Restrictiveness and Phonological Grammar and Lexicon Learning

Gaja Jarosz
Yale University

1 Introduction

A central problem of language learnability is the learning of *restrictive* grammars, grammars that generate all the observed forms but as few others as possible. Given only positive evidence, there are many grammars consistent with the observed data, and the learner must select the most restrictive grammar among these. If the learner mistakenly adopts a broader grammar, no positive evidence will contradict this decision since the broader grammar is consistent with all the positive evidence plus additional data. This is known as the subset problem (Angluin 1978; Baker 1979; Berwick 1986).

Within Optimality Theory (OT; Prince and Smolensky 1993/2004), a well-known solution to the restrictiveness problem is a *ranking bias*, a preference for a particular relative ranking between certain constraint types. Ranking biases have been successfully implemented in algorithms that focus on the learning of a language-particular ranking given the correct underlying forms (Prince and Tesar 2004; Hayes 2004; Tessier 2006). As Alderete and Tesar (2002) point out, however, when the full problem of learning rankings and underlying forms is considered, ranking biases are not sufficient to identify restrictive *combinations* of grammars and lexicons due to the interdependence of grammar and lexicon learning.

This paper presents an alternative solution to the restrictiveness problem: Maximum Likelihood Learning of Lexicons and Grammars (MLG; Jarosz 2006). MLG subsumes the effects of ranking biases and naturally extends to the full phonological learning problem, identifying restrictive grammar and lexicon combinations. Rather than using ranking biases to define the relative restrictiveness of multiple analyses of the same data, MLG relies on the likelihood, or probability, that each grammar and lexicon combination assigns to the data. The likelihood provides an explicit measure of how well the grammar and lexicon explain the data, an *objective function* that may be maximized using a well-known statistical learning algorithm.

The paper is organized as follows. Section 2 briefly overviews the use of ranking biases in OT learning. Section 3 introduces MLG, discusses its ability to identity restrictive grammar and lexicon combinations, and explains how the principles of MLG are implemented in the simulations presented in this paper. Section 4 discusses a series of simulations illustrating MLG's capacity to learn restrictive grammar and lexicon combinations. In Section 4.1, simulations illustrate MLG's capacity to derive the effects of the three ranking biases. Sections 4.2 and 4.3 discuss a system with grammar-lexicon interaction for which ranking biases do not suffice and show that MLG successfully identifies the restrictive grammar and lexicon combination.

2 Ranking biases

The most famous of the ranking biases is the *Markedness » Faithfulness* (M » F) bias (Jakobson 1941/1968; Stampe 1969; Smolensky 1996; Prince and Tesar 2004). In general, a preference for ranking markedness constraints as high as possible will favor grammars that avoid marked configurations at the expense of faithfulness violations. As illustrated in (1), the M » F bias favors the restrictive ranking of NoCoda » Faith for the CV language, the language that admits syllables of the shape CV only. The ranking bias favors the restrictive ranking over the broader Faith » NoCoda ranking, which is crucially also consistent with the observed forms.

(1) The Markedness » Faithfulness Bias – *CV Language:*
  a. Positive evidence: ta ma.pi su.pa.mi…
  b. No negative evidence: *tam *pin.ti ta.mit…
  c. Restrictive grammar: NoCoda » Faith

In addition to the M » F bias, two other general biases have been proposed to ensure restrictive rankings. The *Specific Faithfulness » General Faithfulness* bias favors restrictive rankings that permit a contrast in certain positions only over rankings that permit a contrast everywhere (Smith 2000; Hayes 2004; Tessier 2006). As shown in (2), a high relative ranking of specific faithfulness favors the restrictive ranking that permits voicing contrast in onset position only over the broader ranking permitting voicing contrast in all positions.

(2) Specific Faithfulness » General Faithfulness – *Voicing contrast in onset only:*
  a. Positive evidence: [bat] [pat] [dak]…
  b. No negative evidence: *[bad] *[pag]…
  c. Ident[voi]/Onset » NoVoiObs » Ident[voi]

The third general bias, *Output-Output Faithfulness » Input-Output Faithfulness,* ensures that restrictions across morphologically related forms are reflected in the ranking (McCarthy 1998; Hayes 2004; Tessier 2006). In the example in (3), a high ranking of output-output faithfulness enforces the restriction that roots be bimoraic.

