

Implicational markedness and frequency in constraint-based computational models of phonological learning*

GAJA JAROSZ

Yale University

*(Received 22 December 2008 – Revised 15 August 2009 – Accepted 31 January 2010 –
First published online 22 March 2010)*

ABSTRACT

This study examines the interacting roles of implicational markedness and frequency from the joint perspectives of formal linguistic theory, phonological acquisition and computational modeling. The hypothesis that child grammars are rankings of universal constraints, as in Optimality Theory (Prince & Smolensky, 1993/2004), that learning involves a gradual transition from an unmarked initial state to the target grammar, and that order of acquisition is guided by frequency, along the lines of Levelt, Schiller & Levelt (2000), is investigated. The study reviews empirical findings on syllable structure acquisition in Dutch, German, French and English, and presents novel findings on Polish. These comparisons reveal that, to the extent allowed by implicational markedness universals, frequency covaries with acquisition order across languages. From the computational perspective, the paper shows that interacting roles of markedness and frequency in a class of constraint-based phonological learning models embody this hypothesis, and their predictions are illustrated via computational simulation.

INTRODUCTION

It has been observed that the same structures that are cross-linguistically rare or MARKED are also the structures that are acquired later by children (Jakobson, 1941/1968; Stampe, 1969). In Optimality Theory (OT; Prince

[*] I would like to thank the editors, the guest editor Brian MacWhinney and an anonymous reviewer for their helpful comments, and especially Paul Boersma for his extensive review. Many thanks to Richard Weist for digitizing and sharing the audio-recordings of the Polish CHILDES data, and to Yvan Rose for providing the software and technical support to help with transcription of the data. The development of this work has also benefited by comments from Joe Pater, Karen Jesney, Kathryn Flack, Adam Albright and audiences at SUNY, NYU and the First Northeast Computational Phonology Meeting, where portions of this work were presented. Address for correspondence: Department of Linguistics, Yale University, 370 Temple St., Room 204, P.O. Box 208366, New Haven, CT 06520-8366, USA. Email: gaja.jarosz@yale.edu

& Smolensky, 1993/2004), the relative ranking of universal markedness constraints that penalize marked output configurations and faithfulness constraints that penalize disparity between underlying and surface representations determines the set of allowable surface structures in particular languages, and by permutation, in languages cross-linguistically. If the set of constraints is universal, as is often assumed in the OT literature, then the simplest possible hypothesis about language acquisition is that child grammars and adult grammars are both rankings of the same universal constraints. To explain the relative unmarkedness of child grammars as well as the developmental progression from unmarked to marked, it has been proposed that all markedness constraints are initially ranked above all faithfulness constraints (M»F; Gnanadesikan, 1995/2004; Smolensky, 1996).

The primary focus of this paper is on a particular extension of this hypothesis which maintains a primary role for universal markedness but also assumes a secondary role for frequency – the FREQUENCY HYPOTHESIS – along the lines of Levelt and van de Vijver (1998/2004) and Levelt, Schiller & Levelt (2000). The paper extends the empirical support for the frequency hypothesis from the existing findings on Dutch syllable structure acquisition to four new languages: English, German, French and Polish. The discussion integrates recent findings on the acquisition of syllable structure in English, German and French with novel empirical findings on the acquisition of consonant clusters in Polish. Comparison of acquisition orders with frequencies of syllable types in child-directed speech in these languages reveals that acquisition order covaries with relative frequency, supporting the frequency hypothesis.

As Boersma & Levelt (2000) showed via computer simulations of Dutch syllable structure acquisition, the Gradual Learning Algorithm for Stochastic OT (GLA; Boersma, 1998) embodies exactly the interaction of universal markedness and frequency of the frequency hypothesis. In addition to presenting the predictions of the GLA for Polish and English, the present paper discusses two other learning algorithms, which, by virtue of their sensitivity to frequency during learning, also embody the frequency hypothesis. These three learning models are presented and the way in which their various learning strategies embody the frequency hypothesis is explained. The predictions of these models are exemplified by computer simulations of syllable structure learning, and the predicted learning paths for three languages given input data representative of child-directed speech are shown to correspond to the attested and distinct developmental orders in these languages.

IMPLICATIONAL MARKEDNESS AND THE FREQUENCY HYPOTHESIS

This section briefly reviews Optimality Theory, with emphasis on the role of implicational markedness that it embodies. It then reviews the frequency

hypothesis, focusing on the concrete predictions the hypothesis makes in the domain of basic syllable structure.

Implicational markedness in Optimality Theoretic grammars

Before discussing the frequency hypothesis and its predictions for language acquisition, it is necessary to understand the formal system of OT on which the hypothesis depends. Of particular importance is the role of universal markedness and its implicational structure in OT. It is from this formalization of markedness that the predictions about acquisition order follow.

Optimality Theory formalizes grammars as rankings of universal constraints. A fundamental goal for research within OT is to identify a set of universal constraints that, upon permutation, predict the set of possible (empirically attested) adult languages. Thus, it is inherently a typological theory, and the presence of a constraint is motivated by the cross-linguistic predictions it makes by its interaction with other constraints. Given the universal constraint set, the only permissible systematic difference between languages is the ranking of these constraints. While the universality of constraints is often equated with innateness, it has been proposed that a universal constraint set, or at least part of it, could itself be learned from universally shared experience (Flack, 2007; Hayes, 1999). Whether constraints are innate or acquired is orthogonal to the present discussion. What is crucial for the frequency hypothesis is that constraints be universal and available to the child by the time grammatical development begins.

The predictions of an Optimality Theoretic grammar depend on the set of constraints, and therefore if predictions of the theory are not confirmed, it is always necessary to consider whether the constraint set is to blame. In order to avoid this problem as much as possible, the empirical focus of the present work is in the domain of simple syllable structure for which the predictions of a standard set of constraints have extensive typological support (Blevins, 1995). The set of standard syllable structure constraints that will be used throughout the paper is the same as in Levelt & van de Vijver (1998/2004) and Boersma & Levelt (2000) and is shown in (1). The first four constraints are markedness constraints, which penalize output configurations. The final constraint is a standard faithfulness constraint that penalizes the deletion of underlying material.

- (1) Simple syllable structure constraints:
 - a. ONSET – No vowel-initial syllables.
 - b. NOCODA – No consonant-final syllables.
 - c. *COMPLEXONSET – No syllable-initial consonant clusters.
 - d. *COMPLEXCODA – No syllable-final consonant clusters.
 - e. MAX – No deletion.

Different rankings of these constraints predict different subsets of basic syllable shapes to be permissible. Not all rankings characterize distinct syllable type inventories, however; whether a syllable type is permissible depends only on the relative ranking of the markedness constraints that it violates and the faithfulness constraint. As long as MAX dominates the relevant markedness constraints, the syllable type will be permissible. In light of the diverse views of markedness assumed in the language acquisition literature, the exact definition of markedness characterized by OT grammars warrants a brief discussion. In contrast to some views of markedness as cross-linguistic frequency or structural complexity (Demuth & McCullough, to appear), the type of markedness embodied in OT is IMPLICATIONAL MARKEDNESS, defined in (2).

(2) Implicational markedness:

Given two surface structures A and B, A is MORE MARKED than B iff:

- i. Every language that permits A also permits B.
- ii. There exist languages that permit B and do not permit A.

Thus, in OT, markedness is determined by implicational relations between surface structures cross-linguistically. It is not sufficient for a structure to be infrequent cross-linguistically or to be represented using relatively complex structure to be considered marked. Furthermore, although markedness is defined as a relation between two structures, it is possible to talk about a structure being marked without reference to another structure. In this special case, the presence of this structure is considered marked relative to the absence of this structure. For example, saying that syllable codas are marked means that syllables with codas are more marked than syllables without codas. Implicational markedness follows directly from the structure of the theory and its inherent typological character: the presence of a markedness constraint M penalizing a structure A predicts (at least) two possible languages, one that ranks only a relevant faithfulness constraint above M and therefore permits A, and another that ranks M above all faithfulness constraints and therefore prohibits A. Crucially, whenever a ranking, such as the one with faithfulness high, permits A, it also permits structures without A, thereby establishing the implication.

For any constraint set it is possible to compute the implicational markedness relationships it embodies. In fact, there is software available for doing just this (Anttila & Andrus, 2006). The implicational markedness structure can be represented as a directed graph, in which higher nodes are more marked and imply (point to) lower nodes. Doing this for the syllable structure example results in the graph shown in Figure 1. Here the permissible syllable types are represented in terms of simple consonant-vowel (CV) sequences. If a language permits a structure denoted by a node in the graph, then the language also permits all structures represented by

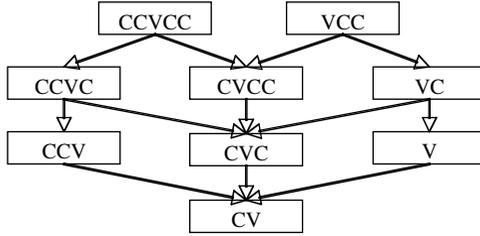


Fig. 1. Implicational markedness relations.

the nodes that are pointed to by that node. For example, the graph for the syllable type constraints shows that any language that permits VC syllable types also permits the less marked V, CVC and CV syllable types. Conversely, if no path along directed edges exists between two nodes, there is no implicational markedness relationship between them, and languages may permit just one type and not the other. For example, no edges connect types with complex onsets such as CCV and types with complex codas such as CVCC. This correctly predicts that there should be languages that have complex onsets but not complex codas, such as Spanish, and languages that have complex codas but not complex onsets, such as Finnish.

The implicational markedness graph captures information about possible languages: a language can be thought of as a subset of the nodes of the graph. A language is permissible according to the implicational structure of the graph if and only if all nodes pointed to by the selected nodes are themselves selected. For example, a language represented by the set {CVC, V, CV} is possible, while the language {CVC, VC, CV} is not since one of its members, VC, points to a node, V, not included in the set. Understanding the implicational markedness predictions embodied in a constraint set is crucial for the development of OT theories since these predictions must be tested against cross-linguistic generalizations. As is shown next, these graphs also make the predictions of the frequency hypothesis for a particular constraint set explicit and transparent.

For the sake of clarity and continuity with previous work, this paper exemplifies the predictions of the hypothesis using the standard constraint set defined above. However, it is important to note that the substantive predictive content of the theory depends only on the implicational relations between the surface forms depicted in the above graph, and this graph is neutral with respect to the kinds of structures used to represent these sequences. Therefore, any alternative representational assumptions and appropriately restated constraints that capture these implicational relations will make the same predictions. To be concrete, although the discussion throughout assumes final clusters are syllabified as complex codas and initial

clusters as complex onsets, this assumption has little substantive, predictive consequence. The same predictions would follow from different representational assumptions as long as they encode the same implicational relations. Furthermore, as argued above, these implicational relations have extensive typological support and, as a result, even theories assuming drastically different representations will generally seek to capture them. In sum, the predictions follow from the implicational relations encoded by the set of constraints, not directly from the constraints and representations they assume.

Before reviewing the frequency hypothesis, one final note is needed. Much recent work has explored the effects of articulatory and morphological factors on phonological development (Kirk & Demuth, 2005; Zydorowicz, 2007; see Demuth (in press) for a review). Even though the predictions of the frequency hypothesis are discussed here in terms of implicational markedness, it is important to note that this includes many morphological and articulatory factors. From the beginning, research in Optimality Theory has been concerned with functional grounding of universal constraints, and many standard constraints have articulatory or perceptual motivations. Universal functional pressures, formalized as constraints, are predicted to have an effect under the frequency hypothesis. The same goes for morphological factors. The interaction of morphology and phonology plays a prominent role in research in OT, and due to the presence of constraints that relate phonological and morphological structures, morphology is also predicted to have an effect on acquisition under the frequency hypothesis. Thus, although the present discussion focuses on the relative markedness of various syllable types, implicational markedness applies equally well to lower-level articulatory and perceptual factors as well as to the interaction of phonology with morphology.

