A revised version of this paper will appear in the proceedings of CLS 39.

Using phonotactics to learn phonological alternations

Bruce Tesar and Alan Prince
Rutgers University

1 Introduction

Linguistic theory presents a search space of possible grammars. Under that theory, learning a language requires navigating that search space in order to find the correct grammar (the most restrictive grammar consistent with the overt data provided to the learner). The computational difficulty of that task depends almost entirely on the formal structure of the search space. Any plausible linguistic theory includes a very large number of possible grammars, too large to be exhaustively searched by a learner. An adequate account of language learning must explain how the learner relates to the formal structure of the grammar space so that it can find the correct grammar much more efficiently than by exhaustive search.

One of the many challenges posed by the formal structure of linguistic theories is the interdependence of underlying forms and grammatical mappings. A learner starts out knowing neither the correct underlying-to-surface mapping, nor the phonological underlying forms for the morphemes of the language. Further, the two work together to produce the overt surface forms, and that interdependence requires that the two be learned simultaneously.

Here is an illustration of this interdependence, using voicing alternations. (See Kager (1999) for exploration of a similar example.) Consider the pair of forms consisting of a bare root, and that root with a suffix, shown in (1).

(1) a. rat
b. rade

There are two obvious analyses of this observation. The first analysis is that the relevant process is one of syllable-final devoicing. This analysis posits /rad/ as the underlying form for the root, and the final obstruent /d/ is devoiced in coda position, in the bare root form. The second analysis is that the relevant process is one of inter-vocalic voicing. This analysis posits /rat/ as the underlying form for the root, and the final obstruent /t/ is voiced inter-vocalically, in the suffixed form.[1]

---

[1] In languages having both syllable-final devoicing and inter-vocalic voicing, voicing is predictable in both environments, and either underlying form will work. Note that voicing might still be contrastive word-initially or post-consonantally in such a language.

Clearly, the mapping, which embodies the choice of active phonological process, and the underlying form are mutually dependent; the choice of one depends upon the choice of the other. It may be that only one of these analyses would ultimately be sustainable, given a sufficient amount of data. But how is the learner to arrive at the correct analysis?

One possible source of information is the phonotactics of the language. Are all codas in the language unvoiced? Are all intervocalic obstruents voiced? The role of phonotactics in determining morphemic alternations has been much discussed (e.g., Kisseberth, 1970; Sommerstein, 1974; Kiparsky, 1980; Goldsmith, 1991). In the context of learning, Pater (2000) has suggested that the phonotactics can identify the correct analysis of the above voicing alternation. Prince and Tesar (1999) and Hayes (1999) proposed generally that learners engage in a period of pure phonotactic learning prior to the direct analysis of phonological alternations, and that the learned phonotactics are used in the learning of underlying forms. The present work fashions those suggestions into a concrete learning algorithm, using an analysis of voicing alternations as a testing ground.

The algorithm presented here first processes the data from a purely phonotactic perspective, learning as best it can a mapping capturing the phonotactics of the language. It then considers morphemic identity and the morphological structure of the words in the data, and attempts to construct a single underlying form for each morpheme. When a feature value in a morpheme alternates, the learner considers each value of the feature in turn, trying an underlying form with each feature value, and seeing which one behaves properly with respect to the phonotactic mapping. If one of the feature values works properly, it is adopted for the learned underlying form. A distinct possibility faced by the algorithm is that neither feature value behaves completely correctly, due to incompleteness of the mapping learned from the phonotactics. In that case, for each feature value, the learner engages in further refinement of the mapping in an effort to fully match the data. The learner then selects the combination of underlying form and (refined) mapping that succeeds in matching the data. In this “back and forth” way, the learner navigates the interdependence between underlying forms and mapping, arriving at a correct grammar.

2 The problem: Voicing phonotactics and alternations

2.1 Background assumptions

The learner is presumed to have a full Optimality Theoretic system available from beginning, so all of the universal constraints are available. The data take the form of words. We assume that learners exhibit an early phonotactic stage of learning, during which they are unaware of the morphological decomposition of the words. At this stage, the learner is attempting to learn the most restrictive mapping that is capable of reproducing the words taken as isolated forms. We further assume that learners later engage in an extensive process of morphological analysis to learn the morphological structure of words. The learning of phonological underlying forms for morphemes, and the co-attendant phonological processes, takes place in the context of the learning of morphological structure.

