonorant nuclei.

(25)

	Input: /lɪ ɾɪl/	*SONNUC
a. ☞	(lɪ́ ɾɪl)	
b.	(lɪ́ ɾl̩)	∗!

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 ɾi/), and (ii) sonorant nucleus (R), as in the second syllable of *actor* (/æk tr̩/). 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* (/hr̩d/).[4] The long variant is a long vowel (VV), as in the second syllable of *kitty* (/kɪ ɾi/), and the super-long variant is a long vowel closed with a consonant (VVC), as in *boot* (/but/). See Table 2 for how these syllabic distinctions compare to the distinctions required by the other representations.

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.

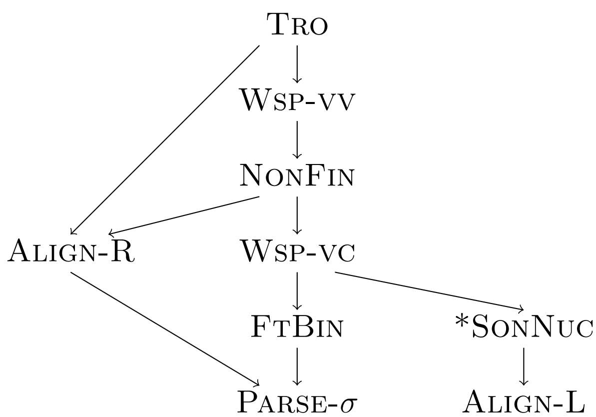

Figure 1: Partial ordering of constraints defining the English grammar.

The tableau below is an evaluation of a grammar satisfying the English constraint rankings on *kitty* (/kɪɾi/), given the five potential candidate analyses that are generated. The optimal candidate has an incorrect stress pattern, primarily because Wsp-vv is highly ranked.

Input: /kɪ ɾi/	Tro	Wsp-vv	NonFin	Wsp-vc	Align-R	FtBin	Parse-σ	*SonNuc	Align-L
(kɪ́ ɾi)		*!	*		*				
(kɪ ɾí)	*!		*						*
(kɪ́) ɾi		*!			*	*	*		
☞ kɪ (ɾí)			*			*	*		*
(kɪ́)(ɾí)			*			**!			

[4]We note that the distinction between syllables with sonorant nuclei (R) and the more general closed syllable (VC) is made because of the interaction between the constraints *SonNuc and Wsp-vc. For example, if *SonNuc is highly ranked, an R syllable will be perceived as a VC syllable, and then Wsp-vc will apply; in contrast, if *SonNuc is not highly ranked, an R syllable will be perceived as a short syllable, and Wsp-vc will not apply.

The tableau below demonstrates the same grammar evaluating the candidates for *little* (/lɪɾɪl/). Because the final /l/ could be the nucleus of the second syllable, eight candidates are generated. For this word form, however, the optimal candidate for the grammar has a stress contour that matches the observed stress contour of *little* (*líttle*).

Input: /lɪ ɾɪl/	Tro	Wsp-vv	NonFin	Wsp-vc	Align-R	FtBin	Parse-σ	*SonNuc	Align-L
(lɪ́ ɾɪl)			*!	*	*				
(lɪ ɾɪ́l)	*!		*						*
(lɪ́) ɾɪl				*!	*	*	*		
lɪ (ɾɪ́l)			*!			*	*		*
(lɪ́)(ɾɪ́l)			*!			**			
(lɪ́ ɾl̩)			*!		*			*	
(lɪ ɾĺ̩)	*!		*					*	*
☞ (lɪ́) ɾl̩					*	*	*	*	
lɪ (ɾĺ̩)			*!			*	*	*	*
(lɪ́)(ɾĺ̩)			*!			**		*	

3.3 Knowledge representation comparison

While these KR theories are not simply notational variants of each other, they do overlap on the linguistic aspects that they consider relevant. Table 1 summarizes the points of overlap, as well as the variables that are unique to each representation. We note that even for the aspects where there is overlap, the instantiation is rarely identical across representations. Thus, these KRs are able to account for the observed constrained cross-linguistic variation by drawing on sets of linguistic variables that are significantly different.

Parametric: HV	**Parametric: Hayes**	**Constraint-ranking: OT**
Extrametricality	Extrametricality	Non-Finality
	Parsing Locality	Parse-σ
Boundedness	Foot Inventory	Foot Binarity
Foot Headedness		Trochaic
Quantity Sensitivity		Weight-to-Stress (VV)
Quantity Sensitivity (VC)	Syllable Weight (VC)	Weight-to-Stress (VC)
Foot Directionality	Foot Directionality	
	Word Layer End Rule	Align Left
		Align Right
	Degenerate Feet	
	Stress Analysis Direction	
		*Sonorant Nucleus

Table 1: A comparison of the three KR theories for metrical phonology, aligning parameters/constraints that refer to similar aspects of metrical structure.

Due to the linguistic variables defined by each KR, there are different syllabic distinctions each requires the learner to make (HV: 3, Hayes: 4, OT: 8). Table 2 highlights the similarities and differences in these syllabic distinctions.