The frequency hypothesis

In order to explain the restricted set of acquisition orders observed in Dutch, Levelt *et al.* (2000) and Levelt & van de Vijver (1998/2004) proposed that when universal markedness is silent with respect to the relative order of acquisition of two structures; the one with higher production frequency in the adult language is acquired first. This proposal, which was also examined in Boersma & Levelt (2000), will be referred to here as the frequency hypothesis. Earlier work indicating a causal role of frequency include Ingram's (1988) findings that order of acquisition of vowel-initial words across languages depends on the frequency of these forms in the ambient language. The assumptions of the frequency hypothesis are summarized in (3) below. Assumption (3)a is inherited from Optimality Theory, which assumes a set of universal constraints and permutation to

explain cross-linguistic variation. As a consequence of continuity and the implicational markedness inherent in OT, implicational markedness universals must be valid at every point during acquisition. This prediction, which is further discussed below, means acquisition order cannot conflict with implicational markedness universals. The next two assumptions, (3)b and (3)c, are motivated by empirical generalizations about the nature of child language acquisition. The initial M»F bias (3)b captures the relative unmarkedness of early grammars (Gnanadesikan, 1995/2004; Smolensky, 1996). Assumption (3)c reflects the uncontroversial assumption that learning is gradual, that grammatical development can be represented as a gradual progression from the initial M»F ranking to the adult ranking via a series of intermediate rankings. Assumption (3)d identifies a secondary role for frequency, along the lines of Levelt & van de Vijver (1998/2004) and Levelt *et al.* (2000). The effect of frequency is secondary to that of markedness: only when no implicational markedness relationship exists between two structures does higher frequency favor earlier acquisition. The final assumption (3)e is provided for completeness: any proposal calling for the role of additional factors, systematic restrictions on the set of attested acquisition orders, is a rejection of the frequency hypothesis.

- (3) Assumptions of the frequency hypothesis:
- a. CONTINUITY: Child grammars and adult grammars are formalized as rankings of the same set of universal markedness and faithfulness constraints.
 - b. M»F BIAS: Initial child grammars can be represented by a ranking with all markedness constraints above all faithfulness constraints.
 - c. GRADUALNESS: grammatical development proceeds from the initial state via a series of intermediate rankings on the way to the target ranking.
 - d. SECONDARY ROLE OF FREQUENCY: When markedness does not determine the relative acquisition order of two structures, the higher frequency structure is acquired earlier.
 - e. TOTALITY: No other factors systematically affect grammatical development.

The predictions of the frequency hypothesis for acquisition of syllable structure are discussed by Levelt & van de Vijver (1998/2004) and Levelt *et al.* (2000) and are reviewed here. In the basic syllable structure system, an initial state with all markedness constraints ranked above all faithfulness constraints corresponds to a ranking of {ONSET, NoCODA, *COMPLEXONSET, *COMPLEXCODA} » MAX. Since the markedness constraints do not conflict with one another and there is only one faithfulness constraint, all rankings compatible with this restriction admit only the maximally unmarked CV syllable type.

Thus, the predicted initial state consists of CV syllables only, which corresponds to the bottommost node of the implicational markedness graph in Figure 1. Subsequent acquisition can also be described in terms of the graph. In particular, acquisition begins in the bottommost node and gradually proceeds to the target language. Intermediate stages must be permissible languages according to the depicted implicational markedness relations. An intermediate stage is legal if the set of syllable types it admits does not entail (point to) any syllable types that are not included. For example, a possible acquisition path for Klamath, which allows the syllable types CV, CVC and CVCC (Blevins, 1995), begins at CV, then adds CVC, and finally adds CVCC. A path in which complex codas are acquired before simple codas is not possible, however, since this path would include an intermediate stage in which complex codas but not simple codas are admitted, which is a language not permitted by the implicational markedness universals. Thus, a learning path in which A is acquired before B is possible only if A is not more marked than B. Put another way, acquisition order is predicted to follow implicational markedness: orders in which the less marked structure is acquired first are possible, whereas orders where the more marked structure is acquired first are not.

Finally, when implicational markedness does not determine a relative acquisition order between two structures, the frequency hypothesis predicts the structure with the higher frequency will be acquired first. Since there is no implicational relationship between complex onsets and complex codas, for example, the frequency hypothesis predicts that in languages that admit both structures their relative order of acquisition will depend on their relative frequency in the ambient language. Thus, if the relative frequency of the same two (equally marked) structures differs across languages, the frequency hypothesis predicts their order of acquisition should likewise vary. The effect of frequency is secondary to that of markedness, however; the frequency hypothesis predicts that earlier acquisition of a more marked structure is not possible, even if its frequency is much higher in the adult language. In sum, the frequency hypothesis predicts a primary role for universal, implicational markedness and a limited effect of language-specific frequency in cases where markedness is silent.

In probabilistic extensions of Optimality Theory (e.g. Stochastic OT: Boersma, 1998), the effect of frequency is mediated by the set of universal constraints. Specifically, frequency of a surface configuration is relevant to the extent that constraints referencing different aspects of that configuration exist and are active in the grammar. The same holds of the frequency hypothesis. In the present example, there are just four markedness constraints, and it is the frequencies of the structures these constraints reference that can affect acquisition order. In a more complex example, each surface configuration would be subject to markedness constraints at various

TABLE 1. *Relative frequencies of syllable types in Dutch*

CV	CVC	CVCC	V	VC	VCC	CCV	CCVC	CCVCC
44.8%	32.1%	3.3%	3.9%	12.0%	0.4%	1.4%	2.0%	0.3%

levels of representation. For example, a complex onset like [st] would be evaluated by constraints on sonority sequencing, sonority distance, voicing agreement, place and voice licensing, not to mention various constraints at the segmental level and many others. In all cases, however, the present set of constraints would still be active and any additional constraints would still be stated over phonological classes at various levels of representation. Thus, it is the frequency of configurations of phonological classes at cross-cutting levels of representation and their interaction that drive order of acquisition under the frequency hypothesis. Clearly, this results in a complex system – the present paper explores in depth the cross-linguistic predictions of the frequency hypothesis at the level of basic syllable structure. This level is complex enough that various intricacies of the interaction of markedness and frequency can be illustrated yet simple enough that the predictions of the hypothesis can be firmly evaluated against recent findings on attested acquisition orders in a number of languages.

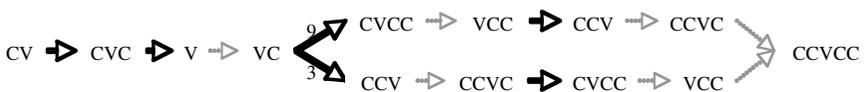
To see what the frequency hypothesis predicts for the acquisition of Dutch syllable types, consider the distribution of syllable types found in Dutch child-directed speech shown in Table 1. This data reflects the frequencies of occurrence of the nine syllable types in primary stressed syllables in a corpus of child-directed speech (Boersma & Levelt, 2000). Levelt *et al.* and Levelt & van de Vijver showed that given this distribution and the restrictions imposed by universal markedness, there are only two possible orders of acquisition for the marked structures coda, empty onset, complex onset and complex coda. The frequency hypothesis predicts that the first structure to be acquired is the unmarked CV syllable type. Review of the implicational markedness graph in Figure 1 reveals that markedness determines the relative order of acquisition between codas and complex codas (codas are less marked than complex codas), but is silent on the relative order for the remaining marked structures. This is where frequency comes in. Inspection of the distribution reveals that a total of 50.1% of the syllables in child-directed speech have codas, 16.3% lack onsets, 4% have complex codas and 3.7% have complex onsets. Levelt & van de Vijver showed that the minute difference in frequency between complex onsets and complex codas is not statistically significant, and therefore, for the purposes of the frequency hypothesis, these two marked structures may be considered equally frequent. Thus, given the restriction that CV must come first and that complex codas must come after singleton codas, there are three candidates

for which of the marked structures should be acquired first: codas, complex onsets or empty onsets. The frequency hypothesis states that the most frequent of these, codas, should come first. The structure predicted to be acquired next is the most frequent of the remaining marked structures, that is, onsetless syllables. Finally, there is a choice between complex onsets and complex codas: since these are equally frequent, the frequency hypothesis predicts both orders should be possible. In sum, the frequency hypothesis predicts the relative orders below:

- (4) Predicted acquisition orders for Dutch (Levelt & van de Vijver, 1998/2004):
- a. unmarked CV → coda → empty onset → complex coda → complex onset
 - b. unmarked CV → coda → empty onset → complex onset → complex coda

These are indeed the two orders found by Levelt *et al.* The two developmental orders identified among the twelve Dutch-speaking children are shown in (5) below. All arrows in the diagram correspond to transitions between developmental stages identified by Levelt *et al.* The larger, black, arrows denote transitions between stages corresponding to the predicted stages in (4), while the smaller, gray, arrows indicate additional order of acquisition differences observed in the data. Nine of the children acquired complex codas before complex onsets, and three showed the reverse pattern. Comparison of the predicted orders to the attested orders reveals that all the predicted relative orders are empirically supported. The frequency hypothesis has correctly restricted the number of predicted orders to the two that are in fact observed. Examining the distribution of syllable types in more detail, it is possible to observe the frequency hypothesis' correct predictions in three distinct situations. First, when markedness and frequency conflict, the frequency hypothesis predicts that markedness should determine relative order of acquisition. This is exactly the situation with V and VC syllable types: VC is more marked than V, but it occurs more than three times as often as V. The frequency hypothesis correctly predicts that the less marked V syllable type is acquired first despite its dramatically lower frequency. Second, if markedness doesn't determine order, then frequency can. It is on this basis that the relative order between codas, onsetless syllables and clusters was established above, and this again is a correct prediction. Finally, in the situation where neither markedness nor frequency favors a relative order, both orders are predicted to be possible. This prediction is supported as well since both orders of the equally marked, equally frequent, cluster types are observed.

- (5) Development of syllable types in Dutch (Levelt *et al.*, 2000):



There are additional order effects that the frequency hypothesis misses, however. While there are a number of possible responses to this observation, this paper will demonstrate in the following sections that some of these additional order effects are in fact expected when gradual learning is combined with frequency sensitivity and implicational markedness. Before turning to a systematic evaluation of the frequency hypothesis cross-linguistically, some issues relating to the frequency hypothesis raised by recent empirical findings are briefly discussed.

Other issues relating to the frequency hypothesis

While it is well known that children's initial productions are unmarked relative to the adult languages, it is not generally the case that all children's initial productions can be described by the SAME unmarked grammar. In particular, it is often observed that children's productions, despite their differences from the adult pronunciations, generally respect the phonotactic restrictions of the target language. For example, children learning Dutch, which has a phonotactic restriction against final voiced obstruents, do not produce word-final voiced obstruents (Zamuner, Kerkhoff & Fikkert, in prep.). Phonotactic restrictions are language-specific and are often conflicting: for example, some languages prohibit voiced obstruents altogether while others can require intervocalic consonants to be voiced. If children's initial productions obey phonotactic restrictions in the ambient language, then initial productions in different languages must be restricted in different ways. The frequency hypothesis, however, does not predict any relationship between initial productions and the phonotactics of the ambient language. Further work examining the relationship between initial production and phonotactic restrictions cross-linguistically is needed, but see Jarosz (2006) for a proposal of how phonotactic learning can result in an initial unmarked state that captures phonotactic restrictions.

