Title
An argument from acquisition: Comparing English metrical stress representations by how learnable they are from child-directed speech

Journal
Language Acquisition, 24(4)

ISSN
1048-9223

Authors
Pearl, Lisa
Ho, Timothy
Detrano, Zephyr

Publication Date
2017-10-02

DOI
10.1080/10489223.2016.1194422

Peer reviewed

An argument from acquisition: Comparing English metrical stress representations by how learnable they are from child-directed speech

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

Abstract

It has long been recognized that there is a natural dependence between theories of knowledge representation and theories of knowledge acquisition, with the idea that the right knowledge representation enables acquisition to happen as reliably as it does. Given this, a reasonable criterion for a theory of knowledge representation is that it be useful for acquisition, particularly in non-trivial learning situations. We propose quantitative learnability metrics meant to capture how useful a representation is for acquisition. We then apply these metrics to the case study of English metrical stress, a language that is notorious for having non-productive aspects in its grammar and so being non-trivial to learn a productive grammar for. We examine three theories of metrical stress representation, assessing their learnability via these metrics from English child-directed speech at different stages of linguistic development. We find that while all three theories are only somewhat useful at the initial stages of stress acquisition, they are far more useful at later stages and define a grammar able to capture the vast majority of English children's acquisitional intake. Interestingly, we also find that the proposed English grammars in each representational theory are *not* the grammars most easily learnable from English child-directed speech. Instead, minor changes to each English grammar yield an English-like grammar that is far more learnable, suggesting that these alternative English grammars should be given further theoretical, experimental, and computational consideration. We discuss implications for both theories of representation and theories of acquisition.

1 Introduction

1.1 Why make an argument from acquisition?

One way to describe a language's grammar is as a compact knowledge system in the human mind that encodes both the productive and non-productive aspects of the language. The productive aspects are typically captured by rules, constraints, or parameters, such as using the past tense suffix $-ed$ as the default in English or knowing that syllable weight matters for English stress. Non-productive aspects include characteristics that can be captured in some compact manner as well as lexical idiosyncrasies that can't be. Examples of "compactable" non-productive aspects in English are irregular past tense paradigms such as *drink-drank/sink-sank/stink-stank* and whether a given suffix in English attracts primary stress (e.g., $-ity$ does while $-ness$ doesn't: *prodúctive* becomes *productívity* vs. *prodúctiveness*). Examples of lexical idiosyncrasies in English are the past tense form of *go* (*went*) and whether medial pretonic heavy syllables preserve stress (e.g., *for* in *ìnformátion* doesn't while *van* in *àdvàntágeous* does (Pater 2000)).

A language's grammar is useful to speakers because it allows them to immediately comprehend and generate novel linguistic items that obey the encoded system of knowledge. For this reason, grammars are viewed as generative systems. An important premise of generative grammar theorizing is that the productive parts of a language's grammar are captured by a finite set of linguistic variables that can take on a restricted range of language-specific values (e.g., specific parameter values, constraint rankings, or finite-state automata instantiations). We will refer to this set of linguistic variables as the productive knowledge representation (**prodKR**). The key idea of this approach is that the right set of linguistic variables, however implemented by the prodKR, can account for the productive aspects of the grammar of any human language.

This is likely one reason that a common criterion for a prodKR is that it be able to explain the constrained cross-linguistic variation we observe in the world's languages. The cognitive basis of this criterion is that it is surprising to see such limited variation if there is no common underlying prodKR that humans are drawing the productive part of their language-specific grammars from. ProdKR theorizing then focuses on identifying the most compact knowledge representation that can account for the observed, limited variation. In this vein, for example, Hayes (1995:55) notes that a successful representation of productive stress knowledge is one that is "maximally restrictive" and "capable of describing all the stress systems of the world's languages".

Interestingly, we typically find that comparative linguistic work leads to several theories of prodKRs that reasonably satisfy the cross-linguistic criterion in any given linguistic domain (e.g., parameters (Halle and Vergnaud 1987; Hayes 1995) and violable constraints (Tesar and Smolensky 2000; Prince and Smolensky 2002) for metrical stress representation). If these prodKRs are simply notational variants of each other, then this is unsurprising – all the prodKRs are effectively versions of the same underlying prodKR. However, it is often not obvious that existing prodKR theories are in fact notational variants. In this case, is there some other way to evaluate them besides their compact cross-linguistic coverage?

We believe there is, due to the natural dependence between theories of representation and theories of acquisition, which has long been recognized by the linguistics community. For example, Pinker (1979) suggested a "learnability condition" for theories of representation, requiring a good representational theory to explain the fact that languages can be learned. Similarly, learnability is considered a fundamental property of knowledge representations by computational learning theorists like Osherson et al. (1986), who state that the right representation describes the "collection of languages that is learnable by children".

The basic assumption behind this learnability criterion is that the prodKR should be useful for acquisition. In particular, children armed with knowledge of the linguistic variables of the prodKR should have a huge advantage when it comes to learning the productive aspect of their language-specific grammars. This is because the right prodKR already defines the hypothesis space of grammars that could encode the productive aspects of the language. So, instead of trying to figure out what variables matter for defining grammars, children can focus on identifying the appropriate grammar from the prodKR based on their language input. In theory, this simplification of the task allows acquisition to occur reliably and fast, even for non-trivial cases (e.g., as suggested by Chomsky 1981; Dresher 1999; Crain and

Pietroski 2002). This also accords with the consensus in computational learning theory that structured hypothesis spaces, such as those defined by prodKRs, are necessary for language acquisition to feasibly occur (Heinz 2014).

Therefore, if the right prodKR is meant to be useful for acquiring the productive aspects of individual languages from children's input, can we formalize this as a way to evaluate different prodKR theories? We are particularly interested in comparing prodKR theories that already satisfy the cross-linguistic criterion. The difficulty, of course, is that acquisition is complicated, and so evaluating how useful a given prodKR is for acquisition is non-trivial. However, if we can find a reasonable way to do so, there is much to be gained for both theories of representation and theories of acquisition.

First, we would have a new metric for evaluating representational theories, and we can see if all the current prodKRs satisfy this learnability criterion equally well, or if some are better than others. Second, because this evaluation requires us to be concrete about how acquisition proceeds using a particular prodKR, we become aware of the learning assumptions each prodKR requires. That is, we more concretely describe the learning theory that accompanies a given representational theory. Third, because we will be evaluating prodKRs on their ability to acquire the productive aspects of specific languages, we have an opportunity to assess the language-specific productive grammars proposed for those languages in each prodKR. For example, if each prodKR is being evaluated on its ability to acquire the productive aspects of English, each prodKR will have a set of values meant to capture English productive aspects. Interestingly, these productive grammars are often derived by careful consideration of adult linguistic knowledge, and so can benefit from being evaluated on how acquirable they are from children's primary linguistic input – particularly when compared against other potential grammars in the prodKR.

1.2 Our approach

A child armed with a prodKR has to solve two problems: identification and filtering. On the identification side, the child is trying to determine what the productive part of the language's grammar looks like, given the available language data. That is, using the prodKR, she is trying to identify the language-specific values of the grammar variables defined by the prodKR. On the filtering side, she is trying to determine the features in the input that mask the productive part of the grammar – and then filter them out when identifying the productive part. This means that at some point, the child needs to recognize that there *are* in fact non-productive aspects of the grammar that are learned separately from the productive aspects defined by the prodKR. Importantly, because (by definition) these non-productive aspects don't pattern like the productive aspects of the language, they can interfere with identifying the productive aspects until they are filtered out.

In terms of the recent acquisition model of Lidz and Gagliardi (2015), identification occurs via inference over the *acquisitional intake*, which is the child's input filtered in various ways. Initially, the input is filtered via the variables of the prodKR, which highlights only those input aspects viewed as acquisitionally relevant. The acquisitional intake may also be filtered by additional initial biases the child has (e.g., an unambiguous data bias: Fodor 1998; Pearl 2007, 2008, 2011). Later, the acquisitional intake may be further filtered by the child's own linguistic experience, which helps determine which input features are productive (e.g., Yang

2005; Legate and Yang 2013).

Under this view, the child's acquisitional intake is determined by the prodKR itself, as well as any active filtering biases. Therefore, we propose to compare prodKR theories by the acquisitional intake they create from children's primary linguistic data. If the acquisitional intake is easy to learn from, then we can argue the prodKR is useful for acquisition; conversely, if the acquisitional intake is hard to learn from, we can argue the prodKR is not useful for acquisition.