For the present study, we will abstract away from the challenges of learning the morphological structure of the words. During the phonotactic stage of learning, each word is treated by the learner as a morphologically unanalyzed form. During the later alternation stage of learning, the learner is provided with a full segmentation of each word into morphemes, with the identity of each morpheme given. Full morphological paradigms of data are provided. We will also assume that the morphology is purely concatenative, and that each morpheme has exactly one phonological underlying form.

2.2 The test system for voicing alternations

The roots of our test system are CVC monosyllables. The suffixes are single vowels. We are only concerned with voicing, so we won't introduce contrasts in other features. The base set of possible inputs for the morphemes is given in (2).

(2) a. Roots: /tat/, /dat/, /tad/, /dad/
b. Suffix: /-e/

The paradigms in the data each consist of two forms, a bare root and a suffixed root. Thus, the potential alternation lies in the voicing of the final consonant of a root. Paradigms look like {tat, tade} and {tat, tate} on the surface. Morpheme identity allows the learner to see the different surface allomorphs of each morpheme: for example, {tat$_1$, tad$_1$-e$_5$}, where the subscripts indicate morpheme identity. The data are of this form during the learning of alternations.

The OT system has four constraints, similar to those of (Ito and Mester, 1997; Lombardi, 1999), listed in (3).

(3)	NoVoi	no voiced obstruents
	NoSFV	no syllable-final voiced obstruents
	IVV	no inter-vocalic voiceless obstruents
	IDVoi	surface voicing must match underlying voicing

These four constraints serve to define a space of six languages. The languages, and some rankings generating them, are described in Table 1.

Table 1 The language typology.

Label	Description	Ranking
A	no voicing	NoVoi ≫ rest
B	voicing always/only IV	{NoSFV, IVV} ≫ NoVoi ≫ IDVoi
C	coda devoicing, IV voicing	{NoSFV, IVV} ≫ IDVoi ≫ NoVoi
D	coda devoicing	NoSFV ≫ IDVoi ≫ {NoVoi, IVV}
E	IV voicing	IVV ≫ IDVoi ≫ {NoSFV, NoVoi}
F	full contrast	IDVoi ≫ rest

2.3 Phonotactic restrictiveness

Properly learning the phonotactics of the language requires not only observing where structures are allowed, but also observing where they aren't allowed. Given that the learner is working on the basis of positive evidence, the learner must infer phonotactic restrictions on the basis of gaps in the data. In the context of grammar learning, the learner's goal is to determine the most restrictive grammar which is consistent with the data: it permits the forms attested in the data, while admitting as few as possible of the forms not attested in the data. this problem is sometimes known as the subset problem (Baker, 1979; Angluin, 1980).

In the system being investigated here, restrictiveness will concern the behavior of voicing in different phonological contexts. The system has a total of eight possible inputs in the base: four bare roots, and four root-suffix combinations. Recall that during the phonotactic learning stage, the learner treats the words as isolated entities, and does not analyze them morphologically. The phonotactic inventory for a language is determined by the set of possible outputs that result when each of the eight possible inputs is run through the mapping for the language. The largest language in this system has eight outputs; any form of neutralization will reduce the number of surface outputs in the language.

The outputs of each language, from the view of phonotactic learning, are shown in Table 2.

Table 2 The words of the languages (the view from phonotactic learning).

Label	Words of the Language							
A	tat				tate			
B	tat					tade		
C	tat		dat			tade		dade
D	tat		dat		tate	tade	date	dade
E	tat	tad	dat	dad		tade		dade
F	tat	tad	dat	dad	tate	tade	date	dade

There are numerous restrictiveness relations (subset relations) among the languages: A ⊆ D ⊆ F, B ⊆ C ⊆ D, and C ⊆ E. All languages are subsets of F. Each language's grammar is consistent with the data of its subset languages. A and B are the most restrictive, permitting no contrast in voicing in any position. F is the least restrictive, permitting full voicing contrast in all positions.