As a consequence of continuity and factorial typology, every observed child grammar should correspond to a possible adult grammar. Put differently, child grammars should be describable in terms of rankings of constraints that are independently motivated by language typology – the constraints and interactions among constraints needed to describe adult grammars should be sufficient to also describe child grammars. However, there are attested processes and restrictions in child grammars that do not seem to have correspondents in adult grammars. For example, consonant harmony is frequently observed in child grammars, but similar processes involving major place features are not found in adult grammars (Pater, 1997; Smith, 1973). Such observations challenge the assumption of constraint universality, and to account for these facts many researchers assume that at least some constraints may be child-specific (Goad, 1998;

Pater, 1997; Pater & Werle, 2001). However, recent work by Fikkert & Levelt (2008) suggests that part of the explanation may reside in the structure of children's developing lexical representations. As shown by Zamuner *et al.* (in prep.), developing lexical representations may have a more general effect on production; much more work along these lines is needed to better understand the interaction of the grammar and lexicon and their shared development. In addition to child-specific processes such as consonant harmony, recent work argues that child-specific restrictions may be observed in intermediate stages of development and that intermediate stages often exhibit cumulative constraint interactions, some of which can be captured by adopting additive constraint interaction rather than ranking (Jesney & Tessier, to appear). As Pater (2009) shows, however, the kinds of cumulative effects possible even in weighted constraint grammars are highly restricted, and not all such interactions in child language can be straightforwardly captured via additive constraint interaction. The final section of the present paper shows how a kind of cumulative effect is expected as a natural consequence of gradual learning and frequency sensitivity.

As discussed above, acquisition of less marked structures can precede but not follow acquisition of more marked structures. On the whole, this prediction has much empirical support, but it is possible to find examples that seem to contradict it. One such example, which now has support from a number of acquisition studies in a number of languages, is the relative acquisition order of different coda consonants. According to well-known typological generalizations, more sonorous consonants are preferred to less sonorous consonants in coda position (Clements, 1990). However, a number of studies in various languages have found that obstruents are the first to appear in coda position (see, e.g., Fikkert (1994) on Dutch, Kehoe & Stoel Gammon (2001) on English, and Hilaire-Debove & Kehoe (2004) on French). There are a number of possible explanations of these findings that are compatible with the frequency hypothesis. For example, liquids are slow to develop in many languages regardless of position – it could be that the slow development of liquids in coda is a symptom of this, though this still leaves open questions about the slow development of other sonorants in coda. Alternatively, the structural development of the rhyme may provide an explanation. Perhaps the affinity of high sonority segments and coda position is tied to rhymal segments' ability to bear weight, and in initial stages children have not yet acquired heavy syllables (Fikkert, 1994). Note, however, that for this explanation to be compatible with the frequency hypothesis, such an intermediate grammar must be warranted by typology. Thus, the development of theoretical phonology and formal analysis of child language are inherently linked, and further work bridging developmental findings and typological generalizations will provide deeper understanding of both.

Given the concrete baseline provided by the frequency hypothesis, recent work has identified specific areas where empirical findings warrant further investigation. The continued interaction between researchers in formal linguistic theory and language acquisition is key to understanding the complex connections between child phonology and typology. The remainder of this paper focuses on evaluating the frequency hypothesis cross-linguistically and via computer simulation.

PREVIOUS WORK ON THE DEVELOPMENT OF WORD-INITIAL AND WORD-FINAL CLUSTERS

Previous work has demonstrated the ability of the frequency hypothesis to model acquisition order of syllable types in a single language, Dutch. Any theory of acquisition must of course be evaluated against empirical findings from many languages. Although the frequency hypothesis is consistent with the orders observed in Dutch, it is not clear that frequency is driving the order of acquisition. For example, given acquisition data from just one language, it is entirely possible that some universal bias explains the attested relative order of acquisition, and it happens to coincide with relative frequency in that language. In order to establish a robust correspondence between frequency and acquisition order, it must be shown that differences in relative frequency for the same structures covary with differences in acquisition order. Accordingly, this section reviews existing work on the acquisition order of syllable types cross-linguistically, focusing on the acquisition of consonant clusters, and shows that the frequency hypothesis is consistent with existing findings in all languages. The next section contributes to these cross-linguistic developmental findings by examining the acquisition order of consonant clusters in Polish.

To review, since no implicational markedness relation exists between complex onsets and complex codas, the frequency hypothesis predicts that the relatively more frequent structure will be acquired first. If the structures are equally frequent, then both orders are predicted to be possible. This is the case in Dutch: the relative frequencies of clusters of both types are around 4%, and the frequency hypothesis predicts both orders to be possible. This prediction is supported by developmental findings as discussed above. If frequency drives acquisition order, then a higher proportion of complex onsets should correspond to earlier acquisition of complex onsets, and a higher proportion of complex codas should correspond to earlier acquisition of complex codas.

Acquisition of consonant clusters in English and German

For English, the frequency hypothesis predicts that complex codas should be acquired first. This prediction follows from the relative frequency of

complex codas versus complex onsets in English child-directed speech. Kirk & Demuth (2005) analyzed the proportion of final versus initial clusters in child-directed speech in the Bernstein-Ratner (1982) and Brown (1973) corpora, which combined consisted of parental speech to twelve children, ages ranging between 1;1 and 4;10. Kirk and Demuth found that word-final clusters accounted for 67% and word-initial clusters accounted for 33% of all consonant clusters occurring at word edges. Thus, for English child-directed speech this study found a significantly higher proportion of complex codas, which according to the frequency hypothesis should correspond to earlier acquisition of complex codas.

The same study also found that English-speaking children's production is more accurate on final clusters than initial clusters. In this study, twelve children's (range 1;5 to 2;7) productions of monosyllabic words with initial and final clusters were elicited in a picture-identification task. Overall accuracy on final clusters was higher than accuracy on initial clusters. In addition, the authors show that accuracy on final clusters is significantly higher than accuracy on initial clusters matched for segmental material and sonority profile (final stop + [s] versus initial [s] + stop and final nasal + [z] versus initial [s] + nasal). While it is difficult to draw conclusions from these comparisons about the relative acquisition orders for individual children, Kirk and Demuth also present the proportion of children that produce each cluster type above a threshold of 75% accuracy. The most accurate final cluster (nasal + [z]) reaches this threshold for nine of the children, while the most accurate initial cluster type (stop + [l]) reaches this threshold for only four of the children. In an earlier study, Templin (1957) found that English-speaking children aged 3;0 and 3;6 produced word-final clusters more accurately than word-initial clusters. Thus, existing work on the acquisition of clusters in English identifies the predominant acquisition order as one with earlier acquisition of complex codas. This is consistent with the predictions of the frequency hypothesis.

For German, existing research suggests earlier acquisition of coda clusters as well (Lleo & Prinz, 1996). A corpus analysis of the proportion of word-initial as compared to word-final clusters in German child-directed speech reveals a significantly higher proportion of final clusters. To determine this, orthographically transcribed parental speech to twenty-two normally developing children, ages ranging between 1;6 and 3;6, in the Szagun corpus of German (Szagun, 2001) was extracted. The CELEX lexical database (Baayen, Piepenbrock & Gulikers, 1995) was used to determine whether words ended or began with bi-consonantal clusters. The analysis revealed that the ratio of final to initial clusters was approximately 70% to 30%.

In sum, developmental findings on the relative order of acquisition of clusters in German and English support the frequency hypothesis. In both languages, final clusters are more frequent in child-directed speech than

initial clusters, and research confirms the earlier acquisition of final clusters in both languages.

Acquisition of consonant clusters in French

English-learning children exhibit earlier acquisition of complex codas, and Dutch-learning children show variation. However, together these results can still be interpreted as showing an overall preference for earlier acquisition of complex codas since, even in Dutch, nine of the twelve children acquired complex codas before complex onsets. Thus, it remains to be shown that higher relative frequency of complex onsets in the ambient language coincides with earlier acquisition of complex onsets. Recent work on the acquisition of clusters in French addresses this question (Demuth & Kehoe, 2006; Demuth & McCullough, to appear).

In a picture-identification task with fourteen French-speaking children (age range 1;10 to 2;9), Demuth & Kehoe (2006) found higher production accuracy on initial obstruent-liquid clusters than final obstruent-liquid clusters. While this is consistent with the frequency hypothesis, the study examines only obstruent-liquid clusters in final position, and the late acquisition of these clusters can also be explained by implicational markedness. In a later, longitudinal study of two French-learning children (ages 1;5 to 3), Demuth & McCullough (to appear) examined the order of acquisition of three cluster types: initial obstruent-rhotic, final rhotic-obstruent and final obstruent-rhotic. The study found earlier acquisition of initial obstruent-rhotic clusters than either of the final clusters for both children. The same study also establishes that word-initial clusters are more frequent than word-final clusters in French child-directed speech. Specifically, the authors found that 70% of clusters occurring at word edges were initial clusters in the child-directed speech to two children (ages ranging from 1;0 to 2;6). This study only examines the acquisition of clusters with obstruents and rhotics, and it is unclear how these results extend to initial and final clusters more generally. For example, while obstruent-liquid clusters are cross-linguistically among the most preferred in initial position, it is generally accepted that more sonorous consonants are cross-linguistically preferred in coda position (Clements, 1990). Thus, it is possible that one of the unexamined final cluster types is acquired earliest of all the clusters.

Although some further examination of the acquisition of other cluster types is warranted, the existing findings suggesting earlier acquisition of initial clusters in French are consistent with the frequency hypothesis. In combination with the earlier research on the acquisition of clusters in the Germanic languages, the findings on acquisition in French provide direct cross-linguistic support for the role of frequency. Together these results

indicate that different relative proportions of initial to final clusters correspond to different acquisition orders.

DEVELOPMENT OF WORD-INITIAL AND WORD-FINAL CLUSTERS IN POLISH

The studies discussed above on the development of clusters in French provide much needed exploration of the predictions of the frequency hypothesis in languages with higher frequency of initial clusters. However, examination of the development of other types of final clusters is needed to rule out the possibility that an unexamined type of cluster develops earliest in final position. This section presents empirical findings on the acquisition of consonant clusters in Polish based on the examination of all types of word-initial and word-final clusters in spontaneous productions of four normally developing, Polish-learning children. As explained below, Polish, like French, exhibits a higher proportion of initial clusters, thereby providing an additional test case for the frequency hypothesis for which earlier acquisition of initial clusters is predicted.

Existing work on the acquisition of clusters in Polish includes an in-depth analysis of the various reductions exhibited in one child's production of target complex onsets (Łukaszewicz, 2007). This work does not compare the relative order of acquisition of initial clusters to final clusters, however. In another study of the productions of one child, Zydorowicz (2007) examines the reductions of clusters falling within morphemes compared to the reduction of clusters falling across morpheme boundaries. Interestingly, the author's findings suggest that reductions are less common for clusters falling across morpheme boundaries. However, this study does not provide measures of accuracy for initial or final clusters and does not discuss their relative order of acquisition.

Predictions of the frequency hypothesis

A corpus analysis of parental speech found in the Weist corpus of Polish, available in CHILDES (MacWhinney, 2000; Weist & Witkowska-Stadnik 1986; Weist, Wysocka, Witkowska-Stadnik, Buczowska & Konieczna, 1984), was performed. The orthographically transcribed child-directed speech in the corpus was automatically phonemicized based on standard pronunciation, which can be reliably extracted from the highly phonemic orthography. This resulted in a corpus of 34,122 words, of which 18.3% had bi-consonantal clusters at one or both edges. The frequencies of various bi-consonantal clusters by sonority profile are shown in Table 2, where the sonority levels are glide (G), liquid (L), nasal (N), fricative (F) and stop (S). Examination of all word-initial and word-final clusters reveals that 13.9% of all words

IMPLICATIONAL MARKEDNESS AND FREQUENCY

TABLE 2. *Bi-consonantal clusters in Polish adult speech by sonority profile*

Initial			Final		
Cluster	Total occurrences	Relative frequency	Cluster	Total occurrences	Relative frequency
FG	798	16.8	FS	1191	72.5
SL	786	16.6	NS	211	12.8
SS	710	15.0	LF	38	2.3
SF	669	14.1	SG	38	2.3
FS	640	13.5	SF	38	2.3
SG	310	6.5	LS	28	1.7
FL	284	6.0	GS	26	1.6
FF	218	4.6	NF	18	1.1
NG	162	3.4	SS	15	0.9
FN	120	2.5	GF	13	0.8
NN	25	0.5	SL	8	0.5
NL	13	0.3	FG	5	0.3
SN	8	0.2	LN	4	0.2
LF	2	0.0	LG	4	0.2
	4745		GN	2	0.1
			NN	2	0.1
			FN	1	0.1
			LL	1	0.1
				1496	

begin with clusters, whereas only 4.4% of words end in clusters. These relative frequencies correspond to a ratio of 76% to 24%, indicating that word-initial clusters are about three times as frequent as word-final clusters.