(3) Output-Output Faithfulness » Input-Output Faithfulness – *Only bimoraic roots allowed:*
  a. Positive evidence: {[kaa], [kaa+ga]} {[bita], [bita+ga]}
  b. No negative evidence: *{[kaa], [ka+ga]} *{[ka], [ka+ga]}
  c. {FtBin, OO-Ident[wt]} » IO-Ident[wt]

In all three cases, the bias is needed to ensure that a ranking accepting additional forms, which is also consistent with the positive evidence, is not selected by the learner. While ranking biases are intuitively straightforward, their implementation in a learning algorithm can be rather complex. For example, as Prince and Tesar (2004) show, simply imposing a Markedness » Faithfulness bias throughout learning is not sufficient: the learner must be able to minimize the number of high-ranked faithfulness constraints overall in a hierarchy with many constraints of each type. In the case of the Specific Faithfulness » General Faithfulness bias, the general-to-specific relationships between constraints must themselves be learned because these relationships are often derived from the language-particular constraint hierarchy (Tessier 2006).

3 MLG: Richness of the base and likelihood maximization

Maximum Likelihood Learning of Lexicons and Grammars (MLG) is a theory of phonological learning that accounts for the learning of phonological grammars and lexicons from unstructured overt phonological forms. This section presents an overview of the theory and discusses its capacity to learn restrictive grammar and lexicon combinations[1].

MLG is a generative, probabilistic model of the acquisition of a phonological grammar and lexicon of underlying forms. As such, it relies on a formal, probabilistic characterization of both the grammar and lexicon: the grammar and lexicon are both probabilistic entities. The grammar is a probability distribution over rankings of OT constraints and assigns a conditional probability to possible surface realizations of a given underlying form[2]. The lexicon is probabilistic as well and associates each morpheme with a set of possible underlying forms, each with its own likelihood. These probabilistic components can express uncertainty (as in the initial stages of learning) or variation by spreading probability over multiple rankings or underlying forms, and they can express certainty (as in the final stages of learning) by assigning to a single ranking or underlying form a probability of one. Together, the grammar and lexicon assign a likelihood, or probability, to the overt forms of the language.

Learning in MLG relies on two general principles: richness of the base (ROTB) and likelihood maximization. According to ROTB, the set of possible underlying forms is universal: there are no systematic, language-particular restrictions on underlying representations, and therefore all language-specific restrictions must be handled by the grammar (Prince and Smolensky 1993/2004). MLG incorporates a probabilistic formulation of ROTB into the learning model. The second learning principle, likelihood maximization, defines the correct

---

[1] For a more in-depth presentation of the structure and properties of MLG see Jarosz (2006).

[2] Various probabilistic variants of OT proposed in previous work, such as Stochastic OT, Partial Order Grammars, and Floating Constraints, are all examples of such a probabilistic grammar (Boersma 1998; Anttila 1997; Reynolds 1994).

grammar and lexicon combination as the one that maximizes the likelihood, or probability, of the overt forms. In other words, likelihood maximization requires that the grammar and lexicon combination generate all and only the observed forms of the target language with high probability, a standard generative perspective cast in a probabilistic setting.

These general principles form the foundation of MLG and are incorporated into a learning model with two stages of learning, phonotactic and morphophonemic learning (see also Prince and Tesar 2004; Hayes 2004):

(4) Two Stage Learning in MLG

a. Phonotactic Learning
   i A fixed, universal rich base is assumed
   ii No morphological awareness
   iii Grammar learning but no lexicon learning

b. Morphophonemic Learning
   i Words are analyzed into component morphemes
   ii Learning of morpheme specific underlying forms occurs
   iii Further learning of the grammar to account for alternations

The phonotactic stage of learning occurs before morphological awareness and prior to the development of a phonological lexicon; it is during this stage that learning of a language-specific phonotactic grammar takes place. Formally, this stage consists of gradual learning of a grammar that maximizes the likelihood of the overt forms, given a (fixed) rich base. The rich base is a representation of all possible underlying forms, each with roughly equal likelihood. This base may be characterized as the expected, unbiased distribution over phonological forms given by the free combination of phonological elements. Under this characterization, phonotactic learning involves maximizing the likelihood of the observed distribution of overt forms, given the expected distribution.

During morphophonemic learning, words are analyzed into component morphemes, and each morpheme is associated with its own probabilistic lexical entry. During this stage, the grammar gradually transitions between the phonotactic grammar learned in the first stage and the target grammar, while the lexicon gradually converges on the target lexicon. Formally, during morphophonemic learning the *grammar and lexicon combination* that maximizes the likelihood of the overt forms is gradually learned. Thus, the crucial difference between the two stages resides in the role of the rich base and lexical learning.

Figure 1 illustrates the outcomes of the two stages of learning with a simple example of a language with two overt forms, CV and CVC, the former occurring twice as often as the latter. During phonotactic learning, a single, equally-distributed rich base is held constant while the grammar that maximizes the likelihood of the overt forms is gradually learned. Phonotactic learning results in a restrictive grammar that matches the frequencies of the overt forms, given the rich base. This grammar is shown in the figure as a mapping from the rich base to the overt forms, with outputs generated in proportion to their frequency of occurrence (darker shading corresponds to higher likelihood). During morphophonemic learning, each morpheme (corresponding to each overt form in this case) is associated with its own probabilistic lexical entry. The lexical entries and grammar are gradually updated until they converge on the target lexicon and grammar shown in the lower portion of the figure. The target lexicon and grammar generate all and only the correct forms for each overt form, as indicated by the black shading of the correct output forms.