Below we present analyses of this kind, representing different stages of the acquisition process that create different acquisitional intakes from the available linguistic data. The first analysis type is meant to capture the initial stages of prodKR grammar acquisition and incorporates no filtering beyond what the linguistic variables of the prodKR provide. Specifically, this learner has the naive assumption that all the language data are generated by the productive grammar. This is useful as an acquisitional baseline: do the different prodKR theories actually need any additional filters or is the language-specific grammar learnable even from a minimally filtered acquisitional intake?[1]

The second analysis type is meant to capture later stages of prodKR grammar acquisition and incorporates derived knowledge of the language-specific non-productive aspects. This serves as a more sophisticated acquisitional evaluation: if a prodKR theory does not enable acquisition to succeed when the learner uses the initial acquisitional intake, does this filtering then create an acquisitional intake that enables acquisition to succeed? If so, which specific filters do so? In general, we might reasonably expect prodKR grammar acquisition to be easier when the child only pays attention to the data that appear to be productive. This is because these are exactly the data that prodKR theories are meant to capture. That is, the non-productive noise has been effectively filtered out and so the child's acquisitional intake should be a much cleaner source for inferring the correct prodKR grammar.

Here we use this approach for the case study of metrical stress prodKR representations, which are evaluated on their ability to make the productive aspects of English stress learnable. Section 2 proposes specific learnability metrics, which are intended to be applicable for evaluating prodKR theories more generally. We then review relevant aspects of English stress in section 3, including the productive aspects typically captured by prodKR theories and non-productive aspects that must also be acquired. In that section, we additionally review empirical evidence about the acquisition trajectory for English stress because this information helps determine when and how knowledge of English-specific non-productive aspects might be integrated into a learner's developing grammar. The developmental trajectory also provides guidance about the exact type of input children are exposed to as their English prodKR grammars are developing, since we know that child-directed speech changes qualitatively with the age of the child (Bernstein Ratner 1984; Gleitman et al. 1984; Kitamura and Burnham 2003). We then discuss the specific set of English child-directed speech data we use for our analyses.

We turn in section 4 to three prodKR theories of stress that have an English prodKR grammar and can be evaluated via the proposed learnability metrics. Section 5 presents

[1] We note that it's probable that additional filters are required when the child's input is particularly noisy (i.e., includes a significant portion of non-productive data), as is the case with English. This is discussed in more detail when we review the relevant aspects of English stress.

results from the two analysis types described above: (i) a learner in an initial state with acquisitional intake that is minimally filtered, and (ii) a learner in a later state with acquisitional intake that is further filtered. We find that while all three prodKR theories struggle to capture all the patterns in the minimally filtered acquisitional intake, they all do quite well with acquisitional intake that is filtered in various ways – that is, they define a productive grammar that can capture the vast majority of the learner's acquisitional intake at that stage. This suggests that all three prodKR theories originally constructed to satisfy the cross-linguistic criterion can also satisfy the learnability criterion under certain learning assumptions – something that did not necessarily need to be true. Interestingly, we also find that the proposed English-specific grammars in each prodKR are *not* the ones that are most easily learnable from the acquisitional intake, no matter what the learning assumptions are. However, minor changes in each prodKR's English grammar yield an English-like grammar that is far more learnable. We suggest that these alternative representations for the productive English grammar be given further theoretical, experimental, and computational consideration as either the true English grammars in each prodKR or as transitory grammars English children using that prodKR would converge on during development. We conclude with implications for theories of representation and theories of acquisition.

2 Learnability metrics

The essence of learnability is simply how easily children could learn a language's grammar when given data from that language to learn from. This is a more targeted form of the general aim of computational learnability theory, which investigates whether it possible to learn a language (or class of languages) from certain types of input data (Heinz 2014).

One key finding in learnability theory has been that feasible learning is possible only when the hypothesis space of possible grammars is restricted and structured appropriately (see Heinz 2014 for an accessible summary of the literature on this point). This is precisely what a prodKR is meant to do, i.e. restrict the child's attention to certain possibilities for the productive grammar of the language. So, in this same spirit, it is reasonable to ask if a specific prodKR theory actually makes learning feasible for a particular language.

Another core finding in learnability theory has been that it is often possible in principle to learn certain types of languages when the input is restricted in various ways (e.g., only data generable by primitive recursive functions (Gold 1967) or only data that are complete and computable (Angluin 1980)). Here, we restrict the input to the data that children would likely encounter in their environment, thus satisfying Pinker's (1979) "input condition", which states that modeled learners should only use the information typically available to children. Moreover, this better represents "the circumstances of actual linguistic development in children" (Osherson et al. 1986) that are of the most interest to developmental linguists.

We focus on how straightforward it would be to learn a particular language's grammar from these data, given a particular prodKR theory. We will assume children are already aware of the prodKR – and importantly, the variables that the prodKR indicates are relevant for determining the language's productive grammar.

2.1 The learnability approach

Many different approaches to assessing learnability exist (e.g., Dresher and Kaye 1990; Clark 1992; Clark and Roberts 1993; Gibson and Wexler 1994; Niyogi and Berwick 1996; Tesar and Smolensky 1996, 1998; Dresher 1999; Tesar and Smolensky 2000; Sakas and Fodor 2001; Pearl 2011; Clark and Lappin 2012; Sakas and Fodor 2012; Legate and Yang 2013; Fulop and Chater 2013; Heinz 2014). Typically, studies following these approaches have assessed a prodKR grammar's learnability by how well it accounts for the language's data. This assessment can be based more directly on the observable form of the language's data (e.g. Clark 1992; Clark and Roberts 1993; Gibson and Wexler 1994; Niyogi and Berwick 1996; Tesar and Smolensky 1996, 1998, 2000; Pearl 2011; Legate and Yang 2013) or via a filtered version of the observable data that includes only the relevant structural cues embedded in the data points themselves (e.g. Dresher and Kaye 1990; Dresher 1999; Sakas and Fodor 2001, 2012). Either way, there seems to be an implicit idea that the right grammar in a prodKR is the one with the best fit to the appropriate acquisitional intake.

This is intuitively similar to other assessment metrics outside the domain of grammar learning, such as the principle of Empirical Risk Minimization (ERM) (Vapnik 1992, 2013) in statistical learning theory. In ERM, the learner picks a hypothesis that minimizes error on the training data. Here, this would mean a learner picks a grammar that minimizes error on the acquisitional intake (i.e., the grammar best fits the acquisitional intake). Similarly, the principle of Minimum Description Length (MDL) (Rissanen 1978; Grünwald 2007) used in both information theory and computational learning theory, includes a component that captures the likelihood of the data, given the hypothesis. Here, this corresponds to the fit of the grammar to the acquisitional intake.

Within the domain of grammar learning, we might ask why the quantity of acquisitional intake data accounted for should be so important. One answer is that this relates to the utility of productive grammars: a productive grammar is useful because it allows the learner to compactly represent the productive aspects of the language data, and so language data captured by the productive grammar do not need to be stored in detail. Instead, the productive aspects of these data can be generated by the compact representation provided by the grammar. So, the more data accounted for by the productive 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 productive grammar is naturally defined as the one that can account for the most data.

The specific learnability approach we pursue here is similar to those taken by Pearl (2011) and Legate and Yang (2013). In particular, we will assess (i) learnability from child-directed speech input and (ii) learnability at the computational level (in the sense of Marr 1982).

By evaluating learnability with child-directed speech input, we can more concretely link learnability to the language acquisition task children actually face. This contrasts with many of the other studies mentioned that investigated learnability over idealized data sets intended to capture salient properties of children's possible acquisitional intake (e.g., Dresher and Kaye 1990; Clark 1992; Clark and Roberts 1993; Gibson and Wexler 1994; Niyogi and Berwick 1996; Tesar and Smolensky 1996, 1998; Dresher 1999; Tesar and Smolensky 2000; Sakas and Fodor 2001, 2012).

By evaluating learnability at the computational level, we can focus on the utility of the hypothesis space defined by the prodKR theory. That is, 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 for 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 productive grammar,[2] and we suggest (following the intuition of many previous learnability studies) that the best productive grammar is the grammar able to account for the most data in children's acquisitional intake.

One important note about this kind of learnability evaluation is that it is focused solely on the practical application of data coverage. It does *not* care about whether a prodKR theory is appropriately restrictive or economical, which is clearly something we believe is important for a prodKR – a prodKR is meant to compactly represent the productive regularities in the data, after all. So, the learnability approach we pursue here is intended for comparing prodKR theories that have already satisfied those other critieria for prodKR "goodness". When we have a set of such prodKR theories, then the learnability metric proposed here can be used to provide support for or against these prodKR theories.

2.2 Specific learnability metrics

Once we define the data in children's acquisitional intake, we can evaluate the productive grammars defined by a prodKR theory on their ability to account for these data. At an individual data point level, a grammar can either be compatible or incompatible with the data point. For example, a metrical stress grammar is compatible with a data point if it can generate/select 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 (e.g., a grammar compatible with 70% of the data set has a raw compatibility of 0.70). When comparing productive grammars within a prodKR, a higher raw compatibility is better since this indicates the grammar is more useful at accounting for the available data. Thus, the best productive grammar will have the highest raw compatibility, and be the most useful.