Thus, assuming that the proportion of initial to final clusters is representative of the proportion of complex onsets to complex codas children are exposed to in the ambient language, complex onsets are dramatically more frequent than complex codas in Polish child-directed speech. Based on this, the predictions of the frequency hypothesis for Polish are clear: initial clusters should be acquired earlier than final clusters.

Participants

The participants in this study are four normally developing Polish-speaking children from the Weist Corpus (Weist & Witkowska-Stadnik, 1986; Weist *et al.*, 1984). The children's ages range from 1;7 to 2;5. Audio-recordings of the sessions as well as orthographic transcriptions are publicly available via CHILDES (MacWhinney, 2000).

Because consonant clusters are just beginning to develop during this time period, and in order to avoid data sparseness problems, the files for sessions were combined into maximally four-month intervals separately for each

child. For convenience, these intervals will be referred to as stages. This resulted in one stage each for Marta (range 1;7-1;8), Kubus (range 2;1-2;4) and Wawrzon (range 2;2-2;5), and two stages for Bartosz (range 1;7-1;8 and 1;11).

Data transcription and coding

The children's speech in each of the audio-recordings was phonetically transcribed using broad phonemic transcription with the help of the ChildPhon software (Rose, 2003). In addition, the existing orthographic CHAT transcripts (Weist & Witkowska-Stadnik, 1986; Weist *et al.*, 1984) were used to identify the children's target words. Finally, the same procedure that was used to automatically translate orthographically transcribed adult speech to broad phonemic transcription was used to create initial phonetic transcriptions of the children's target words, and these phonetic transcriptions were then verified or modified (in a handful of cases) by a trained Polish-speaking transcriber.

All target bi-consonant clusters at word edges were coded according to the sonority of their constituent consonants. The children's productions were coded as correct if the child's production matched the sonority profile of the target cluster and incorrect otherwise; that is, substitutions within the target sonority level were not counted as errors. The same coding was repeated for all target words at a coarser level, grouping all consonants together. In this case the form was considered correct if it was produced as a cluster and incorrect otherwise.

All target cluster types were included in the analysis with the following exceptions. Although the standard pronunciation for the third person singular of the frequent verb *jest* 'to be' ends in a word-final cluster, the actual pronunciation of this word in adult speech is highly variable, with the final [t] or even the entire cluster often deleting. In order to avoid biasing the results, these target words were not included in the analysis. Additionally, although stop-fricative sequences and affricates may be contrastive in Polish (e.g. *trzy* 'three' vs. *czy* QUESTION PARTICLE), the acoustic differences between these two configurations are quite subtle, especially in final position. Therefore it is not clear how reliable the transcriptions are with respect to whether a particular production counts as one or two segments. To avoid this problem, affricates and homo-organic stop-fricative sequences were excluded from the analysis.

Overall accuracy

Results are presented first at the coarse cluster level. The proportion of clusters produced correctly as clusters overall in initial and final position is

TABLE 3. *Correct/total (percent) production of initial and final clusters in Polish*

	Marta 1;7-1;8	Kubus 2;1-2;4	Wawrzon 2;2-2;5	Bartosz 1;7-1;8	Bartosz 1;11
Initial	96/206 (47%)	150/202 (74%)	185/309 (60%)	37/191 (19%)	65/111 (59%)
Final	3/13 (23%)	11/30 (37%)	21/48 (44%)	0/19 (0%)	12/14 (86%)
Fishers's <i>p</i>	<i>p</i> = 0.15	<i>p</i> < 0.001	<i>p</i> < 0.05	<i>p</i> < 0.05	<i>p</i> = 0.078

shown separately for each child in Table 3. With the exception of Bartosz' second stage, the proportion of correctly produced initial clusters is numerically higher than the proportion of correctly produced final clusters for all children. Since the small expected value in a number of cases makes the Chi-square test inappropriate, Fisher's Exact test was used to determine whether the differences in proportions were significant. These results are also shown in the table and indicate that the differences in these proportions are significant in the cases of Kubus ($p < 0.001$), Wawrzon ($p < 0.05$) and the initial stage of Bartosz ($p < 0.05$). Marta's accuracy on initial clusters (47%) is substantially higher than on final clusters (23%) though this difference is not significant. Finally, Bartosz' accuracy on final clusters is numerically higher than on initial clusters in the second stage; however, this difference is not significant ($p = 0.078$). This apparent reversal is clarified when the clusters are broken down by sonority, as discussed next.

In sum, the children exhibit higher production accuracy on initial clusters than final clusters as a group, with all children showing a numerical preference for initial clusters at their earliest stage.

Accuracy by sonority profile

Overall accuracy on clusters by position provides a greater amount of data amenable to statistical analysis but is a crude measure. In particular, it is possible that clusters in final position are produced less accurately overall due to low production accuracy on one frequently attempted final cluster type. To determine whether this is the case, the accuracy of clusters in both positions was examined by sonority profile. Table 4 lists the number of times a cluster type was correctly produced out of the total number of times that cluster was a target, with a corresponding percent correct in parentheses. These proportions are provided for each type of cluster that was attempted at least three times by the child during that stage. The cluster types are presented in decreasing order of accuracy separately for initial and final clusters.

Upon inspection of Table 4, it is immediately clear that there are many more types of clusters produced in initial position than in final position for

TABLE 4. *Correct/total (percent) production of clusters by sonority in Polish*

	Marta 1;7-1;8		Kubus 2;1-2;4		Wawrzon 2;2-2;5		Bartosz 1;7-1;8		Bartosz 1;11	
Initial	SL	28/41 (68)	SL	46/52 (88)	SL	22/26 (85)	NG	3/5 (60)	FG	8/8 (100)
	SG	11/20 (55)	SG	7/9 (78)	SG	17/24 (71)	FG	6/13 (46)	SF	6/7 (86)
	SF	8/25 (32)	NL	3/4 (75)	FG	29/41 (71)	SS	5/15 (33)	NG	4/5 (80)
	FF	3/11 (27)	NG	3/4 (75)	SF	13/34 (38)	FL	1/6 (17)	FF	5/7 (71)
	FG	2/11 (18)	SF	24/35 (69)	FS	19/50 (38)	SG	1/10 (10)	SL	11/18 (61)
	FS	3/50 (6)	SS	5/8 (63)	FL	5/13 (38)	SL	4/55 (7)	FS	17/29 (59)
	SS	0/11 (0)	FG	9/15 (60)	FN	4/12 (33)	SF	1/16 (6)	SG	9/17 (53)
			FL	6/10 (60)	FF	2/8 (25)	FS	3/68 (4)	FN	1/4 (25)
			FS	28/47 (60)	SS	4/26 (15)	FF	0/3 (0)	FL	0/11 (0)
			FF	8/16 (50)					SS	0/7 (0)
Final	NS	3/10 (30)	NS	8/17 (47)	NS	11/19 (58)	NS	0/3 (0)	NS	3/3 (100)
			FS	2/7 (29)	FS	5/22 (23)	FS	0/16 (0)	SF	7/8 (88)
			LF	0/5 (0)						

all children. As shown in Table 2, in the parental input to these children, the number of bi-consonantal cluster types in both positions is comparable: in initial position there are fourteen types, while in final position there are eighteen. Thus, it is noteworthy that, regardless of accuracy, all children produced substantially fewer final clusters than initial clusters. To the extent that production of output structures is indicative of acquisition order, the number of initial cluster types produced alone suggests a preference for clusters in initial position. However, due to the small sample available for each type, not much can be made of the lack of attempts on cluster types that occur infrequently even in the parental speech. Therefore, the accuracy of productions relative to adult targets provides a more reliable measure.

Examination of the production accuracy of the cluster types further supports earlier acquisition of complex onsets. For all stages there are several initial cluster types produced at higher accuracies than the most accurate final cluster type. Specifically, Wawrzon produces initial SL, SG and FG more accurately than he produces final NS, his most accurate final cluster type, and the difference between initial SL (85%) and final NS (58%) is marginally significant (two-tailed Fisher's exact test; $p = 0.086$). For Kubus, all initial cluster types are produced more accurately than all final cluster types, and the difference between initial SL (88%) and the most accurate final cluster (NS; 47%) is highly significant (two-tailed Fisher's exact test; $p < 0.001$). Marta produces three initial cluster types (SL, SG, SF) more accurately than any final cluster type, and the difference between initial SL (68%) and her most accurate final cluster (NS; 30%) is significant (two-tailed Fisher's exact test; $p < 0.05$). Finally, Bartosz, in his earlier stage, produces no final clusters correctly while correctly producing eight initial

cluster types some of the time. Compared to the 0% accuracy on final FS, the proportions correct on initial NG ($p < 0.01$), FG ($p < 0.01$), and SS ($p < 0.05$) are significantly higher (two-tailed Fisher's exact test). Thus, breaking down the clusters by sonority indicates that the most accurate cluster types for all children occur in initial position.

The only exception is in Bartosz' second stage, where the most accurate types in both positions are equally accurate, suggesting that at this stage Bartosz may have already acquired some types in each position. The distribution of clusters and accuracies in Bartosz' second stage further illuminates the results discussed earlier at the level of clusters, where higher accuracy on final clusters was observed. Although overall Bartosz' accuracy on initial clusters (59%) is lower than on final clusters (86%) at this stage, breaking down production accuracy by sonority type reveals that the lower accuracy on initial clusters is a consequence of a broad range of accuracies on a large variety of target cluster types. It is the low accuracy of some of these initial cluster types that brings down the average for initial clusters overall. Since the accuracies of the most accurate types in initial and final position at this stage are comparable and close to 100%, there is no evidence that final clusters are preferred. Indeed, considering the higher accuracy on initial clusters in Bartosz' first stage together with the high accuracy on clusters in both positions in the second stage suggests that, even for Bartosz, an advantage for initial clusters can be ascertained in the overall developmental progression.

Discussion

In sum, examination of the production accuracies of initial and final clusters at two levels of granularity reveals a substantial preference for initial onset clusters. For each child a significant preference for initial onsets was established at one or both of these levels. These results not only indicate a preference for initial clusters overall, but a preference for initial clusters for each individual child. Thus, assuming the development of these children is representative of phonological acquisition of Polish in general, the findings suggest a developmental path in which complex onsets are acquired earlier than complex codas.

Certainly, an analysis indicating earlier acquisition of complex onsets in four children does not decisively establish a single acquisition order for Polish. Further work confirming these findings with additional children is needed. Nonetheless, at this point it is not premature to conclude that the predictions of the frequency hypothesis are consistent with these findings on the acquisition of clusters in Polish.

The results of all the acquisition studies together are consistent with the predictions of the frequency hypothesis and demonstrate that different

orders of acquisition coincide with different relative frequencies for the same two structures. It is important to keep in mind that the markedness considerations under investigation here are limited to basic syllable complexity. Further work is needed to determine to what extent alternative formulations of the markedness pressures, including lower-level segmental as well as morphological factors, are compatible with the existing evidence.

OPTIMALITY THEORETIC LEARNING MODELS COMPATIBLE WITH
THE FREQUENCY HYPOTHESIS

The discussion so far has focused on establishing the predictions of the frequency hypothesis and evaluating those predictions against cross-linguistic findings on acquisition order. The remainder of the paper demonstrates that a number of existing constraint-based computational models of language learning are naturally compatible with the frequency hypothesis. In this section, the learning models compatible with the frequency hypothesis are presented, and the mechanisms by which they capture the frequency hypothesis are discussed. The next section illustrates how the predictions already established above can be derived by computational simulation.

Although the models discussed below differ in a number of important ways, in the present context they can all be treated together due to a fundamental property they share, which makes them compatible with the frequency hypothesis. This property pertains to the way in which the learner's grammatical hypothesis is gradually adjusted in response to input from the ambient language. Although the exact mechanisms by which hypotheses are adjusted in these models vary, they all share the fundamental property that more frequent structures affect the learner's hypothesis more substantially and are therefore acquired more quickly. Moreover, given a universal set of constraints, these models inherit from OT the predictions regarding the role of implicational markedness in grammatical development. Thus, the models capture exactly the interaction of frequency and markedness in the frequency hypothesis.