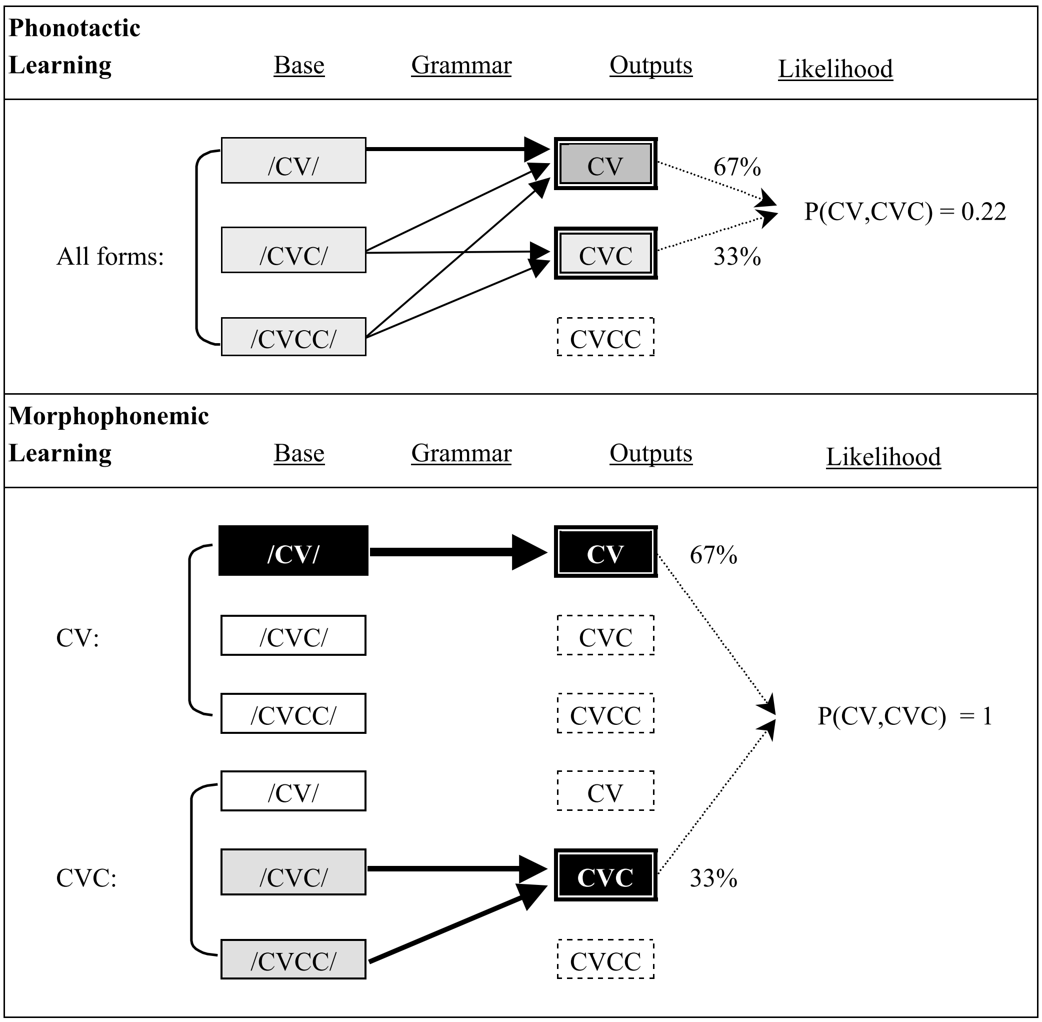

Figure 1: Outcomes of Phonotactic and Morphophonemic Learning in MLG

The division into the phonotactic and morphophonemic stages corresponds to children's phonological development. A large body of literature shows that children acquire at least some phonotactic knowledge by approximately 9 months of age (Jusczyk et al 1993; Friederici and Wessels 1993). On the other hand, learning of alternations and the lexicon occurs much later, roughly between the ages of 2 and 4.5 years, and in some cases even later (Berko 1958; Stager and

Werker 1997; Pater 1997; Pater, Stager and Werker 2004; MacWhinney 1978). The two stages of MLG are based on this overall developmental progression.