Syllable Type	**Example**	**Parametric: HV**	**Parametric: Hayes**	**Constraint-ranking: OT**
V	**ki** tty	short	short	short
R	ac **tor**	closed	potentially closed	sonorant
VR	**ten**	closed	potentially closed	closed-VR
VC	**took**	closed	potentially closed	closed-VC
VCC+	**best**	closed	always closed	closed-VC
RC	**heard**	closed	always closed	closed-RC
VRC	**tent**	closed	always closed	closed-VRC
VV	ki **tty**	long	long	long
VVC+	**boot**	long	long	super-long

Table 2: Syllabic distinctions assumed by each KR (HV: 3, Hayes: 4, OT: 8), indicating how each syllable type is classified by each KR. The relevant syllable in the examples of each syllable type is **bolded**. The syllable type is indicated by using the following abbreviations: V = vowel, R = sonorant consonant, C = non-sonorant consonant, + = 1 or more of the symbol indicated.

A striking difference among the English grammars defined by these KR theories is their differing ability to account for the observable stress contours of English words. Table 3 presents some sample words highlighting these differences. While there are common English words (e.g., *little*) that all three representations' English grammars can account for, there are also common words that each one cannot account for (e.g., HV: *today*; Hayes: *kitty, finished*; OT: *kitty*, sometimes *today*).

Word	**Stress**	**HV**	**Hayes**	**OT**
little	10	✓	✓	✓
kitty	10	✓	No	No
today	01	No	✓	21/26
finished	10	✓	No	✓

Table 3: Analysis of sample English words by the English grammars in the different KR theories. Stress contour is indicated, where 1= a stressed syllable and 0 = an unstressed syllable. English grammars that are capable of accounting for the observed stress contour are indicated with a ✓, or a proportion in the case of OT, which has multiple grammars satisfying the partial ordering of constraints corresponding to English.

4 English metrical phonology

English metrical phonology is an excellent test case for metrical phonology KR theories, since the data are notoriously irregular and therefore 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* (/kjukʌmbəɹ/) is compatible with, shown in (26). Quite strikingly, these two grammars have no parameter values whatsoever in common, yet are able to generate the same stress contour (contour: 110).

(26) Two grammars *cucumber* is compatible with

a. QI, Em-None, Ft-Dir-Rt, B, B-2, B-Syl, Ft-Hd-Left

Analysis	(Ś)	(Ś	S)
Syllables	kju	kʌm	bəɹ

b. QS, QS-VC-H, Em-Some, Em-Right, Ft-Dir-Left, Unb, Ft-Hd-Rt

Analysis	(H́)	(H́)	⟨H⟩
Syllables	kju	kʌm	bəɹ

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 (26) 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 KR theories 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., a word form 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 previously, 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 multisyllabic words, under the assumption that at least two syllables are required to have a stress contour, we find that this issue occurs quite often (see Table 4). Between 37% and 58% of the syllabic word forms (depending on the syllabic distinctions made by a KR theory) have multiple stress contours associated with them. 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 KR theories to account for all the input data, due to these numerous irregularities.

	Total syllabic word forms	**Syllabic word forms with multiple stress contours**
HV	186	95 (51%)
Hayes	149	86 (58%)
OT	452	166 (37%)

Table 4: Syllabic word forms with multiple stress contours for each KR theory.

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 three). 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 KR theories 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 5. 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 theory 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 raw compatibility with the English input data, since grammars that account for more data are more useful to learn. Table 5 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 (by types: 0.485-0.593, by 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 KRs. As Table 5 shows, the parametric English grammars are better than about $\frac{2}{3}$ of the grammars in the hypothesis space (by types: 0.673-0.676, by 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 (by types: 0.817, by 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 (by types: HV=51, Hayes=249; by tokens: HV=51, Hayes=242) while for the constraint-based KR, tens of thousands of grammars are better (by types: OT=66,407, by tokens: OT=78,019). Even if we focus on relative class compatibilities, and simply care about how easy it is to learn a grammar with the compatibility score the (best) English grammar has, the (best) English grammar lags behind (parametric: by types = 0.684-0.697, by tokens=0.667-0.696; constraint-based: by types=0.787, by tokens=0.523).

5.3 Learnability summary

For all three KR theories, there are learnability issues. Using the English input children are likely to encounter, the English grammar defined in each KR is unlikely to be easily learnable from the hypothesis space of grammars defined by the 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 5: Learnability analyses for the three KR theories: HV, Hayes, and OT. The four metrics shown are learnability potential of the KR (KR:Pot), raw compatibility of the (best) English grammar (Eng:Raw), relative compatibility of the (best) English grammar (Eng:Rel), and relative class compatibility (Eng:RelClass) of the (best) English grammar, 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**	**Eng:RelClass**
HV	0.668 (0.739)	0.593 (0.716)	0.673 (0.673)	0.697 (0.667)
Hayes	0.683 (0.750)	0.485 (0.531)	0.676 (0.685)	0.684 (0.696)
OT	0.657 (0.729)	0.573 (0.574)	0.817 (0.785)	0.787 (0.523)

6 Addressing the learnability issues

The learnability problem can effectively be summarized as the learner not being able to reach 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 KR theories could satisfy learnability for English.