According to the principle of richness of the base (Prince and Smolensky, 1993), the same inventory of possible inputs is available to every language. Thus, the restrictiveness relations cannot be solved simply by selecting the proper underlying forms for morphemes. Restrictiveness is the responsibility of the ranking. If the learner sees data from language B, {tat, tade}, and *only* that data, it would be a fundamental mistake to select a ranking consistent with language D. The data of language B suggest that voicing is not contrastive anywhere; the ranking for language D permits voicing to contrast in onsets.

2.4 Overview of the problem

The learner is provided, as data, with labeled outputs. In each output word, the segmentation of the word into morphemes is indicated, and the identity of each morpheme is provided. The goal of the learner is to produce a grammar consisting of a ranking of the constraints and a lexicon containing one underlying form for each morpheme. The learned grammar should correctly reproduce the data and enforce the phonotactic restrictions implicit in the data.

For this system, with four constraints, there are 4! = 24 distinct total rankings of the constraints. The suffix never varies, so its underlying form is transparent. Each of the roots has two plausible underlying forms, one each with the final consonant voiced and voiceless. The number of possible lexica is the product of the number of possible underlying forms for each morpheme. With four possible roots, there are a total of 16 possible lexica. Thus, the number of possible grammars to be considered by the learner is 24 * 16 = 384 grammars.

The goal of the learner is to find the *most restrictive* grammar which successfully *generates the data* of the language. This requires the learner to find a ranking which correctly enforces the phonotactic restrictions of the language (most restrictive), and a lexicon of underlying forms which, when combined with the ranking, correctly analyzes the alternations of the language (generates the data). Further, the learner should do this while actually constructing and evaluating far fewer than the 384 possible grammars. If the learner must overtly evaluate even half of the possible number of grammars, then it is not likely to scale up well to realistic grammar spaces for full human languages.

3 Overcoming mutual dependence

Learning the languages of this system requires the learner to face the mutual entanglement of the ranking and the lexicon. If a hypothesized grammar fails to produce the data, the learner knows that something about the grammar is incorrect. But this failure does not indicate what is incorrect, or how the grammar ought to be changed. Should the learner change the ranking, or the lexicon, or both? If the current lexicon is suspect, how confident can the learner be of their conclusions about the ranking, and vice-versa? Further, if the learner succeeds in finding a grammar that correctly produces the data, can it be confident that the selected grammar is the most restrictive one?

We propose that the learner first goes through a stage of phonotactic learning. In this stage, it treats each word as an isolated entity, as if it were morphologically unrelated to any other words. During this stage, the learner can get away with a crucial simplifying assumption about the lexicon: it can assume that the underlying form of each word is identical to the surface form.[2] The learner then sets about learning a ranking that maps each form to itself: mapping each such underlying form to its (identical) surface form. Further, the learner obtains the most restrictive such ranking. This kind of mapping is known as a restrictive identity map: it maps all attested surface forms to themselves (identity map), while permitting the fewest number of unobserved forms (restricted). An algorithm for phonotactic learning, Biased Constraint Demotion, is presented in section 4.

The identity map analysis during the phonotactic learning stage allows the learner to temporarily pin down the lexicon part of the mutual dependence. The learner is able to do this because it is not attempting to enforce the use of a single underlying form for a morpheme at this stage; each occurrence of a morpheme is treated separately, with a separate underlying form for that word. The learner will not, in general, be able to determine the entire correct ranking at this stage. But they can get a very good start. In particular, they can learn a ranking enforcing the phonotactics of the language.

Once the learner reaches the alternation learning stage, they become morphologically aware. In particular, they realize that the same morpheme may surface in more than one way. The identity map assumption about the lexicon must be loosened at this point, and the learner must consider different possible underlying forms for morphemes. However, the learner now has some grasp of what the ranking is like, in particular what the phonotactics are. The learner now uses the ranking constructed during phonotactic learning to evaluate different possible underlying forms for the morphemes. When a feature value for a morpheme, in this case voicing of the final root consonant, varies across surface forms, the learner will try underlying forms with each of the feature values. The

---

[2] Alderete and Tesar (2002) suggest a case involving stress-epenthesis interaction where it may be desirable to have phonotactic learning consider non-identical underlying forms. We set aside this issue for the present paper.

one that surfaces correctly in every context, using the phonotactic ranking, will be selected as the underlying form in the learner's lexicon. The process of inferring the underlying forms using the phonotactic ranking is illustrated in section 5.