3.1 Restrictiveness in MLG

The combination of likelihood maximization and ROTB favors restrictive grammars that generate *all and only* the observed forms. In order to assign a high likelihood to the observed forms, the grammar must map some underlying form(s) to each of them. This guarantees that the grammar can generate all the observed forms. Furthermore, the likelihood can only be high if few or no additional forms are generated since generating any additional forms takes away probability from the observed forms (the probabilities of all predicted forms sum to 1 by definition). Given ROTB, likelihood maximization favors a restrictive grammar that maps as many hypothetical underlying forms (as much of the rich base) to actual forms as possible.

In MLG, probability provides a principled formalization of the degree of explanation afforded by a grammar and lexicon. An explanatory (restrictive) grammar and lexicon combination is one according to which the observed forms are expected by the model (assigned high probability) and are not treated as accidental. This notion of explanatory power is exactly what likelihood maximization plus ROTB captures, and it is this principle that favors restrictive grammar and lexicon combinations in MLG.

3.2 Implementation

Given the overall structure of the MLG model, there are a number of possible implementations of the actual learning, or likelihood maximization, procedure. In the simulations described here, I employ the standard Expectation-Maximization (EM) algorithm (Dempster et al. 1977). EM is a general-purpose algorithm for likelihood maximization with hidden variables, and it has some properties that make it a suitable candidate for the present task. First, EM is guaranteed to converge on a (local) maximum, and second, it adjusts the grammar and lexicon gradually. In other words, EM transitions gradually from the initial state of the grammar and lexicon to the target states, enabling an examination of the gradual learning path it predicts.

In these simulations, I make the simplifying assumption that the grammar and lexicon are lists of rankings and underlying forms, respectively, with associated probabilities. This simplifies the maximization step of the EM algorithm used here but is not an intrinsic aspect of MLG and crucially does not determine the overall, qualitative predictions of the theory. For a discussion of how more sophisticated representations of the grammar and lexicon may be implemented in MLG, see Jarosz (2006).

4 Simulations

This section presents simulations illustrating MLG's ability to derive the effects of the three ranking biases. It also describes a test case from Alderete and Tesar

(2002) for which ranking biases do not suffice and shows MLG's successful learning for this language as well.

4.1 Deriving the effects of ranking biases

To illustrate how MLG derives the effect of the M » F bias, consider a slightly enhanced version of the CV language introduced earlier, with the five constraints shown in (5).

(5) CV Language Constraints:
ONSET: No vowel initial syllables
NOCODA: No consonant final syllables
*COMPLEXONSET: No syllable-initial consonant clusters
*COMPLEXCODA: No syllable-final consonant clusters
MAX: No deletion

The first simulation focuses on the most restrictive possible language with these constraints: the language that admits only CV syllables. This means the learner is presented only with CV overt forms. As described in Section 3, phonotactic learning proceeds from a probabilistic rich base, which in this case consists of nine syllable types of equal likelihood (shown in Table 1). This probabilistic lexicon is held constant during phonotactic learning while the phonotactic grammar is learned. Learning of the phonotactic grammar proceeds from an initial unbiased state in which all rankings are equally likely (all constraints are ranked equally).

Table 1: Initial (and final) Lexicon for Phonotactic Learning

Overt Form	Rich Base of Underlying Forms								
[CV]	CV 11%	CVC 11%	CVCC 11%	CCV 11%	CCVC 11%	CCVCC 11%	V 11%	VC 11%	VCC 11%

Following phonotactic learning, the model settles on a grammar consistent with the CV language, shown in Table 2. This is the correct, restrictive grammar since NOCODA, ONSET, and *COMPLEXONSET outrank MAX, prohibiting all but CV syllables. In the final state, *COMPLEXCODA is freely rankable since its ranking is irrelevant as long as NOCODA ranks above MAX.

Table 2: Final Ranking after Phonotactic Learning

Final Ranking	Description
{ONSET, *COMPLEXONSET, NOCODA » MAX}, *COMPLEXCODA	CV language

In this simple example, learning is concluded after the phonotactic stage. This is because there are no alternations and because the most restrictive language is being learned. As a result, there is no pressure to select any particular underlying forms during morphophonemic learning – all are mapped to CV by the restrictive grammar. Consequently, the final lexicon remains rich, identical to the starting lexicon.

Why is the restrictive ranking correctly learned? The rich base compels the learner to favor a ranking that maps as much of the rich base as possible onto the observed CV forms. The superset rankings admitting additional forms 'waste' probability on unobserved forms, lowering the overall likelihood of the observed forms. During learning, rankings are rewarded in proportion to how much work they do in explaining the observed data. The wasteful superset rankings are penalized during the learning process while the successful restrictive rankings are rewarded, eventually resulting in a final grammar admitting only CV forms. MLG derives the desired effect of the M » F bias, selecting a restrictive grammar with all active markedness constraints ranked above faithfulness constraints.