6.1 Intermediate knowledge states

Experimental data suggest that there may be several intermediate knowledge states that children pass through when learning English metrical phonology. At age two, English children use a metrical template that operates over syllables (Echols 1993) and which has the leftmost syllable stressed (Gerken 1994, 1996), which Gerken interprets as a syllable-based trochaic template. By age three, children have recognized that the metrical system is quantity sensitive, but not that the rightmost syllable is typically extrametrical (Kehoe 1998). By age four or five, English children seemed to have identified the target English grammar (e.g., Pettinato and Verhoeven 2008; Arciuli et al. 2010).

If we interpret these findings using the KR theories under consideration, it seems the trochaic metrical template used at age two could implement a preference for a quantity insensitive metrical foot spanning two syllables, with stress on the leftmost syllable (HV: QI, B, B-2, B-Syl, Ft-Hd-Left; Hayes: Tro-Syl; OT: Trochaic, FtBin>Wsp-vv,Wsp-vc). By age three, quantity sensitivity is realized, but not extrametricality (HV: QS, Em-None; Hayes: Em-None, LP-Strong; OT: Wsp-VV, Parse-σ highly ranked). It therefore seems possible that there are additional transitory states before the final knowledge state is achieved at age four or five.

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 (Hayes 1995). For example, in *sensátionally*, the derivational affixes *-al* and *-ly* are not stressed, and in *préttiest*, 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, viewing the input in a different way than they had before. In particular, they could ignore inflectional morphology when attempting to determine the analysis underlying an observable stress contour. Thus, *préttiest* (/pɹíɾiɪst/) would be viewed as *prétti* (/pɹíɾi/) 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 (shown in Table 6).[5] This simulates the learner ignoring inflectional morphology in the input when learning the English metrical phonology grammar. Results of this analysis are shown in Table 7.

Table 6: Inflectional morphology ignored once the knowledge that inflectional affixes do not impact English metrical phonology is acquired. Examples come from the Brent corpus of American English child-directed speech.

orthography	**pronunciation(s)**	**examples**
–s	/s/	*minutes*
	/z/	*fingers*
–es	/ɪz/	*glasses*
–ses	/sɪz/	*folkses*
	/zɪz/	*appleses*
–'s	/s/	*cupcake's*
	/z/	*judy's*
–ed	/t/	*finished*
	/d/	*surprised*
	/ɪd/	*decided*
–ing	/ɪŋ/	*listening*
–en	/n̩/	*gotten*
–er	/ɹ̩/	*longer*
–est	/ɪst/	*sweetest*

[5]Note 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.

Table 7: Learnability analyses for the three KRs (HV, Hayes, and OT) once knowledge has been acquired that inflectional morphology does not typically affect the stress contour. The four metrics shown are learnability potential of the knowledge representation (KR:Pot), raw compatibility of the (best) English grammar (Eng:Raw), relative compatibility of the (best) English grammar (Eng:Rel), and relative class compatibility (Eng:RelClass) of the (best) English grammar, 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**	**Eng:RelClass**
HV	0.662 (0.738)	0.605 (0.719)	0.712 (0.673)	0.706 (0.667)
Hayes	0.683 (0.750)	0.550 (0.552)	0.704 (0.685)	0.713 (0.704)
OT	0.677 (0.749)	0.578 (0.575)	0.786 (0.777)	0.627 (0.535)

The learnability potential of the three KRs remains about the same: around $\frac{2}{3}$ of the word types (0.662-0.683) and around $\frac{3}{4}$ of the word tokens (0.738-0.750) can be accounted for by the best grammar defined by each KR. This indicates that the best grammar in each KR is still very useful to learn, since it can account for a large portion of the input. Notably, the acquired knowledge about inflectional morphology does not help the best grammar account for much more data than it could before, no matter what the KR.

However, perhaps the utility of this acquired knowledge is more targeted at improving the coverage of the (best) English grammar within each KR. If this is true, the raw compatibility of the (best) English grammar should be much closer to that of the best grammar. Unfortunately, this is not so: the English grammar still lags behind the best grammar in each KR (by types: English=0.550-0.605 vs. best= 0.662-0.683, by tokens: English=0.552-0.719 vs. best=0.738-0.750).[6] Thus, even with this acquired knowledge about inflectional morphology, the (best) English grammar is still not the best grammar overall, and so a rational learner would not select it based on realistic English input.

But again, what matters more than raw compatibility is relative compatibility: 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? As Table 7 shows, the parametric English grammars are still better than about $\frac{2}{3}$ of the grammars in the hypothesis space (by types: 0.704-0.712, by 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 (by types: 0.786, by tokens: 0.777). This again indicates that the English grammars are better than many other grammars – but there are still a large number of grammars that are better than the English grammars. If we examine relative class compatibility, asking how easy it would be to learn a grammar with the raw compatibility that the (best) English grammar has, we again encounter the same problem – no matter which KR, the (best) English grammar's compatibility with the English data lags behind (by types: 0.627-0.713, by tokens: 0.535-0.704). Once again, the target English grammar is unlikely to be easily learnable from this hypothesis space of grammars, even with this acquired knowledge about some interactions

[6]See Appendix A for more details of the impact of this knowledge on the compatibility of the English grammars defined by the KR theories.

between metrical phonology and morphology.