From a learnability perspective however, what matters more than raw compatibility is

[2]We do recognize that children may not identify the grammar that best captures the data they're exposed to, for various reasons. For example, they may not be capable of optimal inference, due to cognitive resource constraints, and this may lead them to different – though still very useful – answers (e.g., Phillips and Pearl 2015). Or, children may not be seeking the optimal grammar for other reasons, perhaps due to their language learning biases. In this second case, it is important for both theoretical and developmental linguists to be precise about the learning theory that explains how and why sub-optimal language-specific grammars are learned from children's language data. Still, before going down that route, it is useful to first know if such learning mechanisms are necessary, which is exactly what the learnability approach proposed here can identify. That is, before concerning ourselves with exceptional learning mechanisms to explain how a child learns the language-specific grammar, we should first see whether the language-specific grammar is optimal in the sense we describe here. If so, then general learning mechanisms aimed at identifying the optimal grammar should suffice.

how a productive grammar compares to other grammars defined by the prodKR 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 be better than all other grammars, and so its relative compatibility approaches 1 as the number of grammars in the hypothesis space increases. 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 prodKR theory.

However, suppose 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. This situation may occur if more than one grammar can account for exactly the same amount of data. In this case, 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).[3]

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 productive grammar. However, this is not required – the best productive 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 productive grammar to explain the language data in the acquisitional intake. 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.

While the previous metrics focused on evaluating the learnability of grammars within a prodKR, we can also evaluate prodKRs themselves. In particular, we can calculate the *learnability potential* of a prodKR, which is simply the raw compatibility of the best grammar defined by the prodKR. For example, if the best grammar in a prodKR (with relative compatibility and relative class compatibility closest to 1.00) has a raw compatibility of 0.70, then that prodKR has a learnability potential of 0.70. In effect, this metric indicates the utility of the prodKR, as instantiated by the best grammar it defines. This is because the

[3]We note that because these implementations of relative and relative class compatibility depend on the number of grammars being finite, only prodKRs with a finite number of grammars can be evaluated using these metrics. As Heinz (2014) notes, prodKRs with infinite hypothesis spaces can exist and may have interesting learnability implications. For such prodKRs, the intuitions behind relative and relative class compatibility may still be useful, but would have to be implemented a different way.

learnability potential indicates how good the productive grammar variables defined by the prodKR are at accounting for the available data in the child's acquisitional intake.

2.3 Evaluating the language-specific grammar

Language-specific productive grammars have often been derived with the goal of accounting for the language-specific knowledge adults have (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. Still, the language-specific productive grammar defined by the prodKR theory should be learnable from the data children typically encounter, since this is a main motivation from an acquisition perspective for having a prodKR.

To satisfy the learnability criterion in the most straightforward way, we suggest that the language-specific grammar be the grammar that is learned most easily from the language's acquisitional intake. 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 closest to 1.00. This, in turn, would cause this grammar's raw compatibility to be equivalent to the learnability potential of the prodKR that defines it, since it would be the grammar defined by that prodKR that best accounts for the language's acquisitional intake.

3 English stress

3.1 Productive and non-productive aspects of English stress

The description of English stress, including its productive and non-productive aspects, is decidedly non-trivial. Four relevant core patterns of monomorphemic words are as follows:

1. Stress must occur on at least one of the last three syllables (Hammond 1999).

2. Syllable weight impacts stress placement (Chomsky and Halle 1968; Halle and Kenstowicz 1991; Hammond 1999; Pater 2000), and heavy syllables (containing a tense vowel like /*i*/ or closed by at least one consonant like /*ɛn*/) often bear stress. This property is typically referred to as *quantity sensitivity*.

3. Within a trisyllabic window, stress is also a function of lexical factors, and certain exceptional cases are thought to be represented individually rather than being captured by a more general pattern (Hammond 1999).

4. The stress pattern of nouns is different from that of verbs and adjectives (Chomsky and Halle 1968; Hayes 1982; Kelly 1988; Kelly and Bock 1988; Hammond 1999; Cassidy and Kelly 2001). For instance, there are examples like *cónduct*/*condúct* and *désert*/*desért*, where the syntactic category influences the stress pattern for the syllable sequence (e.g., the noun versions are stress-initial while the verb versions are stress-final).

For words containing more than one morpheme, Hammond (1999) notes that there is a class of affixes "outside the domain of stress assignment" (e.g., *-able, -ed, -ing, -s, -or, -er, -ly, -able, -ment, -ness*) that allow modifications to the above patterns. More generally, there are known interactions with inflectional and derivational morphology (Chomsky and Halle 1968; Kiparsky 1979; Hayes 1982, 1995). For example, in *prétty/préttier/préttiest* and *sensátion/sensátional/sensátionally*, adding inflectional and derivational morphology does not shift the primary stress, despite adding syllables to the word.

In terms of representation, the goal of prodKR theories has been to define an English productive grammar that compactly captures as much of this variation as possible, leaving non-productive aspects to be encoded some other way. This ties in nicely to an acquisition perspective: the child equipped with the prodKR should be able to learn the productive aspects of English stress via the prodKR variables while (eventually) filtering out the non-productive aspects. This underscores why it's important to specify the learning assumptions that accompany a particular prodKR theory. More specifically, we can determine two things for a given prodKR theory: (i) how necessary this kind of filtering is when learning the productive English stress system from the data English children typically encounter, and (ii) what effect a plausible implementation of this kind of filtering has on the child's acquisitional intake.

3.2 The developmental trajectory

We turn now to what children seem to know when about English metrical stress. Experimental data suggest that acquisition of the productive aspects progresses in stages. 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). This is useful for capturing the stress pattern of words like *cáptain*, *húngry*, *fíftieth*, *láter*, *ópening*, *zébra*, *ángel*, *grátefully*, and *fábulous*, which are found in the child-directed speech corpus we describe in section 3.3. By age three, children have recognized that the metrical system is quantity sensitive, though they do not recognize the full set of factors that determines how syllable weight impacts stress placement (Kehoe 1998). By age four or five, there is suggestive evidence that English children have identified the target English productive grammar: (i) Arciuli et al. (2010) find that children as young as five override orthographic cues to alternative stress patterns that violate the English productive grammar, and (ii) Pettinato and Verhoeven (2008) find that children as young as four are at ceiling for repeating nonsense words that obey the English productive grammar but not for words that violate it.

These experimental findings provide helpful guideposts for the analyses we wish to conduct. First, in terms of the data children are learning from, we likely want to restrict our analyses to child-directed speech that is encountered before the age of four or five, and perhaps initial state analyses should endeavor to use data encountered before the age of two. Second, in terms of when non-productive aspects may be filtered out of the child's acquisitional intake, we likely want to restrict our analyses to aspects that are recognized before the age of four or five. This is particularly relevant when we consider the aspects known to interact with English stress, such as inflectional morphology, derivational morphology, and syntactic category.

For morphology, recall that there is a class of affixes outside the domain of stress assignment, including both inflectional affixes like *-ing* (*ánswer*/*ánswering*) and derivational affixes like *-ness* (*cránky*/*cránkiness*). When do children recognize these affixes? If it's before the age of four or five, then they may be able to notice that these affixes do not impact the stress contour and filter them out of the acquisitional intake. Experimental evidence suggests that 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)). In contrast, children develop knowledge of inflectional morphology much earlier, often using it productively in their own utterances by age three (Brown 1973). This suggests that a reasonable filter on the acquisitional intake could include ignoring inflectional morphology (e.g., *ánswering* would be filtered to *ánswer*), but probably should not include ignoring derivational morphology (e.g., *cránkiness* is left as is). This is because children do not seem to recognize derivational affixes until after the productive aspects of English stress have been acquired.

For syntactic category, recall that noun stress patterns differ from verb and adjective stress patterns. Experimental evidence suggests that children are aware of noun and adjective categories as early as fourteen months (Booth and Waxman 2003), while the verb category is argued to be available as early as two years old (Kowalski and Yang 2012). Though there is significant debate in the acquisition literature about when the verb category is first recognized, it is likely before the age of five as children produce verb-based overregularization errors before then (e.g., Cazden 1968; Kuczaj 1977; Marcus et al. 1992; Marchman et al. 1997) and show evidence of across-verb priming (e.g., Thothathiri and Snedeker 2008b,a). This suggests they have categorical knowledge of verbs. Therefore, children may recognize that there is an interaction of syntactic category on stress pattern before the age of five, and perhaps filter out words from particular categories if these words obey category-specific patterns while violating the productive grammar of English.

Given these considerations, our analyses will be conducted over speech directed at children under the age of two since English children seem to have learned some of the productive aspects of English stress by that age.[4] The first set of analyses will assume no knowledge of non-productive aspects and so represents the initial stage of English stress acquisition. The second set of analyses will assume cognitively plausible filters on children's intake, based on the experimental data discussed above, and so represents a later stage of acquisition.