The languages requiring Specific Faithfulness » General Faithfulness and OO-Faith » IO-Faith ranking biases are learned in an analogous fashion. Due to space considerations, the details of the other two simulations are omitted here. In short, the MLG learner learns the correct, restrictive ranking of IDENT[VOI]/ONSET » NOVOIOBS » IDENT[VOI] for the language with positional neutralization. The high ranking of specific faithfulness is learned without an explicit bias favoring that ranking, and learning occurs without access to the correct underlying forms; overt forms are the only language-specific information available to the learner. In the final case, MLG learns the restrictive ranking with high output-output faithfulness requiring the bimoraic restriction on base forms be extended to the root of affixed forms. In both cases, the combination of likelihood maximization and ROTB drive the learning of the restrictive grammars and the required underlying forms of the language[3].

The rich base places pressure on the learner to identify rankings that neutralize unobserved distinctions, such as voicing in coda position and weight in roots. MLG relies on a single learning strategy, likelihood maximization and richness of the base, to derive the effects of all three ranking biases. Additionally, the adult rankings are correctly learned in the absence of any language-particular information about the target underlying forms. MLG's capacity to account for the learning of both underlying forms and rankings is key to its capacity to identify restrictive grammar-lexicon combinations, as described in the next section.

4.2 Insufficiency of ranking biases

Ranking biases account for many cases of restrictiveness. However, there are also languages whose restrictiveness depends not on the relative ranking of constraint

[3] An appropriate rich base is specified for each simulation. In principle, the rich base is universal; however, each simple simulation focuses on one small portion of phonology and the base used in the simulation reflects that. For example, the voicing simulation relies on a rich base of forms that vary in the voicing specification of consonants in onset and coda positions, while the weight simulation relies on a base of forms that vary in the weight of roots in base and affixed contexts.

types, but on the interaction between the lexicon and constraint ranking. The following system (Alderete and Tesar 2002) is based on languages with stress-epenthesis interactions such as Mohawk, Selayarese and Yimas.

(6) Stress-epenthesis Interaction (Alderete and Tesar 2002):

a. Regular final stress:
/pakat/ → pakát
/pikat/ → pikát
/pakit/ → pakít
/pikit/ → pikít

b. Penultimate stress to avoid stressing epenthetic vowels:
/pakt/ → pákit
/pikt/ → píkit

c. Crucially, penultimate stress is not possible with underlyingly disyllabic forms:
*pákat
*píkat

In this system, stress is usually final. However, because epenthetic [i] cannot be stressed, penultimate stress occurs in forms with epenthesis in the final syllable. When faced with learning both the lexicon and the ranking for this system, the learner has no access to the correct underlying forms. The learner is exposed to the following overt forms only:

(7) pakát pikát pakít pikít pákit píkit

From this, the learner must deduce the combination of underlying forms and ranking that together account for the surface prohibition against penultimate stress with final [a]. In other words, the correct grammar and lexicon combination admits words with penultimate stress only if the final vowel is [i] (since [i] is the epenthetic vowel).

The challenge faced by the learner is that there are two possible rankings consistent with the data, and these rankings require different underlying forms. The crucial constraints are shown in (8)[4].

(8) Crucial Constraints for Stress-Epenthesis Interaction:
MAINSTRESSRIGHT: Main stress falls on the rightmost syllable
HEADDEP: No epenthesis in stressed syllables
FAITHACCENT: Underlying accents are preserved

---

[4] Of course, there are other constraints that must be correctly ranked in this language, such *COMPLEX and MAINSTRESSLEFT, but the particular challenge at issue here resides in the relative ranking of the three in (8).

The restrictive grammar and lexicon combination requires the learner to posit unfaithful underlying representations for the forms with penultimate stress and rely on HEADDEP to compel violations of MAINSTRESSRIGHT as shown in (9). The ranking requires main stress to be final unless rightmost stress would result in stress on an epenthetic vowel.

(9) Restrictive grammar and lexicon combination:

/pakt/ (/pákt/)	HEADDEP	MAINSTRESSRIGHT	FAITHACCENT
☞ [páki̲t]		*	*
[pakî̲t]	*!		(*)

An alternative grammar-lexicon combination (10) attributes the penultimate stress in forms like pákit to a constraint demanding faithfulness to lexical stress. This ranking is less restrictive since it admits unobserved forms such as *pákat and *pîkat.

(10) Alternative grammar and lexicon combination:

/pákit/	FAITHACCENT	MAINSTRESSRIGHT	HEADDEP
☞ [pákit]		*	
[pakît]	*!*		

The problem is that both rankings involve *Faithfulness » Markedness » Faithfulness* – ranking biases do not identify the restrictive ranking since the faithfulness constraints are unrelated. In other words, restrictiveness does not follow from any general property of the constraints; restrictiveness, as before, depends on what the grammar does with hypothetical underlying forms, the rich base. As shown in (11) and (12), hypothetical underlying forms are mapped to observed forms by the restrictive grammar-lexicon combination but are mapped to unobserved, ungrammatical forms by the alternative grammar-lexicon combination.