So, the same learnability issues persist. One pervasive issue that occurs is that the English grammars in all three KR theories typically want to stress syllables with a long vowel nucleus (e.g., *sweet*).[7] This can be problematic for realistic child-directed speech since many words (and often very frequent words) have unstressed long vowel syllables (see Table 8).

Table 8: Example word types from the Brent corpus that have unstressed long vowel syllables, which the English grammars defined by the KR theories typically have difficulty accounting for. The stress contour for each word is indicated. Unstressed long vowel syllables within words are *italicized*, and the frequency of each word type in the corpus is also shown.

Diminutives	**Compounds**	**Proper names**	**Others**
bá *by* (2158)	pá *tty* + cáke: (190)	mán *dy*: (341)	rá *di* ó (23)
kí *tty* (1261)	bé *lly* + bú tton: (15)	él *mo*: (190)	spa ghé *tti os* (12)
swée *tie* (737)	bá *by* + sí tter: (8)		ób é *di* ent (8)
dá *ddy* (561)	úp *side* + dówn: (6)		
dó *ggie* (376)	slée *py* + héad: (3)		
só *ckie* (10)	hánd + *me* + dówns: (1)		

One way to deal with these problematic data is to acquire additional knowledge that allows the learner to view them differently. For example, perhaps the learner could perceive the diminutive affix /i/ as a kind of inflectional morphology (Legate and Yang 2013): 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 English metrical phonology grammar. 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 KR theories.

Another type of useful knowledge involves recognizing that some words are compound words, and so are comprised of words that may individually obey the English grammar even if the compound word violates the grammar. In effect, the knowledge that a word is a compound word would cause the child to analyze the individual words comprising it separately, rather than analyzing the compound word as an atomic unit. This would be particularly useful for the HV English grammar, for example, which allows the rightmost syllable to be extrametrical. So, an unstressed long vowel syllable at the right edge of a word can be accounted for under this English grammar. This then allows the HV English grammar to account for all the compound words listed in Table 8, since all the individual words comprising those compound words are compatible with that grammar.

A third type of useful knowledge involves recognizing that there are special classes of words that may have different stress patterns from the general grammar of English. Syllable

[7] The HV English grammar allows some exceptions to this, since the rightmost syllable can be extrametrical and so stressless no matter what kind of syllable it is.

word forms with multiple stress contours can signal this, such as bisyllabic nouns and verbs (e.g., *cónduct* vs. *condúct*). The grammatical category of the word is the predictable cue to the stress contour. Once children realize that there are predictable exceptions (which the grammatical category stress contours demonstrate), they may allow other exceptional classes of words for stress. If, for example, proper names then typically have predictable stress contours that violate the English grammar (such as *mandy* and *elmo*), children may ignore these proper name data when learning the English grammar, viewing them as exceptions that will need to be learned separately. It is currently unclear when children recognize that grammatical category interacts with metrical phonology in English, and so it is unclear when they would be able to use this as a signal that exceptional classes exist for English stress. Nonetheless, this general strategy of ignoring certain data in the input 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.

6.2 Selective learning

It could be instead that the learning process is more sophisticated, with the learner having useful prior knowledge that guides learning. Thus, the initial knowledge state would also include helpful learning biases for navigating the hypothesis space defined by the KR theory. Both learning biases we discuss below are predicated on the same basic idea: instead of trying to account for all the input data with a grammar, the learner only tries to account for a subset of the input data that is 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 was 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, quite reasonably, 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. In essence, 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.[8,9] 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 a larger proportion of the relevant 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. But what should it be within each KR? 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 can examine the set of grammars that have the highest raw compatibility (and so also have the highest relative compatibility) to determine how they differ from the current English definition in each KR. In particular, we can calculate the average *proximity score* to quantify (i) how alike the top grammars are to each other and (ii) how different these grammars are from the English grammar(s). The proximity score determines (on average) how many linguistic variables (i.e., parameter values or constraint orderings) these grammars agree on, with 1 being perfect agreement and 0 being perfect disagreement. See Appendix C for details of this calculation.

If we find that the high compatibility grammars have a high average proximity score when compared against each other, this indicates that there are certain linguistic variable values (which these grammars share) that best capture the English data. If the English grammar has a lower average proximity score when compared to this set of high compatibility grammars, perhaps the values that differ between the English grammar and the high compatibility

[8]See Appendix B for a more detailed example of using the Tolerance Principle to implement this bias to learn only from regular data.

[9]Also, it may be useful to apply this kind of filter to data when defining the target grammar for English in the first place, since the same logic of capturing only the patterns of the regular data applies to adult grammars just as much as it does to child grammars.

grammars should be changed.

For each KR, we calculate the average proximity score over word type compatibility for the grammars in the top compatibility class (or classes, in the case of the HV KR, which has only a single grammar in the top compatibility class). We also calculate the average proximity score between the grammars in the top compatibility class(es) and the English grammar (or grammars, in the case of the OT KR). We do this both for a learner who perceives the input with the knowledge that inflectional morphology does not affect stress contour generation as well as for a learner who perceives the input as is. This is because we may be interested in two potential knowledge states: (i) a different target grammar for English when the inflectional morphology interaction would already be known, and (ii) a transitory knowledge state a learner could pass through when this interaction is not yet known.