3.3 English acquisition input

We selected the Brent corpus (Brent and Siskind 2001) from the American English subsection of the CHILDES database (MacWhinney 2000), which contains speech directed at children between the ages of six and twelve months (4780 multisyllabic word types, 99968 multisyllabic word tokens). Word types were syllabified using CELEX 2 (Baayen et al. 1995), the CALLHOME English Lexicon (Kingsbury et al. 1997), and the MRC Psycholinguis-

[4] We recognize that this may underestimate the complexity of children's input as development progresses, since speech directed at three- and four-year-olds may contain more linguistic variety than speech directed at children younger than two. It would therefore be useful for future work to conduct the learnability analyses we propose here on speech directed at older children to see if the results qualitatively change. If they do, this suggests a significant role for the nature of the input as children develop, in addition to children's own developing internal representations and perceptions.

tic Database (Wilson 1988). Stress was estimated using the CALLHOME English Lexicon (Kingsbury et al. 1997).[5] For words not in these resources (typically nonsense words like *fussies* or child-register words like *fishies*), we used native speaker intuitions and comparisons to similar sounding words where possible.

4 Productive knowledge representations for stress

When learning about metrical stress, the observable data are the stress contours associated with words. For example, *óctopus* 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. We note that 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. Given this, all the prodKR theories we examine only include the components of those prodKRs that impact the distinction between stressed and unstressed syllables. In addition, these prodKR theories all define grammars that assume a word has been divided into syllables and those syllables are classified according to their syllable rimes. So, 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 at most a single syllable within a metrical foot being stressed. A grammar defined by a prodKR will be associated with an underlying metrical structure, as shown in (1), whose observable form is the stress contour for the word.

(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

We now briefly review the three prodKR theories we will compare, which include both parametric and constraint-ranking representations.

4.1 Parametric theories of knowledge representation

4.1.1 The HV parametric representation

The first parametric prodKR 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.[6] For a more detailed description of each of the parameters and their interactions with each other, see Pearl (2007).

[5] The syllabified and stressed annotations for the word types are available at *http://www.socsci.uci.edu/~lpearl/CoLaLab/uci-brent-syl-structure-Jul2014.xlsx*.

[6] 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).

4.1.1.1 HV parameters

4.1.1.1.1 Quantity Sensitivity. Quantity sensitivity determines whether syllables are treated identically or instead differentiated by syllable rime weight for the purposes of stress assignment. A language could be quantity sensitive (**QS**), so that syllables are differentiated into (**H**)eavy and (**L**)ight syllables. Long vowel syllables with or without codas (VV(C)) are Heavy, short vowel syllables (V) are Light, and short vowel syllables with codas (VC) can be either Light (**QS-VC-L**) or Heavy (**QS-VC-H**), yielding three syllable type distinctions (long, short, and closed).[7] 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

4.1.1.1.2 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
syllables	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
syllables	æf	təɹ	nun

4.1.1.1.3 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)

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

4.1.1.1.4 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)

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)

4.1.1.1.5 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**), as in (7a), while feet headed on the right have the rightmost syllable of the foot stressed (**Ft-Hd-Rt**), as 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)

4.1.1.2 The HV English grammar. The English grammar proposed 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⟩
syllables	ɑk	tə	pʊs

4.1.2 The Hayes parametric representation

The second parametric system is adapted from Hayes (1995) (**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.

4.1.2.1 Hayes parameters

4.1.2.1.1 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 *sleeping* (/slipɪŋ/):

	VC-H		**VC-L**	
syllable class	H	H	H	L
syllable rime	VV	VC	VV	VC
syllables	sli	pɪŋ	sli	pɪŋ

4.1.2.1.2 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	sli	pɪŋ	sli	pɪŋ

4.1.2.1.3 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**).

4.1.2.1.4 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)

4.1.2.1.5 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-Mor**					
Tro-Syl foot 1	(H́	L)	L	H	H	Tro-Mor foot 1	(H́)	L	L	H	H
Tro-Syl foot 2	(H́	L)	(Ĺ	H)	H	Tro-Mor foot 2	(H́)	(Ĺ	L)	H	H
Tro-Syl foot 3	(H́	L)	(Ĺ	H)	(H́)	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)

4.1.2.1.6 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 ≠ degenerate foot.	L	⟨H⟩	L = degenerate foot.	(Ĺ)	⟨H⟩
Analysis fails.					

4.1.2.1.7 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)

4.1.2.1.8 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. 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 H́)	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)

4.1.2.2 The Hayes English grammar. The English grammar proposed 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.[8] 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 an 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⟩

4.2 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 derived from Hammond (1999) and Pater (2000), there are nine metrical stress 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

[8] We note that WLER-R does matter for distinguishing between primary and secondary stress one metrical layer above.

so entails that their analyses contain at least one metrical foot. Therefore, only candidate analyses that have at least one metrical foot are considered by the learner.

4.2.1 Constraints

4.2.1.1 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)	∗!

4.2.1.2 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)	∗!

4.2.1.3 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)	∗!

4.2.1.4 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)	∗!

4.2.1.5 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	∗!

4.2.1.6 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 (ráj) zən	∗!
c. ☞ hə raj (zʌ́n)	

4.2.1.7 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 (ráj) zən	∗!
c. hə raj (zʌ́n)	∗!∗

4.2.1.8 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	∗!

4.2.1.9 *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)	∗!

4.2.2 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/).[9] 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.

4.2.3 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 derived from 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.

The tableau in Figure 2 is an evaluation of *little* (/lɪɾɪl/) using a grammar satisfying the English constraint rankings. Because the final /l/ could be the nucleus of the second syllable, eight candidates are generated. For this word form, the optimal candidate for the grammar has a stress contour that matches the observed stress contour of *little* (*lı́ttle*).

4.3 Knowledge representation comparison

While these prodKR theories are not obviously 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 prodKRs are able to account for the observed constrained cross-linguistic variation by drawing on sets of linguistic variables that are noticeably different.

[9] 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.

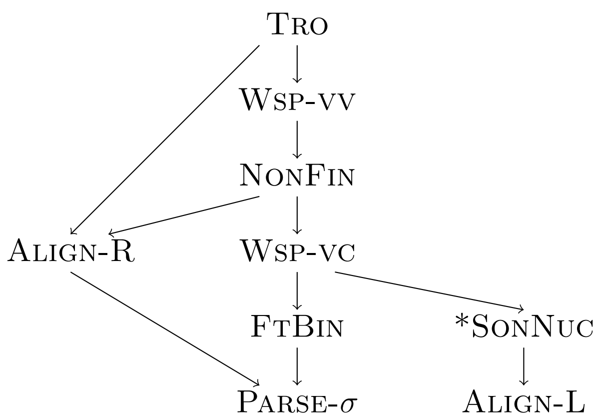

Figure 1: Partial ordering of constraints defining the English grammar.

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)			*!		*	**		*	*

Figure 2: Evaluation of *little* using a grammar that satisfies the English constraint rankings.

Due to the linguistic variables defined by each prodKR, 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.

In addition, the size of the hypothesis space of grammars defined by each prodKR differs, with the parametric prodKRs numbering the possible grammars in the hundreds (HV: 156, Hayes: 768), while the constraint-ranking prodKR numbers its possible grammars in the hundreds of thousands (OT: 362,880). While this does not impact the computational learnability analysis we pursue here, it could very well affect an algorithmic-level approach, which would include how a learner could efficiently search the hypothesis space of grammars.[10]

[10] One helpful attribute is whether the grammars in a hypothesis space are extensionally distinct, i.e., predict different sets of data to be grammatical. As we saw for the English OT "grammar", 26 different grammars are assumed to be extensionally equivalent for English, so that an English learner identifying any one of them would be counted as successful. In essence, when grammars are extensionally equivalent, they cease to be distinct from the learner's perspective. This can aid learning by effectively shrinking the hypothesis space of grammars.

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 prodKR theories for metrical stress, aligning parameters/constraints that refer to similar aspects of metrical structure.

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 prodKR (HV: 3, Hayes: 4, OT: 8), indicating how each syllable type is classified. The relevant syllable in the examples 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 productive grammars defined by these prodKR 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 productive grammars in the different prodKR 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.

4.4 The problem of multiple stress contours

One issue that occurs for English, due to the non-productive aspects of its grammar, is that multiple stress contours can occur for the same syllabic word form. For example, the syllabic word form V VV, which is a short vowel syllable followed by a long vowel syllable, describes the words *kítty*, *awáy*, and *úh óh* within the Brent corpus of English child-directed speech, each of which has a different stress contour. This is problematic because a prodKR 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 child's input. For instance, a grammar that generates/selects the appropriate stress contour for *kítty* will not be able to do so for *awáy* and *úh óh* since it will generate/select the same stress contour for them as well.

But how often does a syllabic word form like V VV have multiple stress contours associated with it in English child-directed speech data? In the Brent corpus, between 37% and 58% of the syllabic word forms (depending on the syllabic distinctions made by a prodKR theory) have multiple stress contours associated with them. This underscores why no single productive grammar can be compatible with all the input data, and also why filtering of the input is likely helpful for identifying the productive grammar of English.