Although these models maintain the predictions of the frequency hypothesis regarding the relationship of developmental grammars and typology, the formalization of grammars in each of these models generalizes the classic OT ranking in various ways. As a result, the models differ from one another and from classic OT in the kinds of grammars they predict to be possible final-state grammars cross-linguistically and, as a consequence of the assumptions of the frequency hypothesis, intermediate grammars in acquisition. The often subtle consequences of the different formulations of grammars across the models are a topic of considerable debate and ongoing investigation (Goldwater & Johnson, 2003; Jäger, to appear; Legendre, Sorace & Smolensky, 2006; Pater, 2009; Prince, 2002; Tesar, 2007).

However, the focus of the present paper is on a property the models all share, and the reader is referred to Pater (2009) for an overview of some of the models' differences. Additionally, as the following section explains, the predictions of the various models for the basic syllable type system considered here are qualitatively very similar.

Gradual Learning Algorithm for Stochastic OT

The Gradual Learning Algorithm (GLA; Boersma, 1998) assumes a probabilistic extension of OT's constraint ranking called Stochastic OT. In Stochastic OT, constraints are not strictly ranked on an ordinal scale. Rather, each constraint is associated with a mean RANKING VALUE along a continuous scale. Formally, each ranking value represents the mean of a normal distribution, and all constraints' distributions are assumed to have equal standard deviations, which are generally arbitrarily set to 2. At evaluation time, a SELECTION POINT is chosen independently from each of the constraints' distributions, and the numerical ordering of these selection points determines the total ordering of constraints, with higher numerical values corresponding to higher relative ranks. In this way, Stochastic OT defines a probability distribution over total orderings of constraints. The farther apart the ranking values of two constraints are, the higher the probability of a particular relative ranking between them. Conversely, when the ranking values for two constraints are close, each relative ranking has a good chance of being selected. This possibility enables Stochastic OT to model free variation: if two active constraints conflict, different rankings will correspond to different outputs being selected as optimal. This is the main typological consequence of Stochastic OT that differs from classic OT: it predicts that final-state grammars can be variable. In sum, Stochastic OT maintains OT's evaluation metric for choosing the optimal output form given a ranking; it differs by allowing a single grammar to vary stochastically among different total rankings.

The Gradual Learning Algorithm for Stochastic OT is ONLINE because it processes one surface form at a time. It is also ERROR-DRIVEN because it compares the actual surface form to the surface form generated by the learner's current grammatical hypothesis, and learning is triggered when the output generated by the learner does not match the observed output. In the case of a mismatch, the algorithm slightly decreases the ranking values of constraints that favor the loser and slightly increases the ranking values of constraints that favor the winner. All constraints are adjusted by the same amount, called the PLASTICITY. The basic insight is that, as learning continues, constraints favoring losers will gradually be pushed lower and lower until errors become diminishingly rare. The algorithm is not guaranteed to converge on a correct grammar, or any grammar for that matter, as shown

most concretely by Pater (2008). In practice, however, the algorithm usually performs quite well assuming it is given pairs of underlying forms and fully structured surface forms as learning data.

How does the GLA embody the frequency hypothesis? Each time the learner is presented with a configuration in the target language that its current grammatical hypothesis cannot generate, the learner makes a small adjustment to the grammar, making that configuration slightly more likely to be generated by the grammar. The more frequent that configuration is in the target language, the more often the learner will make small adjustments to the grammar, and the quicker the learner will get to a grammar that can generate that configuration. As a simple example, consider two marked structures, A and B, two markedness constraints, M_A and M_B , penalizing these two structures, and a faithfulness constraint F penalizing any unfaithful mapping (see Boersma & Levelt (2000) for similar discussion). In the initial state, both markedness constraints are ranked high and the faithfulness constraint is ranked low. In Stochastic OT, this initial state can be represented by assuming much higher ranking values for markedness constraints (e.g. 100) than for faithfulness constraints (e.g. 50). This initial state cannot generate A or B with any reasonable likelihood, so each time the learner processes either one, the grammar is adjusted. Learning proceeds until errors are no longer reliably made. If one of these marked structures (A) occurs more frequently in the data, it will be selected more often and therefore generate errors more often and lead to updates more often. The markedness constraint corresponding to it will move lower toward the faithfulness constraint more quickly. At a certain point, M_A 's ranking value will be close to F's ranking value, while the ranking value of M_B will still be substantially higher. At this point, the learner will start generating the marked structure A because some of the time the selection point for the faithfulness constraint will be higher than the selection point for M_A due to their proximity, resulting in M_A being generated faithfully. At the same time, M_B is still ranked significantly above F such that its faithful generation is much less likely. This intermediate grammar represents a point during learning when A has been (partially) acquired but B has not yet been produced. If the frequencies of A and B are dramatically different, there will be an intermediate grammar that more or less categorically admits A and does not admit B. If the difference in frequencies is not great, the effect will be subtler: there will be an intermediate stage where A will be generated more reliably than B, and A will reach adult-like accuracy before B. Finally, if the frequencies of A and B are very close, the acquisition order of A and B will likewise be close and, since the learning algorithm is non-deterministic, there will be variation across runs, with some runs resulting in slightly earlier learning of A and others with slightly earlier learning of B.

Although the mechanics of grammatical adjustments in response to training data are different in the different learning models, the impact of frequency on the predicted learning paths is essentially the same. The following discussion identifies a number of other models that exhibit the same response to input frequency and explains how the learning strategies reflect this frequency sensitivity.

Maximum Likelihood Learning of Lexicons and Grammars

Maximum Likelihood Learning of Lexicons and Grammars (MLG; Jarosz, 2006) treats constraint-based phonological learning as an optimization problem within the general framework of likelihood maximization. MLG deals with the full problem of learning both the grammar and the lexicon of underlying forms given unstructured surface forms. Learning is defined formally as the gradual optimization of a likelihood function whose domain is the hypothesis space of grammars and lexicons. MLG assumes a grammar is defined as a probability distribution over rankings, as in Stochastic OT. However, learning in MLG is not error-driven. Under gradual maximum likelihood optimization, the rankings of constraints in the hypothesized grammar are adjusted in proportion to how much work they do, or how much probability they assign to the surface forms in the data. Intuitively, maximum likelihood optimization rewards relative rankings of constraints that are able to generate the observed forms, and it rewards relative rankings in proportion to how much of the data they can generate. In this way, more frequent structures in the data lead to more substantial adjustments to the hypothesized grammar, which in turn leads to these structures being learned earlier.

Consider again the abstract example with two marked structures, A and B. Whenever a gradual maximum likelihood learner is exposed to A, it rewards the relative rankings that can generate A. In this simple example, this corresponds to rewarding the relative ranking of $F \gg M_A$. How much a relative ranking is rewarded depends on its frequency in the training data: the rankings favored by frequent structures are rewarded more than those favored by less frequent structures. Thus, if A is more frequent than B, $F \gg M_A$ will be rewarded more than $F \gg M_B$, and a grammar that generates A with some probability will be reached first. In sum, in MLG, learning is not triggered by errors but rather involves rewarding those relative rankings that make correct predictions. Nonetheless, MLG inherently encodes sensitivity to frequency that results in developmental paths that embody the frequency hypothesis. Although the GLA and MLG rely on different learning strategies, they both rely on probabilistic rankings of constraints, and their response to frequency is qualitatively the same (see Jarosz (2006)

for discussion of the important differences between the two learning theories).

Learning models for weighted constraint grammars

There are two main types of weighted constraint grammars differing in how the numerically weighted constraints are interpreted at evaluation time. Both evaluate competing output structures based on their relative HARMONY, which is the weighted sum of constraint violations. The weight of each constraint is multiplied by the number of violations it incurs (expressed as a negative integer), and the results are summed over all constraints. In Harmonic Grammar (HG; Legendre, Miyata & Smolensky, 1990a; 1990b; Smolensky & Legendre, 2006) and its close relatives, such as Linear OT (Keller, 2006), the optimal output form is determined directly from the harmony – the optimal output is defined as the output with highest harmony. In a probabilistic extension of HG, called noisy HG (Boersma & Pater, 2008), the weights of the constraints are selected from independent normal distributions at evaluation time, just as in Stochastic OT. The difference is that in Stochastic OT these numerical weights are interpreted as a strict ranking, whereas in noisy HG they correspond directly to the weights used in evaluation. Thus, noisy HG defines a probability distribution over weightings of constraints in the same way that Stochastic OT defines a probability distribution over rankings. This variation in weights/rankings determines the probability with which different output structures are selected as optimal. In Maximum Entropy (also called log-linear) models, which have recently been applied to phonological learning (Goldwater & Johnson, 2003; Jäger, to appear), the probability associated with an output structure is directly related to the harmony. Maximum Entropy models use a single weighting to define the probability with which different outputs are selected: specifically, the probability of an output is proportional to the exponential of its harmony. In sum, while the stochastic component in noisy HG resides in the weightings themselves being noisy, the stochastic component in Maximum Entropy models exists at the level of candidate output structures directly.

Abstracting somewhat from the differences in constraint interaction between the various models, the focus here is on how learning algorithms for these weighted constraint grammars exhibit a kind of frequency sensitivity that embodies the frequency hypothesis. The reasoning is identical to the reasoning above for the GLA: this is because the gradual learning algorithms for HG, Maximum Entropy models and Stochastic OT are fundamentally the same (Boersma & Pater, 2008). The algorithms for weighted grammars are both error-driven: when there is an error, weights of loser-preferring constraints are slightly decreased, and weights of the winner-preferring

constraints are slightly increased, just as in the GLA. The only difference between learning for Stochastic OT and weighted grammars is that the amount of change for the weight is proportional to the difference between the number of constraint violations it assigns to the winner and loser (Boersma & Pater, 2008; Jäger, to appear; Pater, 2008). This slight difference has little consequence for the algorithms' sensitivity to frequency in the training data, but it does have important formal consequences: the learning algorithms for HG and Maximum Entropy models are provably convergent on the correct target grammar given inputs paired with fully structured outputs (Boersma & Pater, 2008). Starting from the maximally unmarked grammar with all markedness constraints weighted well above faithfulness constraints, the grammar weights gradually change until errors are no longer produced. More frequent marked configurations result in more frequent errors, which in turn result in more frequent slight changes to the corresponding markedness constraints. The speed with which the weights of markedness constraints decrease determines the order in which the corresponding marked structures will be produced.

Summary

This section has introduced three classes of constraint-based learners: error-driven probabilistic ranking, likelihood maximization for probabilistic ranking and error-driven probabilistic weighting. Despite their distinct learning strategies, the learning models all embody the frequency hypothesis when paired with a set of universal constraints and an initial M » F grammar. The predictions of these models are explored via computational simulations in the next section.

SIMULATIONS

This section presents the results of simulations of the three types of learning models discussed above on data representative of child-directed speech in each of Dutch, English and Polish. The simulations with the GLA for Stochastic OT and the GLA for noisy HG, henceforth GLA-OT and GLA-HG, respectively, are carried out using the freely available Praat program (Boersma & Weenink, 2008) and follow the simulation set-up in Boersma and Levelt (2000), who have already presented results of the GLA-OT for Dutch. The Praat simulations employ the standard set of syllable structure constraints introduced above and rely on an initial ranking/weighting with all markedness constraints at 100 and the faithfulness constraint at 50 to capture the initial unmarked state. The noise, or standard deviation, is fixed to 2.0, and the plasticity is set to 0.1. The only difference between the simulations for different languages is the distribution of

syllable types, the training data, to which the learner is exposed. The inventories of all languages under investigation (Dutch, English and Polish) include all nine basic syllable types, but their relative frequencies in child-directed speech vary.