As Table 9 shows, the top grammars in each KR tend to agree with each other more than the English grammars defined by each KR agree with those top grammars.[10] Interestingly, this is true whether the learner has knowledge of inflectional morphology or not (and in fact, for the HV and Hayes KRs, this is because the same grammars are the top grammars in either case). This suggests that there are useful updates that could be made to each KR theory to make the English grammar(s) more compatible with the English data, irrespective of whether the learner has this inflectional morphology knowledge.

Table 9: Proximity score calculations for each KR over word type compatibility, including the number of grammars in the top compatibility class by types (# Top), the average proximity score of that top class(es) of grammars (Prox: Top), and the average proximity score between the top compatibility class(es) and the English grammar(s) (Prox: Top vs. Eng). Proximity scores close to 1 indicate grammars that agree on many linguistic variables, while proximity scores close to 0 indicate grammars that disagree on many linguistic variables. Scores are calculated both for a learner with knowledge of the interaction of inflectional morphology (+infl-knowledge) and for a learner without that knowledge (–infl-knowledge).

	KR	**# Top**	**Prox: Top**	**Prox: Top vs. Eng**
+infl-knowledge	**HV**	16	0.528	0.475
	Hayes	10	0.753	0.363
	OT	48	0.893	0.650
–infl-knowledge	**HV**	16	0.528	0.475
	Hayes	10	0.753	0.363
	OT	648	0.798	0.664

For the HV KR, it turns out that the grammars in the top compatibility classes use a different quantity sensitivity value than the current definition of the English grammar: they

[10]We note that proximity score comparisons are meaningful within a KR, but not across KRs, because the proximity score is affected by the number of linguistic variables in a KR, and this number differs across KRs. This means, for example, that a lower proximity score for the top grammars in the HV KR compared to the top grammars in the Hayes KR does not necessarily indicate that the HV top grammars agree with each other less than the Hayes top grammars do. See Appendix C for details.

use QI or QS-VC-L instead of QS-VC-H. This would allow the grammar to account for words that have unstressed VV or VC syllables, like *béllybútton* and *sátisfied*. Making either of these changes to the HV English definition yields similar effects, whether the learner has knowledge of the inflectional morphology interaction or not. Changing only the quantity sensitivity parameter value boosts the raw compatibility of the English grammar from accounting for about 60% of the data by types (+infl=0.61, –infl=0.59) to accounting for approximately 62–63% (QI: +infl=0.63, –infl=0.64; QS-VC-L: +infl=0.62, –infl=0.62). From the learnability standpoint, these changes boost the relative compatibility so the English grammar is better than approximately 85 (QS-VC-L) to 95 (QI) out of 100 of the other grammars in the hypothesis space (QI, relative comp: +infl=0.96, –infl=0.94; relative class comp: +infl=0.95, –infl=0.95; QS-VC-L, relative comp: +infl=0.86, –infl=0.87; relative class comp: +infl=0.87, –infl=0.86).

Turning to the Hayes KR, it turns out that many of the grammars in the top compatibility class use a different metrical foot value than the current definition of the English grammar: they use syllabic trochees (Tro-Syl) rather than moraic trochees (Tro-Mor). If we alter the English grammar to use the Tro-Syl parameter value, it could account for bisyllabic words with an unstressed heavy syllable at the end, such as *báby* and *kítty*, as well as trisyllabic compound words with unstressed syllables in the middle and heavy syllables as the edge, such as *sléepyhéad*. As with the HV English grammar, making this change to the Hayes English definition yields similar effects, whether the learner has knowledge of the inflectional morphology interaction or not. Changing only the foot inventory parameter value boosts the raw compatibility of the English grammar from accounting for around half of the data by types (+infl=0.55, –infl=0.49) to accounting for nearly two thirds of the data types (+infl=0.64, –infl=0.64). From the learnability standpoint, this change boosts the relative compatibility so the English grammar is better than approximately 9 out of 10 of the other grammars in the hypothesis space (relative comp: +infl=0.87, –infl=0.91; relative class comp: +infl=0.90, –infl=0.90).

Turning to the OT KR, we find that there is a single ordering constraint update that all the top compatibility grammars use, but which the current English grammar definition does not use: ranking NonFin higher than Wsp-vv. This 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. As with the HV and Hayes English grammars, making this change to the OT English definition yields similar effects, whether the learner has knowledge of the inflectional morphology interaction or not. Flipping only this ranking boosts the raw compatibility of the English grammar from accounting for around 58% the data by types (+infl=0.58, –infl=0.57) to accounting for nearly two thirds of the data types (+infl=0.66, –infl=0.66). From the learnability standpoint, this change boosts the relative compatibility so the English grammar is better than at least 9 out of 10 of the other grammars in the hypothesis space, and often better than nearly all other grammars (relative comp: +infl=0.90, –infl=0.99; relative class comp: +infl>0.99, –infl=0.99).

To sum up, there are options for updating the definition of the English grammar for all three KR theories that significantly aid learnability. This motivates us to examine whether these updated English grammars can similarly account for adult metrical phonology knowledge better than the previous English grammar definitions. If it turns out that they can, then this is support for these being the true target states for English metrical phonology knowledge. If instead it turns out that these updated grammars are not as good at accounting for adult knowledge, they could instead represent transitory knowledge states that children pass through, as discussed in section 6.1. It would then be useful to determine if children converge on these updated grammars at some point during acquisition, before moving on to the true target grammar for English.