(11) The restrictive ranking with rich base inputs:

/pákat/	HEADDEP	MAINSTRESSRIGHT	FAITHACCENT
[pákat]		*!	
☞ [pakát]			**

(12) Alternative ranking with rich base inputs:

/pákat/	FAITHACCENT	MAINSTRESSRIGHT	HEADDEP
💣 [pákat]		*	
[pakát]	*!*		

This example underscores the importance of underlying forms as well as the insufficiency of ranking biases in phonological learning. The first issue concerns the identification of unfaithful underlying forms in the absence of alternations. In order to identify the restrictive grammar-lexicon combination the learner must be able to consider unfaithful underlying forms, those with a final cluster, in the absence of alternations. The second issue concerns the insufficiency of ranking biases to identify restrictive rankings in general. In this example, ranking biases cannot distinguish the restrictive ranking from the superset ranking. As the next section shows, MLG provides a solution to both these problems.

4.3 Restrictiveness in grammar and lexicon learning

This section presents the simulation applying MLG to the stress-epenthesis interaction system described above. The full set of constraints used in the simulation are shown in (13). This set of constraints can generate languages such as the target language with stress-epenthesis interaction, the superset language with lexical stress, as well as a number of other languages. The task of the learner is to select the correct grammar and the correct lexicon based on overt forms alone with no access to the target underlying forms.

(13) Constraints for Stress-Epenthesis Interaction:
MAINSTRESSRIGHT: Main stress falls on the rightmost syllable
MAINSTRESSLEFT: Main stress falls on the leftmost syllable
HEADDEP: No epenthesis in stressed syllables
FAITHACCENT: Underlying accents are preserved
*COMPLEX: Syllable margins may not contain multiple segments

As in the simulations described earlier, the phonotactic stage begins with the learning of a phonotactic grammar given a fixed rich base. In the simplified universe of this simulation, the rich base consists of forms varying in the presence or absence of underlying vowels and in the specification of underlying stress. The entire base of twenty, equally distributed forms is shown in Table 3.

Table 3: Rich Lexicon for Phonotactic Learning

Overt Form	Rich Base of Underlying Forms							
ALL	pákat 5%	pîkat 5%	pákit 5%	pîkit 5%	pkat 5%	pkit 5%	pikt 5%	pakt 5%
	pakát 5%	pikát 5%	pakît 5%	pikît 5%	pkát 5%	pkît 5%	pîkt 5%	pákt 5%
	pakat 5%	pikat 5%	pakit 5%	pikit 5%				

At the onset of the phonotactic stage, the initial grammar is unbiased as well, with all rankings having equal probability. The gradual learning of the restrictive grammar with stress-epenthesis interaction relative to the alternative, superset grammar is shown in **Figure 2**. During phonotactic learning, the probability of the restrictive grammar gradually increases, while the probability of the alternative grammar with lexical stress goes to zero. It is during the phonotactic stage that the full force of the rich base compels the learner to reject any superset grammars.