The mutual dependence between mapping and lexicon is handled by alternately fixing one, and using it to learn about the other. During phonotactic learning, the learner holds the underlying forms fixed, and learns something about the ranking. During alternation learning, the learner holds the phonotactic ranking fixed, and learns about the underlying forms. In the simplest case, the phonotactic ranking is in fact the correct ranking for the language, and alternation learning need only test underlying forms to find the correct one. In more complex cases, phonotactic learning alone cannot reveal the entire correct ranking. In these cases, the learner must modify the underlying forms and then further modify (here, refine) the ranking. Crucially, when the learner further modifies the ranking, it holds on to essential information it learned about the phonotactics. An illustration of this more complex kind of case is given in section 6.

4 Phonotactic learning

4.1 Characterizing restrictiveness: Markedness over faithfulness

The learner needs a way of characterizing restrictiveness in terms of identifiable properties of grammars, so that the learner may select the most restrictive grammar consistent with the data. One commonly made observation is that, in Optimality Theory, restrictiveness correlates with the extent to which markedness constraints dominate faithfulness constraints; having more markedness constraints dominating more faithfulness constraints often results in greater restrictiveness (Sherer, 1994; Demuth, 1995; Gnanadesikan, 1995; van Oostendorp, 1995; Smolensky, 1996). Faithfulness constraints typically act to preserve underlying distinctions, and their dominance will reduce restrictiveness in favor of preserving more types of structures in more environments. Markedness constraints typically act to neutralize underlying distinctions, and their dominance will increase restrictiveness by preventing marked structures from surfacing.

The markedness over faithfulness relationship can be illustrated with three of the constraints from the voicing alternation analysis (to keep the illustration simple, we will not include any intervocalic environments). Three grammars are shown in Table 3. The left column shows the ranking, while the other columns show the forms permitted by the ranking. Each grammar has one faithfulness constraint, IDVoi, and two markedness constraints. The most restrictive grammar, with an inventory of only two of the considered forms, has both markedness constraints dominating the faithfulness constraint; it does not permit voiced obstruents anywhere. The least restrictive grammar, containing all of the considered forms, has the faithfulness constraint dominating both markedness constraints; it preserves voicing in all environments. The intermediate grammar has the faithfulness constraint in between the two markedness constraints; it permits voiced obstruents syllable-initially, but still bans them syllable-finally.

Table 3 Higher faithfulness (IDVoi) yields a less restrictive grammar (more forms).

Ranking	**Inventory**					
{NoVoi, NoSFV} ≫ IDVoi	ta		tat			
NoSFV ≫ IDVoi ≫ NoVoi	ta	da	tat		dat	
IDVoi ≫ {NoVoi, NoSFV}	ta	da	tat	tad	dat	dad

Prince and Tesar (1999) have proposed quantifying the markedness over faithfulness relationship with the 'r-measure'. The r-measure of a constraint hierarchy is determined by adding together, for each faithfulness constraint, the number of markedness constraints that dominate it. Thus, higher r-measures are expected to correlate with greater restrictiveness.

The r-measure gives the learner a concrete, easily computed property of grammars to use. During phonotactic learning, the learner can now attempt to find, from among those rankings that successfully map each of the observed surface forms to themselves, the one with the highest r-measure. The maximization of the r-measure acts a well-defined realization of the learner's mandate to enforce attested phonotactic restrictions.

4.2 Biased Constraint Demotion

Biased Constraint Demotion, or BCD, is an algorithm for inferring the most restrictive ranking consistent with a set of ranking arguments (Prince and Tesar, 1999).[3] It works by attempting to locally maximize the r-measure of the hierarchy it constructs, while ensuring the hierarchy's consistency with the ranking arguments.

A ranking argument is here termed a winner-loser pair, because it consists of a comparison between a candidate deemed grammatical, the winner, and a competing candidate, the loser. For a ranking to be consistent with a winner-loser pair, it must evaluate the winner as more harmonic than the loser. A winner-loser pair indicates what must be true of the ranking via a comparison of the number of violations of each constraint.

Table 4 An example winner-loser pair.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi
/tat/, /-e/	tate ~ tade	W		L	W

---

[3] See also Hayes (1999) for a similar proposal.