6.4 Summary

The basic issue is that all three KR theories appear to have learnability issues when it comes to learning the English grammar they define from realistic English child-directed input. So, while these KR theories 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 is required for unbiased learners to learn the target English grammar in each KR. Experimental work may be able to determine what other 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 theory. Two potentially useful biases involve learning a grammar that accounts for a subset of the available input data, 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 all three KR theories, there were minor updates, altering a single parameter value or constraint ranking, 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 KR theories, 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

We have established a methodology for quantitatively evaluating different theories of knowledge representation (KR), based on the learnability of their language-specific grammars from realistic acquisition input. This computational analysis represents the first step for making an argument from acquisition for any KR theory. If and when we find that it is possible for a KR to satisfy the learnability criterion proposed here, we can then proceed to the next step: Is it possible for children – with all their cognitive limitations – to learn the language-specific grammar defined by the KR from realistic language input? That is, if the language-specific grammar is learnable in principle, is it also learnable in practice? If so, we then have a strong argument from acquisition for that KR theory.

Here, we have used this approach to investigate KR theories in metrical phonology, evaluating them on their ability to make the target English grammar easily learnable 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 KR theories 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 learner using a given KR and possible changes to the definition of the target grammars for English within existing KR theories. 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.

Catharine Echols. A perceptually-based model of children's earliest productions. *Cognition*, 46(3):245–296, 1993.

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.

A Impact of morphological knowledge

Allowing the learner to be aware of some of the interactions between morphology and English metrical phonology has different effects on the KRs. Before this knowledge is available, the parametric KRs are able to account for different subsets of the ten most frequent stressed syllabic word forms, as shown in Table 10.

Syl word form HV/Hayes	**Stress**	**# types**	**Examples**	**HV**	**Hayes**
LC/LP	10	592	*water, going, doing*	✓	✓
XC/XP	10	472	*little, getting, coming*	✓	✓
LL	10	334	*baby, sweetie, mommy*	✓	
XL	10	309	*kitty, daddy, very*	✓	
CC/AP	10	235	*goodness, handsome, helper*	✓	✓
LL	11	188	*okay, bye-bye, tv*		✓
CL/AL	10	172	*window, birdie, only*	✓	
LC/LA	10	171	*peanuts, secrets, highest*	✓	
XC/XA	10	170	*biggest, buckets, hiccups*	✓	
XL	01	145	*below, today, hurray*		✓

Table 10: Ten most frequent stressed word forms (by type) in the parametric KR input. The syllabic word forms are abbreviated with X (short syllable: V), C (closed syllable: VC+ [HV]), P (potentially closed syllable: VC [Hayes]), A (always closed syllable: VCC+ [Hayes]), and L (long syllable: VV(C)). Both HV and Hayes representations of syllabic word forms are shown. Stress contour is indicated, where 0 = unstressed syllable and 1 = stressed syllable. The number of word types corresponding to the syllabic word form with the indicated stress contour is shown. A ✓ indicates that the KR can account for the stressed syllabic word form.

One striking difference is the ability of the HV English grammar to already account for many of the stressed syllabic word forms that have the most word types (8 out of 10), compared with the Hayes English grammar (5 out of 10). Notably, the frequent stressed syllabic word forms that the HV grammar cannot account for are unlikely to be helped by the knowledge that inflectional morphology can be ignored for the purposes of generating a stress contour – words like *okay* (LL: ‘11’) and *today* (XL: ’01’) do not involve inflectional morphology. Thus stripping off inflectional morphology does not help the HV English grammar account for these word types any more than it could before. In fact, of the 552 word types reduced to monosyllabic forms due to morphological knowledge, *all* of them were already compatible with the HV English grammar (e.g., *highest* (LC: '10') and *biggest* (XC:'10')). This is why there is little increase in overall compatibility for the HV English grammar (0.593 to 0.605 by types, and 0.716 to 0.719 by tokens). There are only a few new word forms that can be accounted for with morphological knowledge in the HV representation: 14 trisyllabic words like *sillier* (XLC: '100') and 17 trisyllabic words like *coloring* (XCC: '100'). Thus, the addition of morphological knowledge does not obviously aid the HV English grammar.

The Hayes English grammar, in contrast, is unable to account for five of the most frequent stressed syllabic word forms, some of which involve inflectional morphology, like *highest* (LA: '10') and *biggest* (XA: '10'). Ignoring inflectional morphology clearly is helpful, as the raw compatibility of the English grammar goes from 0.485 to 0.550 by types (though only 0.531 to 0.552 by tokens). This increase occurs because the Hayes English grammar is able to account for 332 more word types than before: 28 troublesome bisyllabic forms becoming monosyllabic (e.g., *cleanest* → *clean*), 100 troublesome LA forms becoming LP (e.g., *pockets* → *pocket*), 112 troublesome XA forms becoming XP (e.g., *apples* → *apple*), and 92 changes in less common syllabic word forms (e.g., *messages* → *message*).

When we turn to the OT English grammar, we find similar behavior to the HV English grammar.