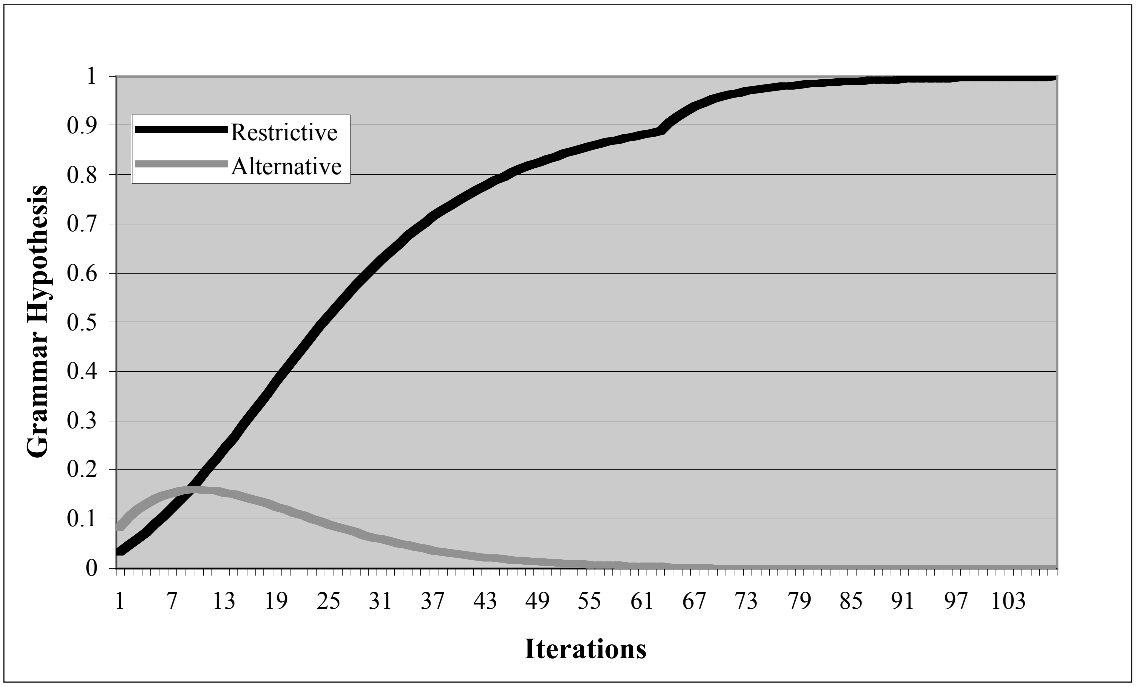

Figure 2: Grammar Learning of Stress-Epenthesis Interaction System

At the start of morphophonemic learning (at around iteration 63), learning of the lexicon begins, and grammar learning continues. As the figure indicates, at the conclusion of the simulation, the learner has selected the correct, restrictive grammar, HEADDEP » MAINSTRESSRIGHT » FAITHACCENT, assigning a probability of one to this hypothesis and zero to all others.

The initial lexicon is unbiased, with all forms in the rich base serving as potential underlying forms for each overt form. The final lexicon that is learned during the morphophonemic stage is shown in Table 4. The final lexicon confirms that the correct, restrictive grammar-lexicon combination has been learned: this is evidenced by the richness that has been retained in the lexicon. The particular distribution of underlying forms selected for each overt form is not of great consequence. What is key is that no particular underlying stress has been selected for any of the overt forms: this underscores the fact that the learner has correctly discovered that stress is grammatical, not lexical, in this system. Furthermore, the lexical information crucial to this system has been learned since all and only the forms with surface penultimate stress are given underlying representations of the shape CVCC. As discussed above, the restrictive grammar maps /CVCC/ underlying forms to [CVCVC] forms with penultimate stress.

Table 4: Final Lexicon after Morphophonemic Learning

Overt Form	Learned Lexicon				
[pakát]	pakát 38.3%	pákat 23.5%	pakat 38.3%		
[pikát]	pikát 21.7%	pikat 21.7%	pîkat 13.3%	pkat 21.6%	pkát 21.5%
[pakît]	pakît 38.3%	pákit 23.5%	pakit 38.3%		
[pikît]	pikît 21.7%	pikit 21.7%	pîkit 13.3%	pkit 21.6%	pkît 21.5%
[pákit]	pákt 50.5%	pakt 49.5%			
[pîkit]	pîkt 50.5%	pikt 49.5%			

In sum, MLG successfully learns the restrictive grammar-lexicon combination that captures the observed surface restriction. As before, the learner relies on a single learning strategy: ROTB and Likelihood Maximization. The rich base available during phonotactic learning enables the learner to distinguish between two grammars that cannot be distinguished by ranking biases. The same rich base that compels the identification of restrictive rankings provides a base from which underlying forms are selected, enabling the learning of unfaithful mappings in the absence of alternations.

5 Discussion

The stress-epenthesis system exemplifies two real challenges faced by children acquiring the phonologies of their native languages. The first challenge is the identification of restrictive grammars that cannot be distinguished from broader grammars on the basis of general ranking biases. The second challenge relates to the learning of underlying forms themselves. Learning the restrictive grammar for the stress-epenthesis system requires the capacity to consider underlying forms that are crucially distinct from surface forms even though the overt forms have a single surface realization. McCarthy (to appear) identifies several cases in which an unfaithful mapping is required for forms that do not alternate on the surface. Focusing on Sanskrit, McCarthy demonstrates that the restrictive grammar-lexicon combination requires all long mid vowels in the language to be derived via coalescence from underlying [ai] or [au] sequences. In other words, the learner must be able to consider [ai] and [au] sequences as potential underlying representations of nonalternating surface mid vowels. The Sanskrit case and other examples McCarthy describes underscore the significance of underlying forms and their interaction with the grammar in phonological learning.

In addition to their role in learnability, biases have been extensively employed in the acquisition literature to account for the initial state of acquisition as well as order of acquisition effects (Gnanadesikan 1995/2004; Tessier 2006). MLG can in fact derive the effects of ranking biases in the modeling of acquisition stages as well. The phonotactic stage in MLG results in a grammar that captures the phonotactics of the target language but is biased against infrequent and marked forms. The gradual transition from the phonotactic grammar to the target grammar results in gradual introduction of marked forms. See Jarosz (to appear) for details.