In all the Praat simulations, learning proceeds according to the steps in (6). The learner first samples a syllable type randomly from the distribution of syllable types in the target distribution. This sampling represents the fact that in child-directed speech, syllable types occur in an arbitrary order but together form a representative sample of the syllable types in the ambient language. With its current grammar, the learner generates an output using the syllable type's CV sequence as input and adjusts the grammar if there is a mismatch. Learning iterates in this fashion until the rate of errors has reached some prespecified threshold or the maximum number of iterations has been reached.

(6) Iterate:

- a. Randomly select a syllable type (TARGET FORM) according to the distribution of syllable types in the training data.
- b. Use the current stochastic grammar to generate an output (ACTUAL FORM) for that syllable type.
- c. If the actual form does not match the target form:
 - i. Increase the ranking/weighting value of each constraint that assigns more violation marks to the actual form than to the target form by 0.1
 - ii. Decrease the ranking/weighting value of each constraint that assigns fewer violations to the actual form than to the target form by 0.1

At any point during learning, it is possible to evaluate the current (noisy) grammar by using the grammar to generate the outputs for a large random sample from the target distribution. This provides a measure of accuracy for each of the syllable types. The gradual changes in accuracy on the various syllable types are used to model the acquisition order of the syllable types.

The simulations with MLG are performed using the software developed by Jarosz (2006), and the general procedure is reviewed here. Learning occurs via the Expectation-Maximization algorithm (Dempster, Laird & Rubin, 1977), which is summarized in (7). The algorithm first calculates the contribution of each ranking given the training data and the current grammar. The contribution of each ranking is simply the sum of its conditional probability given each data point, weighted by that data point's frequency. It is here that frequency plays a role: higher frequency training items carry more weight with respect to the calculation of a ranking's overall contribution. The algorithm then updates the grammar, setting the probability of each ranking in proportion to its relative contribution. Like the GLA algorithms,

updates to the grammar are gradual, making it possible to model acquisition paths. In contrast to the GLA algorithms, this algorithm runs in BATCH, processing all the training data before making an update to the grammar. This makes MLG somewhat less psychologically plausible, but see Jarosz (2006) for discussion of some of its advantages in settings where underlying representations and prosodic structure are unknown. In any case, the present focus is on the similarity between all these models in the way they respond to frequency.

(7) Iterate:

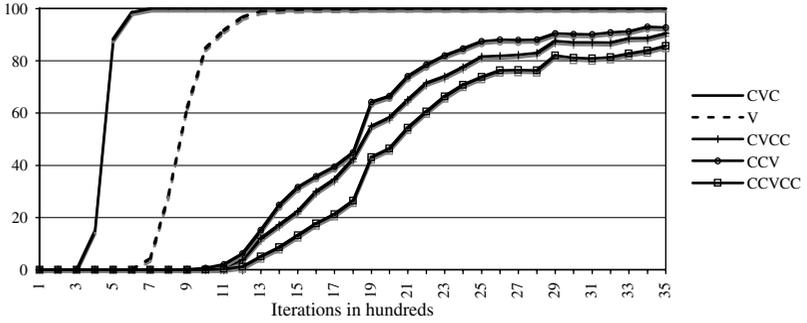
- a. Expectation Step: calculate the expected counts of each total ranking given the current grammar and the distribution of syllable types in the training data.
- b. Maximization Step: set the probability of each total ranking in proportion to its expected count.

Jarosz advocates an early stage of phonotactic learning that provides an initial state for the phonological learning modeled here and the learning of underlying representations. However, in an effort to make the MLG simulations as comparable to the Praat simulations as possible, the MLG simulations presented here assume a simple M » F initial bias. In particular, the initial state is set such that high ranking of markedness constraints is strongly favored with the probability of re-ranking being just 0.01. The results of the simulations for each model are discussed next for each language in turn, starting with Dutch.

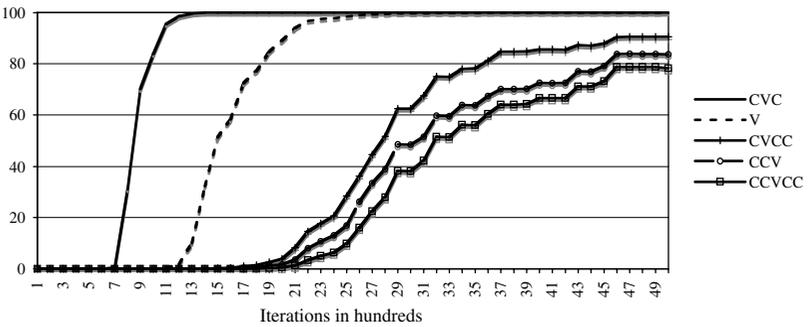
Dutch

Simulations using GLA-OT to model the learning of Dutch syllable types have already been reported on in previous work (Boersma & Levelt, 2000). For completeness, this section replicates the earlier simulations and presents the results of simulations with GLA-HG and MLG. For the Dutch simulations, the distribution of syllables discussed earlier and presented in Table 1 is used. Figure 2 shows a sample learning path, representing a predicted acquisition path for one child, for each of the three algorithms. The curve for CV is not shown since predicted accuracy for this syllable type is always 100% no matter what the constraint ranking is. The curves corresponding to the syllable types VC, VCC and CCVC are not shown because they are virtually identical to the curves for V, CVCC and CCV, respectively. Each curve shows how the accuracy of a syllable type changes over time, expressed in iterations for MLG and in hundreds of iterations for GLA-OT and GLA-HG.

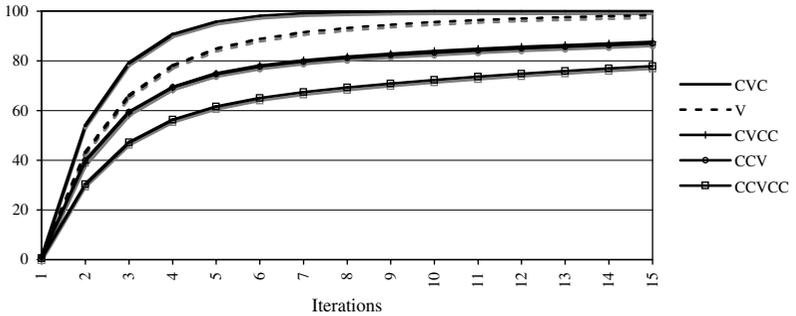
It is possible to use some threshold of accuracy to establish a predicted order of acquisition. In Figure 2a the first syllable to reach an 80% accuracy



(a) Sample learning path for GLA-OT



(b) Sample learning path for GLA-HG



(c) Learning path for MLG

Fig. 2. Dutch learning paths.

threshold (after CV) is CVC, then V, then CCV, then CVCC and finally CCVCC. Thus, for this run of GLA-OT, the predicted order is $CV \rightarrow CVC \rightarrow \{V, VC\} \rightarrow \{CCV, CCVC\} \rightarrow \{CVCC, VCC\} \rightarrow CCVCC$, where braces indicate simultaneous learning. The results of simulations with GLA-HG,

for which a representative simulation is shown in Figure 2b, are very similar. Finally, the Dutch simulation with MLG is shown in Figure 2c. Since MLG is a deterministic algorithm, it does not predict distinct outcomes on different runs. The frequency tie between complex onsets and complex codas therefore results in near simultaneous learning for the two structures. Otherwise, the predicted order of acquisition is the same as for GLA-OT and GLA-HG.

Due to the stochastic nature of the GLA algorithms and the nearly identical frequencies of complex codas and complex onsets in the Dutch distribution overall, different runs result in slightly different outcomes. If the simulation is repeated many times, some of the time complex onsets are acquired first and other times complex codas are acquired first. Running the simulation 10,000 times for 20,000 iterations (a point at which learning is essentially complete) reveals that 63.1% of the runs result in a slight preference for complex codas, 27.8% with slight preference for complex onsets, and 9% result in a tied ranking value for the two corresponding markedness constraints. Running GLA-HG 10,000 times results in similar proportions with 60.2% of the runs favoring complex codas, 30.2% of the runs favoring complex onsets, and 9.6% of the runs resulting in tied weights. This coincides well with the proportions reported by Boersma & Levelt (2000) and with Levelt *et al.*'s (2000) finding that nine out of twelve children exhibited this order. The acquisition orders predicted for Dutch by the three algorithms are summarized in (8), where the double arrow indicates variation.

- (8) Predicted orders of acquisition for Dutch:
- a. (GLAs) $CV \rightarrow CVC \rightarrow \{V, VC\} \rightarrow \{CCV, CCVC\} \leftrightarrow \{CVCC, VCC\} \rightarrow CCVCC$
 - b. (MLG) $CV \rightarrow CVC \rightarrow \{V, VC\} \rightarrow \{CCV, CCVC, CVCC, VCC\} \rightarrow CCVCC$

The predicted acquisition orders for MLG and the other two algorithms are essentially the same, showing that the response to frequency is qualitatively similar. However, examination of Figure 2 makes clear that the learning curves look quite different: the effect of frequency appears to be weaker in MLG. In MLG, the learning curves are all relatively close together, predicting that some learning of all syllable types happens simultaneously. In contrast, as the separation of the curves for CVC and V in the graphs for GLA-OT and GLA-HG reveals, these models predict the two syllable types should be acquired in sequence, with acquisition of CVC complete by the time acquisition of V begins. All three algorithms favor more frequent forms, but the different learning strategies have somewhat different effects given similar starting conditions. In particular, the GLA algorithms update ranking/weighting values in proportion to relative frequency, but ranking

values don't directly correspond to production probability. Recall that production is determined by independent normal distributions centered around the ranking/weighting values. When the markedness and faithfulness constraints get within a window of approximately two standard deviations of one another, large changes in production accuracy occur. In contrast, in MLG, the updates to the grammar are consistently proportional to the relative frequency, resulting in more gradual curves. Most acquisition work establishes acquisition orders by comparing production accuracy, and differences in production accuracy are consistent with both disjoint and overlapping curves. Thus, it is difficult to know whether attested acquisition orders correspond to truly disjoint learning curves as in the GLA algorithms or partially overlapping ones as in MLG. These interesting consequences of the learning strategies should be explored in future work.

Before moving on to the predictions for English, one remaining aspect of these simulations warrants further discussion. As discussed above, the syllable type CCVCC is learned last by all three algorithms, after learning of CCV and CVCC. What is particularly interesting about this prediction is that no ranking of these constraints can capture a language that admits CVCC and CCV but not CCVCC. If CCV is admitted, this means MAX ranks above *COMPLEXONSET. If CVCC is admitted, MAX must rank above *COMPLEXCODA. But the ranking with MAX above *COMPLEXONSET and *COMPLEXCODA also admits CCVCC. How can this be? This appears to be a cumulative constraint interaction, which ranking does not permit. The source of this emergent cumulativity, also discussed by Jäger & Rosenbach (2006), is the stochastic constraint ranking and the proximity of the rankings of these three constraints. To see where this cumulativity comes from, consider a simple Stochastic OT grammar where all three constraints have exactly the same ranking value. This means the probability of each of the six rankings of the three constraints is exactly one-sixth. Since in three of these rankings MAX ranks above *COMPLEXONSET, 50% of the time CCV is selected as optimal. The same goes for CVCC. The situation is different for CCVCC, however, because it incurs violations of both markedness constraints. CCVCC is selected as optimal only if MAX dominates BOTH markedness constraints, which happens in just two of the rankings. Thus, the accuracy of CCVCC is only one-third. The same logic applies when the ranking values are close but not identical: CCVCC surfaces faithfully only if both markedness constraints are dominated. Therefore, if there is a significant chance of either of the markedness constraints re-ranking relative to MAX, then CCVCC's accuracy will be lower than the accuracies of CCV and CVCC.