Consider the winner-loser pair shown in Table 4, in the format of a comparative tableau (Prince, 2000). To satisfy this pair, a ranking must ensure that the winner, *tate*, is more harmonic than the loser, *tade*. The W mark in the columns of constraints NoVoi and IDVoi indicate that these constraints are violated more by the loser than by the winner. If it were up to a W-marked constraint to choose between the two candidates, they would choose the winner. The L mark in the column for constraint IVV indicates that it favors the loser, and would choose the loser if given the opportunity. What this winner-loser pair indicates is that at least one of NoVoi and IDVoi must dominate IVV. More generally, a winner-loser pair is satisfied by a constraint hierarchy if and only if *at least one* of the W-marked constraints dominates *all* of the L-marked constraints in the hierarchy.

A learner gains information about the correct ranking by accumulating and retaining appropriately informative winner-loser pairs. How the learner obtains appropriate pairs will be discussed below. Given a list of winner-loser pairs, here termed a 'support', the phonotactic learner's job is to find the most restrictive ranking that is consistent with them. That is the role of BCD.

BCD is based upon the Recursive Constraint Demotion (RCD) algorithm for learning constraint rankings (Tesar and Smolensky, 1994, 2000). BCD builds a constraint hierarchy top down, first placing the highest-ranked constraints in the hierarchy, then the next-highest constraints, and so forth. BCD, like RCD, enforces consistency with the winner-loser pairs of the support by making sure not to place a constraint into the hierarchy when it would mistakenly prefer a loser in one of the pairs. The 'bias' in BCD is realized by only placing markedness constraints into the hierarchy until no more can be placed without violating consistency with the support.

The actions of BCD can be illustrated using the support in Table 5.

Table 5 The support.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi
/dat/	dat ~ tat	L			W
/tat/, /-e/	tate ~ tade	W		L	W

On the first pass through the support, BCD determines what constraints are available to be placed into the ranking. Constraints are available if they have no L's in their column. Ranking a constraint with an L, like IVV, would violate the corresponding winner-pair; in this case, ranking IVV at the top would cause *tade* to (incorrectly) beat *tate*. In this support, two constraints are available for ranking, NoSFV and IDVoi. BCD then determines if any of the available constraints are markedness constraints. In this case, NoSFV is a markedness constraint, so only it is placed into the ranking at this point. IDVoi, a faithfulness constraint, is not placed into the ranking, as part of the effort to maximize the r-measure. After the first pass, the hierarchy has the constraint NoSFV at the top.

BCD then eliminates from further consideration any winner-loser pairs that are accounted for by the hierarchy thus far constructed. Pairs eliminated from consideration are not deleted from the support. A pair is accounted for if the hierarchy ensures that the winner will beat the loser, which will be the case once a constraint preferring the winner has been placed into the hierarchy. NoSFV has no W's in its column; it prefers none of the winners, and so cannot account for any pairs. Thus, after the first pass, both pairs remain in consideration.

On the next pass, only IDVoi is available for ranking. No markedness constraints are available, so IDVoi, a faithfulness constraint, must be placed into the second stratum of the hierarchy, yielding a partial hierarchy NoSFV ≫ IDVoi.

BCD then observes that placing IDVoi into the hierarchy is sufficient to account for both winner-loser pairs, so both may be eliminated from consideration. Once that is done, no L marks remain in the list, so both remaining constraints, NoVoi and IVV, are now available to be placed into the ranking. The result is the final constraint hierarchy returned by BCD, given in (4).

(4) NoSFV ≫ IDVoi ≫ {NoVoi, IVV}

4.3 Error-driven learning

The data provided to the learner do not contain winner-loser pairs; they only contain grammatical surface forms. The winners are full structural descriptions, combining both inputs and surface forms, but will match attested surface forms, with each winner having a surface form corresponding to actually observed data. During phonotactic learning, the learner constructs a full winner by simply copying the surface form for use as the input as well (during alternation learning, constructing the appropriate input for a winner is not as simple, as will be discussed below).

The learner must construct losers that will form winner-loser pairs providing useful information about the ranking. The informativeness of a pair is not a property it holds in isolation, but one determined relative to other pairs already held in a support by the learner. In other words, a pair is informative if it tells the learner something that it does not already know.