In sum, a probabilistic formulation of ROTB enables restrictiveness to be approached directly. Rather than encoding ranking biases explicitly, the present proposal employs a single learning strategy, likelihood maximization combined with richness of the base. Because of its generality, the solution derives the desired effects of all three ranking biases and additionally extends to the learning of restrictive grammar and lexicon combinations for which ranking biases do not suffice. MLG addresses an additional problem of phonological learning: the learning of underlying forms. Because MLG learns both underlying forms and rankings, it handles cases where underlying forms are crucial to identifying restrictive rankings. Finally, the probabilistic formulation of ROTB also enables MLG to select unfaithful underlying forms in the absence of alternations.

References

Alderete, John and Bruce Tesar. 2002. Learning Covert Phonological Interaction: an Analysis of the Problem Posed by the Interaction of Stress and Epenthesis. RuCCS Technical Report TR-72, Rutgers University.

Angluin, Dana. 1978. Inductive inference of formal languages from positive data. *Information and Control* 45.117 135.

Baker, C. L. 1979. Syntactic Theory and the Projection Problem. *Linguistic Inquiry* 10.4.533- 581.

Berwick, Robert. 1986. *The acquisition of syntactic knowledge*. MIT Press, Cambridge, MA.

Berko, Jean. 1958. The child's learning of English morphology. *Word* 14.150-177.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry* 32(1).45-86.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. Maximum Likelihood from incomplete data via the EM Algorithm. *Journal of Royal Statistics Society*. 39(B).1-38.

Friederici, Angela D. and Jeanine E. Wessels. 1993. Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics* 54.287-295.

Gnanadesikan, Amahlia. 1995/2004. Markedness and faithfulness constraints in child phonology. In R. Kager, W. Zonneveld, J. Pater, eds., *Fixing Priorities: Constraints in Phonological Acquisition*, Cambridge, Cambridge University Press.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: the early stages. Appeared in Kager, Rene, Pater, Joe, and Zonneveld, Wim, (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press.

Jakobson, Roman. 1968. *Child Language, Aphasia and Phonological Universals* (A.R. Kuler, Trans.), Mouton, The Hague. (Originally published as Jakobson, Roman (1941). Kindersprache, Aphasie, und algemeine Lautgesetze. Uppsala: Almqvist & Wiksell.)

Jarosz, Gaja. 2006. *Rich Lexicons and Restrictive Grammars – Maximum Likelihood Learning in Optimality Theory*. PhD dissertation, Johns Hopkins University.

Gaja Jarosz. to appear. Stages of Acquisition without Ranking Biases: the Roles of Frequency and Markedness in Phonological Learning. In M. Becker (ed.), *UMass Occasional Papers in Linguistics* 36.

Jusczyk, Peter W., Angela D. Friederici, Jeanine M.I. Wessels, Vigdis Y. Svenkerud, and Ann Marie Jusczyk. 1993. Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language* 32.402-420.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. Appeared in M. Catherine Gruber, Derrick Higgins, Kenneth Olson, and Tamra Wysocki, eds., *Proceedings of the Chicago Linguistic Society* 5, Vol. II: The Panels, Chicago, CLS.

McCarthy, John. to appear. Taking a Free Ride in Morphophonemic Learning. In *Catalan Journal of Linguistics* 4. Special issue on phonology in morphology, edited by Maria-Rosa Lloret and Jesús Jiménez.

Pater, J., C. Stager, and J. Werker. 2004. The Perceptual Acquisition of Phonological Contrasts. *Language* 80.3.

Prince, Alan, and Paul Smolensky. 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.

Prince, Alan and Bruce Tesar. 2004. Learning Phonotactic Distributions. In Kager, Rene, Pater, Joe, and Zonneveld, Wim, (eds.), *Fixing Priorities: Constraints in Phonological Acquisition*. Cambridge University Press.

Smith, Jennifer L. 2000. Positional faithfulness and learnability in Optimality Theory. In Rebecca Daly and A. Rehl, eds. *Proceedings of ESCOL99*, Ithaca, CLC Publications.

Smolensky, Paul. 1996. *The Initial State and 'Richness of the Base'*. Technical Report JHU-CogSci-96-4.

Stager, Christine, and Janet Werker. 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 388.381-382.

Stampe, David. 1969. The acquisition of phonetic representation. In Papers from the Fifth Regional Meeting, CLS. pp 443-454. Chicago: Chicago Linguistic Society.

Tesar, Bruce. 1995. *Computational Optimality Theory*. Ph.D. dissertation, University of Colorado, Boulder.

Tesar, Bruce, and Alan Prince. to appear. Using phonotactics to learn phonological alternations. Revised version will appear in The Proceedings of CLS 39, Vol. II: The Panels. ROA-620.

Tessier, Anne-Michelle. 2006. *Biases and Stages in Phonological Acquisition*. PhD dissertation, Univeristy of Massachusetts, Amherst.