	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 in English child-directed speech with multiple stress contours for each prodKR theory.

5 Learnability analyses

For each of the learnability analyses reported below, an algorithm was run that evaluated the compatibility of each grammar in a prodKR against the data in the English learner's acquisitional intake (e.g., all the word types abstracted into their respective syllabic word forms, with stressed and unstressed syllables identified, so that *báby* would become VV VV with stress contour 10). This allowed each grammar in a prodKR to be evaluated against each data point in the acquisitional intake, and either be able to account for that data point or not. This calculation was then used to generate the raw compatibility, relative compatibility, relative class compatibility, and learnability potential scores.[11]

5.1 Initial state analyses

The first set of analyses represents the initial state of metrical stress acquisition, when the child is not yet aware of the non-productive aspects of English. In other words, the learner being modeled has the naive assumption that all the data in the input are generated by the productive English grammar.

5.1.1 Learnability potential

Given how many syllabic word forms have multiple stress contours, it is reasonable to wonder how well any one grammar within these prodKR theories could possibly do. In particular, what is the largest quantity of data that any single productive grammar can account for? This represents the learnability potential of the prodKR. It turns out that all three prodKRs have a grammar that is able to account for about $\frac{2}{3}$ of the word types (0.654-0.683), as shown in Table 5. This suggests that the best grammar in each representation is fairly 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 prodKR theory is capable of defining a productive grammar that would be useful for the child to acquire at this stage.

5.1.2 English grammar compatibility

Since it is possible to learn a useful productive 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 prodKRs (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, accounting for 0.485-0.593 of word types. 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 identifying 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

[11]Code for the compatibility algorithm, along with documentation on how to use it and sample input files, are available at *https://github.com/lisapearl/compatibility-metphon*.

have a better chance of identifying the English grammar, especially if they are not perfectly optimal 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 child-directed speech input.

Unfortunately, this turns out not to be true for any prodKR. As Table 5 shows, the parametric English grammars are worse than about $\frac{1}{3}$ of the grammars in the hypothesis space (0.673-0.676) and the best constraint-based grammar is worse than about $\frac{1}{5}$ of the grammars in the hypothesis space (0.816). This indicates that while the English grammars are better than many other grammars, there are a large number of grammars that are better than the English grammars. For the parametric prodKRs, tens or hundreds of grammars are better able to account for the English input (HV=51, Hayes=249) while for the constraint-based prodKR, tens of thousands of grammars are better (OT=66,886). 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 is nowhere near the best (parametric: 0.684-0.697; constraint-based: 0.744).

5.1.3 Learnability summary

Table 5: Learnability analyses for the three prodKR theories: HV, Hayes, and OT. The four metrics shown are learnability potential of the prodKR (prodKR: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 in English child-directed speech.

	prodKR:Pot	**Eng:Raw**	**Eng:Rel**	**Eng:RelClass**
HV	0.668	0.593	0.673	0.697
Hayes	0.683	0.485	0.676	0.684
OT	0.654	0.570	0.816	0.744

While all three prodKR theories have a productive grammar that is useful to learn for English input, there are learnability issues when it comes to selecting the proposed English grammar for each one. This is because the proposed English grammar is not the one most compatible with this acquisitional intake, given the hypothesis space of grammars defined by each prodKR. More specifically, a rational learner (or even a mostly rational one) looking for the productive grammar best able to account for this English child-directed speech input would not select the target English grammar in any of these prodKRs.

Nonetheless, we can still say something comparative about the prodKR theories, based on these results. Both the HV and OT prodKRs have English grammars that are more useful to acquire, since their raw compatibility with the data is much higher than that of the Hayes prodKR's English grammar. The OT prodKR may have the most learnable English grammar at this stage of development since its relative compatibility with the data is the highest (relative: 0.816, relative class: 0.744). So, given this, the OT prodKR may have a slight advantage on the other two, though the core learnability issue remains for all three.

More generally, these results suggest that something additional is required to make the proposed English grammars in each prodKR learnable from typical English input. We next explore the impact of cognitively plausible filters on the child's acquisitional intake.

5.2 Analyses with further filtering

The analyses below involve further filtering of the acquisitional intake, and therefore represent a learner who does not assume all the data in the input are generated by the productive grammar. We specifically consider the impact of derived knowledge of different kinds, first discussing the knowledge itself that might be useful and then a learning bias that could help derive that useful knowledge over time.

5.2.1 Derived knowledge of inflectional morphology

Because children appear to derive knowledge of inflectional morphology before they finish acquiring the productive grammar of English (see section 3.2), they may notice that inflectional morphology does not affect a word's stress contour. To investigate the impact of this knowledge, we re-analyzed the English input data for their compatibility with the various grammars after removing inflectional morphology (shown in Table 6).[12] This simulates the child ignoring inflectional morphology in the input when learning the English productive grammar (i.e., the child's acquisitional intake has word types stripped of their inflectional morphology). Results of this analysis are shown in Table 7.

The learnability potential of the three prodKRs remains about the same: around $\frac{2}{3}$ of the word types (0.662-0.683) can be accounted for by the best productive grammar defined by each prodKR. This indicates that the best grammar in each prodKR is still very useful to learn, since it can account for a large portion of the filtered intake. Interestingly, this derived knowledge about inflectional morphology does not help the best grammar account for much more data than it could before, no matter which prodKR.

However, perhaps the utility of this derived knowledge is targeted more at improving the coverage of the (best) English grammar within each prodKR. 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 prodKR (English=0.550-0.605 vs. best= 0.662-0.683).[13] Thus, even with this derived inflectional morphology knowledge, the (best) English grammar is still not the best grammar overall, and so a rational learner would not select it based on English child-directed speech 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 derived knowledge about inflectional morphology? As Table 7 shows, the parametric English grammars are still worse than about $\frac{1}{3}$ of the grammars in the hypothesis space (0.704-0.712) and the best constraint-based grammar is worse than about $\frac{1}{5}$

[12] 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.

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

Table 6: Inflectional morphology ignored once the child has derived the knowledge that inflectional affixes do not impact English stress. 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	/r̩/	*longer*
–est	/ɪst/	*sweetest*

Table 7: Learnability analyses for the three prodKRs (HV, Hayes, and OT) once children derive the knowledge that inflectional morphology does not typically affect the stress contour. The four metrics shown are learnability potential of the prodKR (prodKR: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 in English child-directed speech.

	prodKR:Pot	**Eng:Raw**	**Eng:Rel**	**Eng:RelClass**
HV	0.662	0.605	0.712	0.706
Hayes	0.683	0.550	0.704	0.713
OT	0.674	0.575	0.786	0.641

of the grammars in the hypothesis space (0.786). This again indicates that while the English grammars are better than many other grammars, 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 prodKR, the (best) English grammar's compatibility with the English data lags behind (0.641-0.713). As before, the target English grammar is unlikely to be easily learnable from this hypothesis space of grammars, even with this derived knowledge about some interactions between metrical stress and morphology.

So, the same learnability issues persist. However, as before, we can still say something about the prodKR theories relative to each other. With this inflectional knowledge, the HV prodKR now has the most useful English grammar to acquire since its raw compatibility is highest. The OT prodKR has the most learnable English grammar by relative compatibility, while the HV and Hayes prodKRs have more learnable English grammars with respect to relative class compatibility. So, all three prodKR theories have English grammars that are more or less equally learnable once the learner has this inflectional knowledge, and the OT prodKR theory no longer has a clear learnability advantage.

5.2.2 Other derived knowledge

One pervasive issue even with this more filtered acquisitional intake is that the English grammars in all three prodKRs typically want to stress syllables with a long vowel nucleus (e.g., *sweet*).[14] This can be problematic for English 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 prodKR 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 perceive them differently – and effectively filter them out. 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 productive English metrical stress grammar. They could then use this knowledge to perceive these diminutive data differently, ignoring the diminutive affix for purposes of metrical stress. The diminutives then become compatible with the English grammars in all three prodKR theories, because the diminutive affix is filtered out.

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 productive

[14] 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.

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 non-productive classes of words with different stress patterns from the productive grammar of English. Syllable word forms with multiple stress contours can signal this, such as bisyllabic nouns and verbs (e.g., *cónduct* vs. *condúct*). The syntactic category of the word is the predictable cue to the stress contour. Once children realize that there are predictable exceptions (which the syntactic category stress contours demonstrate), they may allow other exceptional classes of words for stress. If, for example, proper names typically have predictable stress contours that violate the English productive 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. More generally, this strategy of filtering out non-productive data in the input could be very helpful, given the abundance of non-productive aspects in English. In the next section, we discuss a way children might implement this type of selective learning strategy on English input.

5.2.3 Selective learning

Yang (2005) and Legate and Yang (2013) describe a learning bias to filter out non-productive data, implemented formally via the Tolerance Principle. This principle is used to estimate how many exceptions a rule can tolerate before it's no longer useful for the learner to represent 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.