This observation leads to the intriguing possibility that some of the attested cumulative interactions in child language can be attributed to this kind of cumulativity. This possibility is supported by the fact that acquisition

orders are often established on the basis of differences in production accuracy. As noted, this STOCHASTIC CUMULATIVITY is possible only when the rankings of all three constraints are relatively close: probabilistic ranking cannot express CATEGORICAL CUMULATIVITY, where the singly marked structures are generated with perfect accuracy and the doubly marked structures are never generated. Weighted constraint grammars, like HG, are capable of expressing such interactions, but as Pater (2009) shows, this can only occur under certain conditions. HG cannot model cumulative effects with this particular constraint system because deletion of onset and coda consonants incurs separate violations of MAX, and therefore simplification of complex onsets and of complex codas is independent. As in the GLA-OT simulation, it is stochastic cumulativity that accounts for the cumulative effect seen in the GLA-HG simulation. Thus, if the attested cumulative interactions are categorical in nature, some additional mechanism is necessary to capture it. For a proposal along these lines, see Albright, Magri & Michaels (2008). Experimental work that can reliably compare the production accuracies of two structures will likely be needed to determine the extent to which cumulative effects in child language are stochastic or categorical.

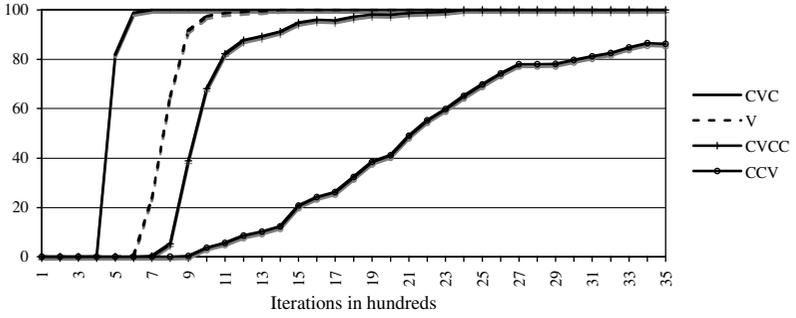
English

To illustrate the predictions for acquisition order in English, the relative frequencies of all syllable types were estimated from child-directed speech. Specifically, the frequencies of the various syllable types in primary-stressed monosyllabic words in the CHILDES Parental Corpus (MacWhinney, 2000; Li & Shirai, 2000) were extracted. The Parental Corpus combines parental speech to English-learning children across a large number of CHILDES corpora. The words were automatically transcribed using the CMU Pronouncing Dictionary (Weide, 1994). The resulting estimate of the relative proportions of all basic syllable types in English child-directed speech is shown in Table 5. These estimates confirm Kirk & Demuth's (2005) findings that complex codas are more frequent than complex onsets in English child-directed speech.

The result of the simulations using these frequencies, with settings otherwise identical to those for Dutch, is depicted in Figure 3. As before, the curves corresponding to the syllable types CV, VC, VCC and CCVC are not shown. Accuracy on CV is always perfect, while the curves for VC, VCC and CCVC are virtually identical to those for V, CVCC and CCV, respectively. Additionally, the syllable type CCVCC is not shown for the GLA simulations as its learning curves are virtually identical to those of CCV. As explained above, stochastic cumulativity is only possible when the rankings for all three constraints are close together. Since the relative frequency of complex codas is substantially higher than complex onsets,

TABLE 5. *Relative frequencies of syllable types in English*

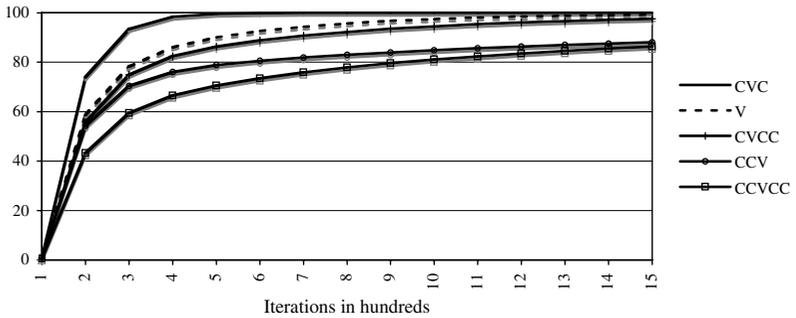
CV	CVC	CVCC	V	VC	VCC	CCV	CCVC	CCVCC
24.4%	40.5%	10.1%	4.7%	13.0%	3.5%	0.9%	2.2%	0.6%



(a) Sample learning path for GLA-OT



(b) Sample learning path for GLA-HG



(c) Learning path for MLG

Fig. 3. English learning paths.

complex codas are learned relatively quickly, and by the time complex onsets are developing the ranking of *COMPLEXCODA is too low to affect the accuracy of CCVCC relative to CCV. In MLG, however, because the learning curves are closer together, a cumulative effect is present, and the curve for CCVCC is shown.

Because the frequencies of equally marked structures are sufficiently distinct, different trials of GLA-OT and GLA-HG algorithms nearly always result in the same acquisition order, which is summarized in (9). Specifically, in both GLA-OT and GLA-HG onset-less syllables are acquired before complex codas in 99.9% of 10,000 identical runs, while the reverse order occurs in less than 0.1% of the runs.

- (9) Predicted order of acquisition for English:
- a. (GLAs) CV → CVC → {V, VC} → {CVCC, VCC} → {CCV, CCVC, CCVCC}
 - b. (MLG) CV → CVC → {V, VC} → {CVCC, VCC} → {CCV, CCVC} → CCVCC

The primary role of implicational markedness can be observed in these simulations. In Table 5, syllables with codas are overall more frequent than syllables without codas. If frequency were the only factor, it would predict earlier acquisition of CVC syllables than CV syllables. Under the frequency hypothesis, however, frequency's role is secondary to markedness: since all rankings that admit CVC syllables also admit CV syllables, it is impossible to model the earlier acquisition of CVC under the frequency hypothesis.

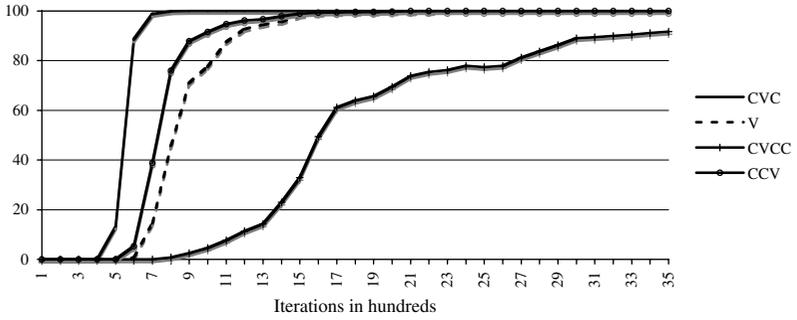
Polish

Finally, the relative frequencies of all syllable types in Polish child-directed speech were estimated based on the combined parental speech in the same corpus used above to establish the developmental order for Polish (Weist & Witkowska-Stadnik, 1986; Weist *et al.*, 1984). The orthographic transcriptions were automatically converted to a phonemic standard pronunciation as before. The proportions of initial and final consonant clusters of lengths 0, 1 and 2 were used to estimate the proportion of whole syllable types by assuming independent combination of onsets and codas. For example, the relative frequency of CVCC is the product of the probability of an initial C and the probability of final CC. Crucially, the resulting relative frequencies, shown in Table 6, reflect the fact that complex onsets are more frequent than complex codas in Polish.

The predicted learning curves for Polish are shown in Figure 4, and the corresponding predicted acquisition orders are summarized in (10). The predicted orders are complementary to that of English, with complex onsets developing earlier than complex codas. As in the English simulations, the

TABLE 6. *Relative frequencies of syllable types in Polish*

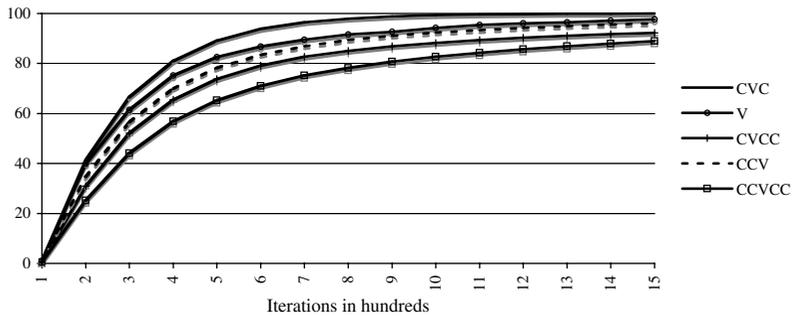
CV	CVC	CVCC	V	VC	VCC	CCV	CCVC	CCVCC
50.3%	20.9%	3.3%	8.5%	3.5%	0.6%	9.6%	4.0%	0.6%



(a) Sample learning path for GLA-OT



(b) Sample learning path for GLA-HG



(c) Learning path for MLG

Fig. 4. Polish learning paths.

substantial difference in frequencies between complex onsets and complex codas results in simultaneous learning of the second cluster (in this case CVCC) and CCVCC. In MLG, because of the closeness of the learning curves, CCVCC is learned later and this is shown in Figure 4. As before, the learning curves for VC, VCC and CCVC are virtually identical to those for V, CVCC and CCV, respectively, and these are not included in Figure 4. Finally, since the relative frequency of complex onsets overall is higher than the relative frequency of onsetless syllables, and because no implicational markedness relations exist between these two structures, the predicted order for Polish indicates a preferred order with complex onsets acquired earlier than onsetless syllables. However, because the relative frequencies are fairly close, the development of the two structures is predicted to be partially overlapping. Indeed, of 10,000 identical runs of GLA-OT, 79.7% slightly favored complex onsets and 16% slightly favored onsetless syllables, while in 4.3% of the runs the ranking values resulted in a tie. Likewise, for 10,000 runs of GLA-HG, the proportions favoring complex codas, favoring onsetless syllables and resulting in a tie were 77.9%, 17% and 5.1%, respectively. Further work on the development of Polish syllable structure is needed to test this prediction.

(10) Predicted orders of acquisition for Polish:

- a. (GLAs) $CV \rightarrow CVC \rightarrow \{CCV, CCVC\} \leftrightarrow \{V, VC\} \rightarrow \{CVCC, VCC, CCVCC\}$
- b. (MLG) $CV \rightarrow CVC \rightarrow \{CCV, CCVC\} \rightarrow \{V, VC\} \rightarrow \{CVCC, VCC\} \rightarrow CCVCC$

Discussion

This section has illustrated the predictions of the three learning models via computational simulation. In all cases, the predictions of the models correspond to the predictions of the frequency hypothesis as discussed above, which in turn correspond to attested acquisition orders for these languages. Additionally, it was shown that computational simulation sheds light on predictions that are otherwise hard to foresee and may help explain some of the discrepancies between intermediate child grammars and adult grammars. Even with this manageable constraint set, some of the complex interactions are difficult to anticipate. Further work is needed to examine the predictions of the frequency hypothesis at finer-grained levels, considering the joint effects of syllable structure, segmental content and morphological complexity, among others. Computational simulations such as these will undoubtedly be crucial to working out predictions for finer-grained, more complex systems with more interacting constraints.

Additionally, this paper has focused on a commonality of several existing constraint-based models of phonological learning and an empirical domain

in which differences between their predictions are minimal. Despite the effect of frequency common to all these models, they differ in important ways. The simulations revealed differences in the learning curves, rooted in the distinct learning strategies, that should be explored in future work, both computational and empirical. Also, the way each of these models generalizes classic OT is distinct and has consequences not explored here. Considerable progress has been made in recent work (Boersma & Pater, 2008; Goldwater & Johnson, 2003; Jäger, to appear; Jesney & Tessier, to appear; Legendre *et al.*, 2006; Pater, 2009; Prince, 2002; Tesar, 2007), yet further work comparing the predictions of these theories for typology, acquisition and learnability is essential.

CONCLUSION

This study examines the interacting roles of implicational markedness and frequency formally, empirically and computationally. From the perspective of formal linguistic theory, the paper discusses the interacting roles of universal markedness and language-specific frequency in making predictions for order of acquisition and phonological typology. From the empirical perspective, the paper reviews existing work on the acquisition of consonant clusters cross-linguistically and argues that the findings are consistent with the frequency hypothesis. The study also provides novel empirical support for the frequency hypothesis based on an analysis of the acquisition of consonant clusters by four Polish-learning children. The cross-linguistic findings in combination provide evidence that differences in relative frequency for the same structures correspond to differences in acquisition orders. Finally, from the computational perspective, the study examines the effect of frequency on the way grammatical hypotheses are gradually updated in three related computational models of phonological learning. Despite the differences in learning strategies and somewhat different formulations of constraint interaction, the models' response to frequency embodies the frequency hypothesis, and these predictions are illustrated via computational simulations for three languages with distinct distributions of syllable types.