Syl word form	**Stress**	**# types**	**Examples**	**OT**
XL	10	242	*kitty, ready, very*	
LM	10	198	*going, doing, trying*	✓
LL	10	198	*baby, sweetie, mommy*	
XR	10	196	*little, Dillon, other*	✓
XC	10	178	*hiccup, jacket, kisses*	✓
LR	10	174	*over, water, open*	✓
XM	10	130	*getting, looking, coming*	✓
XS	01	127	*about, around, supposed*	21/26
LC	10	123	*mooshas, peaches, mama's*	✓
XN	10	121	*didn't, doesn't, isn't*	✓

Table 11: Ten most frequent stressed word forms (by type) in the constraint-based KR input. The syllabic word forms are abbreviated with X (short syllable: V), R (sonorant nucleus: R), C (closed syllable with non-sonorant nucleus: VC), M (closed syllable with non-sonorant nucleus and sonorant consonant: VR), N (closed syllable with sonorant nucleus and non-sonorant consonant: RC), L (long syllable with no coda: VV), and S (super-long syllable: VVC). Stress contour is indicated, where 0 = unstressed syllable and 1 = stressed syllable. The number of word types corresponding to the syllabic word form with the indicated stress contour is shown. A ✓ indicates that all 26 OT English grammar instantiations can account for the stressed syllabic word form. A proportion indicates how many of the 26 English grammar instantiations can account for the stressed syllabic word form.

Adding knowledge of inflectional morphology doesn't seem to help raw compatibility much (from 0.573 to 0.578 by types, from 0.574 to 0.575 by tokens). When we examine the OT English grammar's ability to account for the most frequent stressed syllabic word forms (see Table 11), we see a similar pattern to the HV English grammar: the OT English grammar instantiations can already account for 7 to 8 out of 10 of them (depending on the grammar instantiation), and the ones that cannot be accounted for don't seem to involve inflectional morphology (e.g., *ready* and *baby*). So, it is perhaps unsurprising that little improvement is seen once the learner has some knowledge about the interaction of inflectional morphology with English metrical phonology.

Nonetheless, when this morphological knowledge is added, 552 word types are reduced to monosyllabic forms which are assumed to be accounted for by every grammar. Why then do we not see that reflected in the English grammars' raw compatibility? One reason (similar to the HV representation) is that many of these now-monosyllabic word forms were already accounted for by the English grammars even before knowing about the morphological interaction (e.g., *going* (LM: '10') and *getting* (XM: '10')). Thus, reducing them to monosyllabic forms doesn't allow the English grammar to account for any additional word forms. Moreover, many of the word forms without inflectional morphology become XL with contour '10' (74) and LL with contour '10' (48), which the English grammar still can't account for (see Table 11). Thus, the addition of inflectional morphology knowledge does not obviously aid the OT English grammar.

B Using the Tolerance Principle to filter the input

The Tolerance Principle (Yang 2005; Legate and Yang 2013) can be used to determine whether there is a productive rule in a set of data. Here we demonstrate the process of using the Tolerance Principle to identify a useful subset of metrical phonology data to learn from, and briefly show its impact for each of the KRs.

Suppose the learner is considering the syllabic word form V VV (which includes words such as *kítty, awáy*, and *úh óh*). This syllabic word form is perceived as a short vowel syllable (X) followed by a long vowel syllable (L). For the HV and Hayes KRs, which do not distinguish between long (VV) and super-long (VVC+) syllables, there are 506 lexicon items in the input that are of the form XL: 325 with stress contour '10' like *kítty*, 162 with stress contour '01' like *awáy*, and 19 with stress contour '11' like *úh óh*.

The Tolerance Principle predicts that a rule that should apply to N items can tolerate $\frac{N}{ln\ N}$ exceptions. So, if there are 506 XL words, a stress contour is considered the productive stress contour for XL if it has $\frac{506}{ln\ 506} = 81$ or fewer exceptions. This means that for any given stress contour, the number of XL lexical items that have some *other* stress contour associated with them must be 81 or less. As Table 12 shows, no matter which stress contour is considered, there are always too many exceptions for that stress contour to be considered the productive stress contour ('10' has 181 exceptions, '01' has 344 exceptions, '11' has 487 exceptions).

This is actually helpful for both KRs, since this means the learner should ignore all the XL syllabic word form data when trying to learn the English grammar. The HV English grammar could not account for the '01' and '11' stress contours, which comprise 181 word types of the input. The Hayes English grammar could not account for the '10' and '11' stress contours, which comprise 344 word types of the input. Thus, these 181 and 344 word types, respectively, would not "count against" the English grammars in these knowledge representations anymore. Instead, these data are ignored during acquisition.

For the OT KR, which distinguishes between long (VV) and super-long (VVC+) syllables, there are 355 lexicon items in the input that are of the form XL: 316 with stress contour '10' like *kítty*, 25 with stress contour '01' like *awáy*, and 14 with stress contour '11' like *úh óh*.

For the Tolerance Principle application, if there are 355 XL words, a stress contour is considered the productive stress contour for XL if it has $\frac{355}{ln\ 355} = 60$ or fewer exceptions. As Table 12 shows, the '10' stress contour is considered productive, since it has only 39 exceptions (as compared with the '01' and '11' contours, which have 330 and 341 exceptions, respectively).