One way children might use this bias when learning metrical stress 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 productive stress contour while the others are exceptions. Children could then apply the Tolerance Principle when considering a 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 productive 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*). Other data for that syllabic word form (e.g., *awáy*, *úh óh*) would be ignored when trying to learn the productive grammar. The other option is that no contour is productive according to the Tolerance Principle, and so all the data for that syllabic word form are ignored when trying to learn the grammar.[15,16] It may be that this bias to filter out non-productive data helps the child perceive the input in a way that causes the English grammar in each prodKR to be compatible with a larger proportion of the relevant data. This could then surmount the apparent learnability problem for English in each prodKR.

Table 9 shows the learnability results for a learner using the Tolerance Principle to filter the acquisitional intake to contain only those stress patterns that seem productive. Notably, whether or not the learner has knowledge about the interaction with morphology, the learner's intake is considerably less than what's available in the input (1188-2331 word types from an original 4780 types).

Table 9: Learnability analyses for the three prodKR theories (HV, Hayes, and OT) for a learner using the Tolerance Principle to decide whether input data are productive and therefore useful to learn from. The quantity of word types in the acquisitional intake is shown for each prodKR. The four metrics shown are learnability potential of the prodKR (prodKR: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 in English child-directed speech. Scores are calculated both for a learner with knowledge of the interaction of inflectional morphology (+infl) and for a learner without that knowledge (–infl).

		Intake	**prodKR:Pot**	**Eng:Raw**	**Eng:Rel**	**Eng:RelClass**
+infl	**HV**	1188	0.930	0.858	0.603	0.514
	Hayes	1125	0.896	0.858	0.829	0.832
	OT	1816	0.860	0.521	0.762	0.520
–infl	**HV**	1637	0.949	0.874	0.622	0.484
	Hayes	1664	0.933	0.699	0.680	0.626
	OT	2331	0.843	0.634	0.798	0.708

One promising result is that the learnability potential for all three prodKR theories skyrockets – all three prodKRs have a grammar that will account for at least 85% of the productive data (0.843–0.949). This suggests that the linguistic variables of these prodKR theories are very useful for capturing the productive data in English. However, the coverage of the English grammar in each prodKR is still less than this potential, which indicates the English grammar is again not the most useful grammar to learn for these data.

For a learner with inflectional knowledge, the HV and Hayes English grammars are far more useful than the OT English grammar (capturing 85.8% of the data, rather than 52.1%). However, the Hayes English grammar is much more learnable (Hayes: 0.829 relative compatibility, 0.832 relative class compatibility). Thus, the Hayes English grammar is the best of the

[15] See Appendix B for a more detailed example of using the Tolerance Principle to implement this bias to filter out non-productive data.

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

three from a learnability standpoint. Still, there are over a hundred other grammars in the Hayes KR that are more learnable (i.e., with higher relative and relative class compatibility), which means the learnability problem persists.

For a learner who does not yet have inflectional knowledge, the HV English grammar is clearly the most useful grammar, accounting for 87.4% of the data intake. Yet, the HV English grammar is actually the least learnable of the three prodKR theories (0.622 relative compatibility, 0.484 relative class compatibility). For this learner, the OT English grammar is the most learnable (OT: 0.798 relative compatibility, 0.708 relative class compatibility), despite actually accounting for the smallest portion of the acquisitional intake (OT: 0.634 raw compatibility). Nonetheless, the learnability problem again persists, as those relative compatibilities indicate that over 73,000 other OT grammars are more compatible with the English data.

In summary, the Hayes English grammar is the most learnable for a selective learner with inflectional knowledge, while the OT English grammar is most learnable for a selective learner without this knowledge. Yet, because there are still many grammars defined by these prodKR theories that are better able to account for even this more filtered data set, it is worth considering another alternative for solving the learnability problem: a different target grammar for English.

5.3 Different productive grammars for English

What alternative productive grammars should we consider for English in each prodKR? One idea is to look at the grammars within each prodKR that are more compatible with the English child-directed speech 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 grammar definition in each prodKR.

5.3.1 Updating the HV English grammar

For the HV prodKR, 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: many use QI 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 this change to the HV English definition yields similar effects, whether the learner has knowledge of the inflectional morphology interaction or not, as shown in Table 10.

Changing only the quantity sensitivity parameter value boosts the raw compatibility of the English grammar from accounting for about 60% of the data (+infl=0.61, –infl=0.59) to accounting for 63–64% (+infl=0.63, –infl=0.64). From the learnability standpoint, this change boosts the relative compatibility so the English grammar is better than approximately 95 out of 100 of the other grammars in the hypothesis space (relative comp: +infl=0.96, –infl=0.94; relative class comp: +infl=0.95, –infl=0.95).

Interestingly, this beneficial effect is mitigated if the learner is using a selective learning strategy like the Tolerance Principle, though the effects are again similar whether the learner has knowledge of the inflectional morphology interaction or not. There is still a very beneficial impact in grammar coverage: making this change to the quantity sensitivity parameter value boosts the raw compatibility of the alternative English grammar to 86–88% (+infl=0.86, –infl=0.88). However, the relative compatibility of the alternative English grammar is not as good, accounting for fewer data than a large proportion of the grammars in the hypothesis space (relative comp: +infl=0.72, –infl=0.71; relative class comp: +infl=0.60, –infl=0.53).

So, the HV alternative English productive grammar is only a real improvement if the learner is attempting to account for all the input, rather than the productive subset of it. From a developmental standpoint, this means this alternative grammar would only be easier to learn in the initial stages of acquisition before the child has realized there are non-productive components generating a portion of the English data. Given this, this alternative grammar could be a transition grammar children identify early on before they derive additional knowledge about English.

5.3.2 Updating the Hayes English grammar

Turning to the Hayes prodKR, we find 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 productive 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 at 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, as shown in Table 10.

Changing only the foot inventory parameter value boosts the raw compatibility of the English grammar from accounting for around half of the data (+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).

Unlike the HV English grammar, the updated Hayes English grammar remains this learnable even for a learner using a selective learning strategy like the Tolerance Principle. As before, the effects are similar whether the learner has knowledge of the inflectional morphology interaction or not. Updating the Hayes English grammar this way allows it to account for around 90% of the productive data (+infl=0.87, –infl=0.93), and boosts the relative compatibility so that the alternative English grammar is still better than at least 9 out of 10 of the other grammars in the hypothesis space (relative comp: +infl=0.91, –infl=0.96; relative class comp: +infl=0.93, –infl=0.97).

So, the Hayes alternative English productive grammar is an improvement whether the learner is attempting to account for all the input or only the productive subset of it. From a developmental standpoint, this means this alternative grammar would be useful in both the initial stages of acquisition and later stages when the child has realized there are non-productive components generating a portion of the English data. It therefore may be a viable alternative target productive grammar for the end state of English acquisition.

5.3.3 Updating the OT English grammar

Turning to the OT prodKR, 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).[17] The current definition of the English grammar has the opposite ranking (Wsp-vv higher than 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, as shown in Table 10.

Flipping only this ranking boosts the raw compatibility of the English grammar from accounting for around 57% of the data (+infl=0.575, –infl=0.570) to accounting for nearly two thirds of the data (+infl=0.649, –infl=0.650). From the learnability standpoint, this change boosts the relative compatibility so the English grammar is better than nearly 9 out of 10 of the other grammars in the hypothesis space, and often better than nearly all other grammars (relative comp: +infl=0.982, –infl=0.988; relative class comp: +infl=0.886, –infl=0.974).

Interestingly, the updated OT English grammar's performance varies somewhat for a learner using a selective learning strategy like the Tolerance Principle. In particular, a learner who does not yet have inflectional morphology knowledge fares much better than a learner who does. Without inflectional morphology knowledge, the alternative OT English grammar can account for 84% of the productive data (0.840), and is more learnable than nearly all other grammars (relative comp: 0.985, relative class comp: 0.973). However, once the learner has inflectional morphology knowledge, data coverage drops to 71% (0.706) and while the relative compatibility is still high (0.933), the relative class compatibility is much lower (0.752), suggesting that the exact learning process could determine how learnable this updated English grammar is. For example, it would matter whether the learner is simply comparing grammars, which relative compatibility captures, or instead identifying grammars that have high raw compatibilities, which the relative class compatibility captures.

So, the OT alternative English productive grammar seems most useful for a learner not using a selective learning strategy, or for a selective learner who does not yet have inflectional morphology knowledge. From a developmental standpoint, this suggests this alternative grammar may be a transition grammar that children identify before they realize there are non-productive components in English or before they recognize the interaction between metrical stress and inflectional morphology.

[17] Interestingly, this ranking may not only be more useful for the implementation of OT we investigate here, but also corresponds to the English grammar identified for these data by the Maximum Entropy approach of Hayes and Wilson (2008), an alternative OT implementation that evaluates candidates according to their cumulative weighted constraint violations. See Appendix C for more discussion.