Collaborative efforts connecting research in computational modeling, linguistic theory and typology, and formal analysis of acquisition result in deeper understanding of the formal and computational underpinnings of the system of language and its acquisition by children. The present work is an effort in this vein. This paper connects related work in formal linguistic theory and developmental findings on acquisition orders cross-linguistically with a class of learning models for constraint-based phonology. The paper has focused on a domain, basic syllable structure, for which the availability of existing work in all three disciplines makes the connection possible. The

present work examines the frequency hypothesis and shows that a class of learning models embodies this exact interaction of markedness and frequency. Much further work is needed, however. As discussed above, empirical findings supporting language-specific restrictions on early production and a divergence between child phonology and phonological typology challenge the frequency hypothesis. There is great potential for continued collaboration across these disciplines to lead to answers to these challenges and other outstanding questions.

REFERENCES

- Albright, Adam, Magri, Giorgio & Michaels, Jennifer (2008). Modeling doubly marked lags with a split additive model. In Harvey Chan, Heather Jacob & Enkeleida Kapia (eds), *BUCLD 32: Proceedings of the 32nd annual Boston University Conference on Language Development*, 36–47. Somerville, MA: Cascadilla Press.
- Anttila, A. & Andrus, C. (2006). *T-Order Generator*. Software package, Stanford University. Retrieved from www.stanford.edu/~anttila/research/software.html.
- Baayen, R. H., Piepenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database (Release 2)* [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [Distributor].
- Bernstein-Ratner, N. (1982). Acoustic study of mothers' speech to language-learning children: An analysis of vowel articulatory characteristics. Unpublished doctoral dissertation, Boston University.
- Blevins, J. (1995). The syllable in phonological theory. In J. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Cambridge, MA: Blackwell.
- Boersma, P. (1998). *Functional phonology: Formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.
- Boersma, P. & Levelt, C. (2000). Gradual constraint-ranking learning algorithm predicts acquisition order. In Eve V. Clark (ed.), *The proceedings of the thirtieth annual child language research forum*, 229–37. Stanford, CA: CSLI.
- Boersma, P. & Pater, J. (2008). Convergence properties of a Gradual Learning Algorithm for Harmonic Grammar. Unpublished ms, University of Amsterdam and University of Massachusetts, Amherst.
- Boersma, P. & Weenink, D. (2008). *Praat: Doing phonetics by computer* (Version 5.0.17) [Computer program]. Retrieved from www.praat.org/. Developed at the Institute of Phonetic Sciences, University of Amsterdam.
- Brown, R. (1973). *A first language: The early stage*. Cambridge, MA: Harvard University Press.
- Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (eds), *Papers in laboratory phonology I: Between the grammar and physics of speech*, 283–333. New York: Cambridge University Press.
- Dempster, A., Laird, M. & Rubin, D. (1977). Maximum Likelihood from incomplete data via the EM Algorithm. *Journal of Royal Statistics Society*, 39(B): 1–38.
- Demuth, K. (in press). The prosody of syllables, words and morphemes. In E. Bavin (ed.), *Cambridge handbook on child language*. Cambridge: Cambridge University Press.
- Demuth, K. & Kehoe, M. (2006). The acquisition of word-final clusters in French. *Journal of Catalan Linguistics* 5, 59–81.
- Demuth, K. & McCullough, E. (to appear). The longitudinal development of clusters in French. *Journal of Child Language*.
- Fikkert, P. (1994). *On the acquisition of prosodic structure*. Dordrecht: Holland Institute of Generative Linguistics.

- Fikkert, P. & Levelt, C. C. (2008). How does place fall into place? The lexicon and emergent constraints in children's developing phonological grammar. In P. Avery, B. Elan Dresher & K. Rice (eds), *Contrast in phonology: Theory, perception, acquisition (Phonology and Phonetics 13)*, 231–70. Berlin: Mouton.
- Flack, K. (2007). Sources of phonological markedness. Unpublished doctoral dissertation, University of Massachusetts, Amherst.
- Goad, H. (1998). Consonant harmony in child language: An Optimality-Theoretic account. In S. J. Hannahs & Martha Young-Scholten (eds), *Focus on phonological acquisition*, 113–42. Amsterdam: John Benjamins.
- Goldwater, S. & Johnson, M. (2003). Learning OT constraint rankings using a maximum entropy model. In Jennifer Spenader, Anders Eriksson & Östen Dahl (eds.), *Proceedings of the Stockholm workshop on variation within Optimality Theory*, 111–20. Stockholm: Stockholm University.
- Gnanadesikan, A. (1995/2004). Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater & W. Zonneveld (eds), *Constraints in phonological acquisition*, 73–109. Cambridge: Cambridge University Press.
- Hayes, B. (1999). Phonetically-driven phonology: The role of Optimality Theory and inductive grounding. In Michael Darnell, Edith Moravcsik, Michael Noonan, Frederick Newmeyer & Kathleen Wheatly (eds), *Functionalism and formalism in linguistics, Volume 1: General papers*, 243–85. Amsterdam: John Benjamins.
- Hilaire-Debove, G. & Kehoe, M. (2004). Acquisition des consonnes finales (codas) chez les enfants francophones: Des universaux aux spécificités de la langue maternelle. In *Actes de la 25ème Journée d'Études sur la Parole*, 265–68. Fez: Morocco.
- Ingram, David (1988). The acquisition of word-Initial [v]. *Language and Speech* 31(1): 77–85.
- Jakobson, R. (1941/1968). *Child language aphasia and phonological universals*. The Hague: Mouton.
- Jarosz, G. (2006). Rich lexicons and restrictive grammars – maximum likelihood learning in Optimality Theory. Unpublished doctoral dissertation, Johns Hopkins University.
- Jäger, G. (to appear). Maximum entropy models and Stochastic Optimality Theory. In Jane Grimshaw, Joan Maling, Chris Manning, Jane Simpson & Annie Zaenen (eds), *Architectures, rules, and preferences: A festschrift for Joan Bresnan*. Stanford, CA: CSLI.
- Jäger, G. & Rosenbach, A. (2006). The winner takes it all – almost. Cumulativity in grammatical variation. *Linguistics* 44, 937–71.
- Jesney, K. & Tessier, A. (to appear). Biases in Harmonic Grammar: The road to restrictive learning. *Natural Language and Linguistic Theory*.
- Kehoe, M. & Stoel Gammon, C. (2001). Development of syllable structure in English-speaking children with particular reference to rhymes. *Journal of Child Language* 28, 393–432.
- Keller, F. (2006). Linear Optimality Theory as a model of gradience in grammar. In Gisbert Fanselow, Caroline Féry, Ralph Vogel & Matthias Schlesewsky (eds), *Gradience in grammar: Generative perspectives*, 270–87. Oxford: Oxford University Press.
- Kirk, C. & Demuth, K. (2005). Asymmetries in the acquisition of word-initial and word-final consonant clusters. *Journal of Child Language* 32(4), 709–34.
- Legendre, G., Miyata, Y. & Smolensky, P. (1990a). Harmonic Grammar – a formal multilevel connectionist theory of linguistic wellformedness: An application. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 884–91. Cambridge, MA: Lawrence Erlbaum.
- Legendre, G., Miyata, Y. & Smolensky, P. (1990b). Harmonic Grammar – a formal multi-level connectionist theory of linguistic wellformedness: Theoretical foundations. In *Proceedings of the twelfth annual conference of the Cognitive Science Society*, 388–95. Cambridge, MA: Lawrence Erlbaum.
- Legendre, G., Sorace, A. & Smolensky, P. (2006). The Optimality Theory–Harmonic Grammar connection. In P. Smolensky & G. Legendre (eds), *The harmonic mind: From*

- neural computation to *Optimality-Theoretic grammar*, 339–402. Cambridge, MA: MIT Press.
- Levelt, C. C., Schiller, N. O. & Levelt, W. J. (2000). The acquisition of syllable types. *Language Acquisition* 8, 237–64.
- Levelt, C. & van de Vijver, R. (1998/2004). Syllable types in cross-linguistic and developmental grammars. In R. Kager, J. Pater & W. Zonneveld (eds), *Constraints in phonological acquisition*, 204–218. Cambridge: Cambridge University Press. Original version available on Rutgers Optimality Archive, ROA-265.
- Li, P. & Shirai, Y. (2000). *The acquisition of lexical and grammatical aspect*. Berlin & New York: Mouton de Gruyter.
- Leo, C. & Prinz, M. (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment. *Journal of Child Language* 23, 31–56.
- Lukaszewicz, B. (2007). Reduction in syllable onsets in the acquisition of Polish: Deletion, coalescence, metathesis, and gemination. *Journal of Child Language* 34(1), 52–82.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pater, J. (1997). Minimal violation and phonological development. *Language Acquisition* 6, 201–53.
- Pater, J. (2008). Gradual learning and convergence. *Linguistic Inquiry* 39(2), 334–45.
- Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science* 33, 999–1035.
- Pater, J. & Werle, A. (2001). Typology and variation in child consonant harmony. In Caroline Féry, Antony Dubach Green & Ruben van de Vijver (eds), *Proceedings of HILP5*, 119–39. Potsdam: University of Potsdam.
- Prince, A. (2002). Anything goes. In Takeru Honma, Masao Okazaki, Toshiyuki Tabata & Shin-ichi Tanaka (eds), *New century of phonology and phonological theory*, 66–90. Tokyo: Kaitakusha.
- Prince, A. & Smolensky, P. (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.
- Rose, Y. (2003). ChildPhon: A database solution for the study of child phonology. In Barbara Beachley, Amanda Brown & Frances Conlin (eds), *Proceedings of the 27th Annual Boston University Conference on Language Development*, 674–85. Somerville, MA: Cascadilla Press.
- Smith, N. (1973). *The acquisition of phonology: A case study*. Cambridge: Cambridge University Press.
- Smolensky, P. (1996). The initial state and ‘richness of the base’. Technical Report, Department of Cognitive Science, the Johns Hopkins University, Baltimore, Maryland.
- Smolensky, P. & Legendre, G. (2006). *The harmonic mind: From neural computation to Optimality-Theoretic grammar*. Cambridge, MA: MIT Press.
- Stampe, D. (1969). The acquisition of phonemic representation. In Alice Davidson, Georgia Green & Jerry Morgan (eds), *Papers from the 5th regional meeting of the Chicago Linguistics Society*, 433–44. Chicago: Chicago Linguistics Society.
- Szagan, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language* 21, 109–141.
- Templin, M. (1957). *Certain language skills in children: Their development and interrelationships* (Monograph Series No. 26). Minneapolis: University of Minnesota, The Institute of Child Welfare.
- Tesar, B. (2007). A comparison of lexicographic and linear numeric optimization using violation difference ratios. Unpublished ms, Rutgers University.
- Weide, R. L. (1994). *CMU pronouncing dictionary*. www.speech.cs.cmu.edu/cgi-bin/cmudict.
- Weist, R. & Witkowska-Stadnik, K. (1986). Basic relations in child language and the word order myth. *International Journal of Psychology* 21, 363–81.

- Weist, R., Wysocka, H., Witkowska-Stadnik, K., Buczowska, E. & Konieczna, E. (1984). The defective tense hypothesis: On the emergence of tense and aspect in child Polish. *Journal of Child Language* **11**, 347–74.
- Zamuner, T. S., Kerkhoff, A. & Fikkert, P. (in preparation). Children's knowledge of how phonotactics and morphology interact.
- Zygorowicz, P. (2007). Polish morphonotactics in first language acquisition. In Florian Menz and Marcus Rheindorf (eds), *Weiner Linguistische Gazette* **74**, 24–44.