Unlike with the parametric KRs, this turns out to be harmful for the OT English grammar. This is because most of the OT English grammars can account for the stress contour '01' (21 of 26) while the rest can account for the stress contour '11' (5 of 26). However, as neither of these is viewed as the productive stress contour, those data are ignored (39 lexicon items total). Instead, the only data used for learning the grammar are the very data that none of the OT English grammars are compatible with: the '10' stress contour data (with 316 lexicon items). Thus, these 316 word types "count against" the OT English grammar, and would likely make it more difficult to learn from realistic English input.

Stress	**# Types**	**Example**	**N**	$\frac{N}{ln\ N}$	**# Exceptions**	**Productive?**
			HV and Hayes			
10	325	*kí tty*			162+19 = 181	No
01	162	*a wáy*	506	81	325+19 = 344	No
11	19	*úh óh*			325+162 = 487	No
			OT			
10	316	*kí tty*			25 +14 = 39	Yes
01	25	*a wáy*	355	60	316+14 = 330	No
11	14	*úh óh*			316+25 = 341	No

Table 12: The Tolerance Principle applying to the stress contours associated with the XL syllabic word form for parametric and constraint-based representations. Stressed syllables are indicated with 1 and unstressed syllables are indicated with 0.

C Calculating proximity scores

A proximity score is calculated between two grammars defined by a KR, G_1 and G_2. For each linguistic variable (e.g., a parameter like Ft-Dir-Left/Ft-Dir-Rt or a constraint ordering like NONFIN>WSP-VV), if G_1 has the same value as G_2, v_{same} is 1; otherwise v_{same} is 0. The proximity score for $G1$ and $G2$ ($prox_{G_1,G_2}$) is the number of variable values that are the same, normalized so that the maximum score is 1 (see 27).

(27) Proximity score calculation between two grammars, G_1 and G_2, with v linguistic variables

For an individual parameter or constraint ordering variable v_i:

if $G_1(v_i) = G_2(v_i)$, $v_{i,same} = 1$

if $G_1(v_i) \neq G_2(v_i)$, $v_{i,same} = 0$

$$prox_{G_1,G_2} = \frac{\sum_{i=1}^{v} v_{i,same}}{v}$$

The normalization factor is the maximum score possible, and depends on the variables defined by the KR. If there are sub-parameters, the normalization factor is the number of main parameters (28a, 28b). If there are no sub-parameters, the normalization factor is simply v (28c, 28d).

(28) Proximity score calculation examples for each KR

a. **HV**

G_1: QS (QS-VC-L), Em-None, Ft-Dir-Left, Unb, Ft-Hd-Left

G_2: QS (QS-VC-H), Em-Left, Ft-Dir-Rt, B (B-2, B-Syl), Ft-Hd-Rt

total $v_{same} = 0$

$prox_{G_1,G_2} = \frac{0}{5} = 0.0$

b. **HV**

G_1: QS (QS-VC-H), Em-None, Ft-Dir-Rt, B (B-3, B-Mor), Ft-Hd-Left

G_2: QS (QS-VC-H), Em-Left, Ft-Dir-Rt, B (B-2, B-Syl), Ft-Hd-Rt

total $v_{same} = 2$ (QS (QS-VC-H), Ft-Dir-Rt)

$prox_{G_1,G_2} = \frac{2}{5} = 0.40$

c. **Hayes**

G_1: VC-H, Em-RtCons, Ft-Dir-Rt, LP-Strong, Tro-Mor, DF-Strong, WLER-R, Bot-Up

G_2: VC-L, Em-Rt, Ft-Dir-Left, LP-Strong, Tro-Mor, DF-Weak, WLER-L, Bot-Up

total $v_{same} = 3$ (LP-Strong, Tro-Mor, Bot-Up)

$prox_{G_1,G_2} = \frac{3}{8} = 0.375$

d. **OT**

G_1: TRO, WSP-VV, NONFIN, WSP-VC, ALIGN-R, FTBIN, PARSE-σ, *SONNUC, ALIGN-L

G_2: WSP-VV, TRO, NONFIN, WSP-VC, ALIGN-R, FTBIN, PARSE-σ, *SONNUC, ALIGN-L

total $v_{same} = 35$ (WSP-VV>NONFIN, ..., *SONNUC>ALIGN-L)

$prox_{G_1,G_2} = \frac{35}{36} = 0.972$

The average proximity score can be calculated for a set of grammars by calculating the proximity score for each grammar pair in the set, and then taking the average. For example, if there are 3 grammars in the set (G_1, G_2, and G_3), the average is taken of the 3 pair scores ($prox_{G_1,G_2}$, $prox_{G_1,G_3}$, and $prox_{G_2,G_3}$).

We note additionally that proximity scores are useful as a quantitative comparison within a KR, but not necessarily across KRs because their calculation depends on the number of linguistic variables in a KR. As demonstrated above, this number can differ quite significantly from KR to KR. For example, two grammars in the HV representation that differ by a single parameter value would have a proximity score of $\frac{4}{5}$=0.800, while two grammars in the OT representation that differ by a single constraint ranking would have a proximity score of $\frac{35}{36}$=0.972. Thus, proximity scores are only directly comparable within a KR.