5.3.4 Summary: Alternative English grammar definitions

Table 10: Learnability analyses for the three alternative prodKR English grammars, given knowledge of inflectional morphology (+/–infl) and a selective bias to learn from data perceived as productive (+/–sel learn). The three metrics shown are raw compatibility of the (best) English grammar (Raw), relative compatibility of the (best) English grammar (Rel), and relative class compatibility (RelClass) of the (best) English grammar, which are computed over word types in English child-directed speech.

		+sel learn			–sel learn		
		Raw	**Rel**	**RelClass**	**Raw**	**Rel**	**RelClass**
+infl	**HV: QI**	0.86	0.72	0.60	0.63	0.96	0.95
	Hayes: Tro-Syl	0.87	0.91	0.93	0.64	0.87	0.90
	OT: NonFin>>WSP-VV	0.71	0.93	0.75	0.65	0.98	0.89
–infl	**HV: QI**	0.88	0.71	0.53	0.64	0.94	0.95
	Hayes: Tro-Syl	0.93	0.96	0.97	0.64	0.91	0.90
	OT: NonFin>>WSP-VV	0.84	0.99	0.97	0.65	0.99	0.97

To sum up, there are options for updating the definition of the English productive grammar for all three prodKR theories that significantly aid learnability, as shown in Table 10. The alternative Hayes English grammar appears to be the most consistently useful, as it improves whether the learner is using a selective learning strategy or not, and whether the learner has inflectional morphology knowledge or not. Nonetheless, the alternative OT English grammar is also useful for many learning scenarios, and the alternative HV English grammar is useful for a learner attempting to account for all the input, rather than filtering the acquisitional intake to a productive subset.

This motivates us to examine whether these alternative English grammars can similarly account for English adult metrical stress knowledge better than the current English grammar definitions. If it turns out that they can, then these definitions are worth further theoretical consideration as the target English productive grammars. If instead it turns out that these updated grammars are not as good at accounting for adult productive knowledge, they could instead represent transitory grammars children identify during acquisition, as discussed above. It would then be useful to determine experimentally if children converge on these alternative grammars at some point during acquisition, before moving on to the true target grammar for English.

6 General discussion

In this paper, we proposed new quantitative evaluation metrics for theories of representation, based on the natural link between linguistic representation and linguistic acquisition. We then demonstrated these metrics' use by comparing three theories of metrical stress representation meant to capture the productive component of the grammar (the prodKRs). The first key finding is that the prodKRs themselves are more or less equivalently good when it comes to their ability to define a productive grammar that is useful for English acquisition input, with the best prodKR depending on the exact learning assumptions involved. This didn't have to be true *a priori*, as each prodKR was designed with compact cross-linguistic coverage in mind rather than acquisition utility. Moreover, it's a welcome finding that all are equivalently *good*, and not merely equivalent – whether a productive grammar can account for $\frac{2}{3}$ or $\frac{9}{10}$ of the acquisitional intake, that portion is non-trivial and so such a grammar seems useful for a learner to acquire.

But how then do we interpret this equivalence? Recall that these prodKRs do not appear to be obvious notational variants of each other, and so are not transparently the same underlying representation. One option is that they are, in fact, equivalent at some level of computational analysis, and we simply have yet to discover the appropriate translation from one to another. If this is true, learnability considerations will never be able to distinguish them because they should not actually be distinguished. Another option is that they are, in fact, truly different and learnability considerations could distinguish between them, but a more sophisticated approach is needed that incorporates additional aspects of the developmental trajectory. Both options offer interesting opportunities for future research by theorists (option 1) and cognitive modelers (option 2).

The second key finding involves articulating how acquisition needs to proceed in English for a learner using each prodKR option. By using these evaluation metrics, we have empirically motivated proposals for the learning assumptions that need to be true for an English child to identify the current definition of the English grammar in each prodKR. For example, if children are not filtering out non-productive data, it is difficult to learn any of the prodKR English grammars. When children are filtering out the non-productive data, knowledge of the interaction between inflectional morphology and metrical stress is much more helpful for the Hayes English grammar than the HV or OT English grammars. This is therefore a useful contribution to the acquisition theory that accompanies each representational theory.

The third key finding is that there may be alternative versions of the English productive grammars in each prodKR that, while minimally different from the existing definitions, are much more learnable from English child-directed speech. These alternative grammars were identified by examining the prodKR grammars that were most learnable according to our evaluation metrics and seeing how they differ from the current English grammar definitions. Notably, whether these alternative grammars are very learnable depends on the stage of acquisition the child is at because children's acquisitional intake changes as they derive linguistic knowledge over time. All three alternative English grammars are much more learnable at an earlier stage of development before non-productive data are filtered out. The OT and Hayes alternative English grammars are also much more learnable after the non-productive data are filtered out but before knowledge of the impact of inflectional morphology on stress is recognized. So, these could represent transitory grammars that can be identified in children via experimental methods. Additionally, the Hayes alternative English grammar continues to be more learnable after the impact of inflectional morphology is recognized, and so could also represent a viable target grammar for English – or at least a later stage of acquisition if it is not the target grammar. These results highlight the contribution that acquisition considerations can make to theories of representation, as well as predictions that can be investigated experimentally.

More generally, when choosing among prodKR theories, it matters what each prodKR needs to satisfy learnability. 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 at the appropriate stage of development. 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 productive knowledge representations (prodKRs), based on the learnability of their language-specific grammars from the data children typically learn from. This computational analysis represents the first step for making an argument from acquisition for any prodKR theory. If and when we find that it is possible for a prodKR and its language-specific grammar 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 prodKR from typical child-directed 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 prodKR theory.

Here, we have used this approach to investigate three prodKR theories of metrical stress, evaluating them on their ability to make the target English productive grammar easily learnable from an English child's acquisitional intake at different developmental stages. English is an excellent test case for metrical stress learnability, as it contains many non-productive aspects and therefore represents a difficult acquisition scenario. So, if a prodKR allows a child to successfully acquire the English productive grammar, that prodKR truly is useful for acquisition.

While we found that all three prodKR theories define useful productive grammars for English, we also found that the current definitions of the English grammar in each one were, in fact, *not* those grammars. This led us to propose potential changes to the way acquisition must proceed for an English child using each prodKR and potential changes to the definition of the target grammars for English within existing prodKR theories. Thus, this computational approach allows us to suggest useful updates to both the theories about how acquisition occurs in this domain and the theories about how knowledge in this domain is represented.

Acknowledgements

We are especially grateful to Joanna Lee for her help in preparing the child-directed speech corpus and to Jeff Lidz for numerous suggestions and general faith in this enterprise. We have additionally benefited from comments and suggestions from Jeff Heinz, Pranav Anand, Matt Wagers, Sandy Chung, Armin Mester, Jim McCloskey, Adrian Brasoveanu, Mits Ota, several anonymous reviewers, 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, the Berkeley Linguistics Society 2014 meeting, and the GALANA 2015 special session on learning in generative grammar. All errors are of course our own and not at all their fault.

8 References

Dana Angluin. Inductive inference of formal languages from positive data. *Information and Control*, 45(2):117–135, 1980.

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.

R Baayen, R Piepenbrock, and L Gulikers. CELEX2 LDC96L14. `https://catalog.ldc.upenn.edu/LDC96L14`, 1995.

Nan Bernstein Ratner. Patterns of vowel modification in mother–child speech. *Journal of Child Language*, 11(3):557–578, 1984.

Amy Booth and Sandra Waxman. Mapping words to the world in infancy: On the evolution of expectations for nouns and adjectives. *Journal of Cognition and Development*, 4(3): 357–381, 2003.

Michael R Brent and Jeffrey Mark Siskind. The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2):B33–B44, 2001.

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.

Courtney B Cazden. The acquisition of noun and verb inflections. *Child Development*, 39: 433–448, 1968.

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.

Robin Clark. The selection of syntactic knowledge. *Language Acquisition*, 2(2):83–149, 1992.

Robin Clark and Ian Roberts. A computational model of language learnability and language change. *Linguistic Inquiry*, 24(2):299–345, 1993.

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.

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.

Edward Gibson and Kenneth Wexler. Triggers. *Linguistic Inquiry*, 25(4):407–454, 1994.

Lila R Gleitman, Elissa L Newport, and Henry Gleitman. The current status of the motherese hypothesis. *Journal of Child Language*, 11(1):43–79, 1984.

Mark E. Gold. Language identification in the limit. *Information and Control*, 10(5):447–474, 1967.

Sharon Goldwater, Thomas Griffiths, and Mark Johnson. A Bayesian Framework for Word Segmentation: Exploring the Effects of Context. *Cognition*, 112(1):21–54, 2009.

Peter D Grünwald. *The minimum description length principle*. MIT press, 2007.

Morris Halle and Michael Kenstowicz. The Free Element Condition and cyclic versus noncyclic stress. *Linguistic Inquiry*, 22:457–501, 1991.

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.

Bruce Hayes and Colin Wilson. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry*, 39(3):379–440, 2008.

Jeffrey Heinz. Computational Theories of Learning and Developmental Psycholinguistics. In Jeffrey Lidz, William Synder, and Joe Pater, editors, *The Cambridge Handbook of Developmental Linguistics.* Cambridge University Press, 2014. To appear.

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 Kingsbury, Stephanie Strassel, Cynthia McLemore, and Robert MacIntyre. CALLHOME American English Lexicon (PRONLEX). https://catalog.ldc.upenn.edu/LDC97L20, 1997.

Paul Kiparsky. Metrical structure assignment is cyclical. *Linguistic Inquiry*, 10(4):421–441, 1979.

Christine Kitamura and Denis Burnham. Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, 4(1):85–110, 2003.

Alix Kowalski and Charles Yang. Verb Islands in Child and Adult Language. In Alia K. Biller, Esther Y. Chung, and Amelia E. Kimball, editors, *BUCLD 36: Proceedings of the 36th annual Boston University Conference on Child Language Development*, pages 281–289. Cascadilla Press, Somerville, MA, 2012.

Stan A Kuczaj. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600, 1977.

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.

Jeffrey Lidz and Annie Gagliardi. How Nature Meets Nurture: Universal Grammar and Statistical Learning. *Annual Review of Linguistics*, 1(1):333–352, 2015.

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

Virginia A Marchman, Kim Plunkett, and Judith Goodman. Overregularization in English plural and past tense inflectional morphology: A response to Marcus (1995). *Journal of Child Language*, 24(03):767–779, 1997.

Gary F Marcus, Steven Pinker, Michael Ullman, Michelle Hollander, T John Rosen, Fei Xu, and Harald Clahsen. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57(4):i–178, 1992.

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.

Partha Niyogi and Robert C. Berwick. A language learning model or finite parameter spaces. *Cognition*, 61:161–193, 1996.

Daniel N. Osherson, Michael Stob, and Scott Weinstein. *Systems that learn: An introduction to learning theory for cognitive and computer scientists*. The MIT Press, Cambridge, MA, 1986.

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, 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.

Lawrence Phillips and Lisa Pearl. The utility of cognitive plausibility in language acquisition modeling: Evidence from word segmentation. *Cognitive Science*, doi:10.1111/cogs.12217, 2015.

Steven Pinker. Formal models of language learning. *Cognition*, 7:217–283, 1979.

Alan Prince and Paul Smolensky. *Optimality Theory: Constraint Interaction in Generative Grammar*. ROA, New Brunswick, NJ, 2002.

Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

William Sakas and Janet Fodor. Disambiguating Syntactic Triggers. *Language Acquisition*, 19(2):83–143, 2012.

William Sakas and Janet D. Fodor. The Structural Triggers Learner. In Stefano Bertolo, editor, *Language Acquisition and Learnability*, pages 172–233. Cambridge University Press, Cambridge, UK, 2001.

Elaine Stotko. Investigating children's use of English derivational morphology. *First Language*, 14(42–43):348–349, 1994.

Bruce Tesar and Paul Smolensky. Learnability in optimality theory (short version). *ROA. Johns Hopkins University*, 1996.

Bruce Tesar and Paul Smolensky. Learnability in optimality theory. *Linguistic Inquiry*, 29 (2):229–268, 1998.

Bruce Tesar and Paul Smolensky. *Learnability and optimality theory*. The MIT Press, Boston, MA, 2000.

Malathi Thothathiri and Jesse Snedeker. Give and take: Syntactic priming during spoken language comprehension. *Cognition*, 108(1):51–68, 2008a.

Malathi Thothathiri and Jesse Snedeker. Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language*, 58(2):188–213, 2008b.

Andrea Tyler and William Nagy. The acquisition of english derivational morphology. *Journal of Memory and Language*, 28(6):649–667, 1989.

Vladimir Vapnik. Principles of risk minimization for learning theory. In *Advances in Neural Information Processing Systems*, pages 831–838, 1992.

Vladimir Vapnik. *The nature of statistical learning theory.* Springer Science & Business Media, 2013.

Michael Wilson. MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, 20(1):6–10, 1988.

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 stress has different effects for each prodKR theory. Before this knowledge is available, the parametric English productive grammars are able to account for different subsets of the ten most frequent stressed syllabic word forms, as shown in Table 11.

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 11: Ten most frequent stressed word forms (by type) in the parametric prodKR intake. 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 English productive grammar can account for the stressed syllabic word form.

One striking difference is the ability of the HV English productive 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 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 contain inflectional morphology. Thus, stripping off inflectional morphology does not help the HV 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 grammar (e.g., *highest* (LC: '10') and *biggest* (XC:'10')). This is why there is little increase in overall compatibility for the HV grammar (0.593 to 0.605). 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 productive grammar.

The Hayes English productive grammar, in contrast, is unable to account for five of the most frequent stressed syllabic word forms, some of which contain inflectional morphology, like *highest* (LA: '10') and *biggest* (XA: '10'). Ignoring inflectional morphology clearly is helpful, as the raw compatibility of the Hayes grammar goes from 0.485 to 0.550. This increase occurs because the Hayes 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 productive grammar, we find similar behavior to the HV grammar (Table 12). Adding knowledge of inflectional morphology doesn't seem to help raw compatibility much (from 0.570 to 0.575). When we examine the OT grammar's ability to account for the most frequent stressed syllabic word forms (see Table 12), we see a similar pattern to the HV grammar: the OT 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 contain 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 stress.

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 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 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 OT 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 OT grammar still can't account for (see Table 12). Thus, the addition of inflectional morphology knowledge does not obviously aid the OT English productive 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 productive subset of metrical stress data to learn from,

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 12: Ten most frequent stressed word forms (by type) in the constraint-based prodKR intake. 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 productive grammar instantiations can account for the stressed syllabic word form. A proportion indicates how many of the 26 grammar instantiations can account for the stressed syllabic word form.

and briefly show its impact for each prodKR theory.

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 prodKR theories, 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 which 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 13 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 prodKR theories' English grammars, since this means the learner should ignore all the XL syllabic word form data when trying to learn the English productive grammar. The HV grammar could not account for the '01' and '11' stress contours, which comprise 181 word types of the input. The Hayes grammar could not account for the '10' and '11' stress contours, which comprise 344 word types of the input. Thus, these

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 13: The Tolerance Principle applying to the stress contours associated with the XL syllabic word form for parametric and constraint-based representations. 1 = stressed syllable and 0 = unstressed syllable.

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 prodKR theory, 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 13 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 prodKR theories' English grammars, this turns out to be harmful for the OT grammar. This is because most of the OT 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 productive 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 productive grammar, and would likely make it more difficult to learn from English child-directed speech input.

C Maximum Entropy OT

Based on compatibility with English child-directed speech data, we proposed an update to the OT English productive grammar definition that would ameliorate many learnability issues: NonFin ranked higher than Wsp-vv. This updated ranking in fact corresponds to the ranking in the English grammar identified for the English child-directed speech data by the Maximum Entropy (**MaxEnt**) approach of Hayes and Wilson (2008), an alterative OT implementation that evaluates candidates according to their cumulative weighted constraint violations. This contrasts with an implementation that selects the candidate that violates the fewest important constraints, which is the traditional OT implementation.

Under the MaxEnt approach, a grammar is a set of weighted constraints (e.g., NonFin = 0.969, Wsp-vv = 0.701, ...), and software is available implementing the MaxEnt algorithm[18] to identify the best set of constraint weightings for a given data set. While weights are real values, any given set of weightings can be mapped to a strict ordering of constraints (e.g., the weights above correspond to NonFin being ranked higher than Wsp-vv), such as the grammars we evaluated here for the OT prodKR. So, we applied the MaxEnt approach to the English child-directed speech input to identify the best set of constraint weights for that data set, which can then be translated to a constraint ordering (Table 14).

Constraint	Tro	FtBin	Align-L	NonFin	Wsp-vv	Wsp-vc	Parse-σ Align-R *SonNuc
Weight	1.317	1.263	1.058	0.969	0.701	0.048	0.000

Table 14: Constraint weights discovered by the MaxEnt algorithm for the English child-directed speech data from the Brent corpus.

Notably, the set of weights discovered corresponded to a somewhat different grammar than the OT English definition we used here, which is perhaps not unexpected given the very different way the MaxEnt OT implementation evaluates candidates. Nonetheless, there were several similarities between the MaxEnt-identified English productive grammar and the English grammar definition derived from Hammond (1999) and Pater (2000): (i) Tro remains the most important constraint, (ii) Wsp-vv is higher than Wsp-vc, and (iii) both parse-σ and *SonNuc are unimportant. Most importantly, the same relative ordering we suggest here for NonFin and Wsp-vv (i.e., NonFin is higher than Wsp-vv) is also the ordering of those constraints for the MaxEnt English productive grammar.

[18] At *http://www.linguistics.ucla.edu/people/hayes/MaxentGrammarTool/*.