The construction of informative pairs can be done using error-driven learning (Tesar, 1998). When a learner processes an observed word, they take the input form for the observed word and compute the surface form predicted by their current grammar (the predicted surface form is the surface form of the optimal candidate). If the predicted surface form is identical to the observed surface form, no error has occurred (in the sense of error-driven learning), and the learner does not construct a winner-loser pair. However, if the predicted surface form is different from the observed surface form, an error has occurred. The learner then constructs a new winner-loser pair. The winner is constructed using the observed surface form, while the loser is constructed using the surface form predicted by the learner's current grammar. This newly constructed winner-loser pair is then added to the support, and BCD is re-applied to the support to derive a new constraint hierarchy.

In this way, the learner proceeds through a series of grammars. A new grammar is constructed each time the learner adds a winner-loser pair to the support. Initially, the learner has an empty support, resulting in an initial grammar in which all markedness constraints dominate all faithfulness constraints. The key repository of information retained by the learner is the support; it always contains the ranking arguments that the learner has constructed based upon the observed data. BCD is used to generate a working grammar hypothesis based upon the support. The technique of accumulating winner-loser pairs one at a time via error-driven learning, and constructing a new ranking at each step, is known as Multi-Recursive Constraint Demotion, or MRCD (Tesar, 1997a, 2000).

4.4 Phonotactic learning with language D

We will illustrate the learning proposal of this paper using language D of the system. Here, we illustrate the phonotactic learning stage. The surface forms of language D were given in Table 2, and are repeated here for convenience.

(5) tat, dat, tate, tade, date, dade

The learner starts out with an empty support. Applying BCD to an empty support yields the initial constraint hierarchy shown in (6).

(6) {NoVoi, IVV, NoSFV} $\gg$ {IDVoi}

If the first surface form processed by the learner is *tat*, the learner will use /tat/ as the underlying form, and observe that no constraints are violated by this candidate. That means that it cannot lose; the optimal output for /tat/ will always be *tat*. This is fully uninformative; it tells the learner nothing about the ranking. Because the optimal candidate matches the observed form, no winner-loser pair is formed, and the support is unchanged.

The next word processed by the learner is *dat*. Here, an error occurs, because the optimal candidate, *tat*, does not match the observed form, *dat*. This is shown in the tableau in Table 6.

Table 6 An error: the candidate for the observed form, *dat*, loses to *tat*.

/dat/	NoVoi	NoSFV	IVV	IDVoi
dat	*			
tat				*

The error causes the learner to construct a winner-loser pair, using the observed surface form to build the winner, and the incorrectly optimal surface form to build the loser. This results in a support with one pair, shown in Table 7. BCD is then applied to the support, yielding the constraint hierarchy in (7).

Table 7 The support after the first error.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi
/dat/	dat ~ tat	L			W

(7) {NoSFV, IVV} ≫ IDVoi ≫ NoVoi

In phonotactic terms, the learner has learned that voiced obstruents must be permitted in onset position. The learner’s current hierarchy is still enforcing all the restrictions allowable by the data; it bans voiced obstruents from codas, and requires obstruents to be voiced intervocalically. No positive data have yet contradicted those restrictions.

Suppose the next word processed by the learner is *tate*. This causes another error, as shown in Table 8.

Table 8 The second error: the observed *tate* loses to *tade*.

/tate/	NoSFV	IVV	IDVoi	NoVoi
tate		*		
tade			*	*

The learner constructs a second winner-loser pair from this comparison and adds it to the support. The resulting support, shown in Table 9, yields the constraint hierarchy in (8).

Table 9 The support after the second error.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi
/dat/	dat ~ tat	L			W
/tate/	tate ~ tade	W		L	W

(8) NoSFV ≫ IDVoi ≫ {NoVoi, IVV}

This is the final result of phonotactic learning. The learner will not encounter any further errors on the forms of language D; this constraint hierarchy maps all of the observed surface forms to themselves. The restricted identity map has been obtained, as reflected by the dominated status of IDVoi. This is due to the bias of BCD, not to any direct observations. The constraint hierarchy at this point captures the learner's knowledge of the phonotactics: voiced obstruents are banned syllable-finally.

5 Alternation learning

5.1 Identifying alternating features

Once the learner reaches the alternation learning stage, they are morphologically aware. They perceive the words not as isolated, monolithic forms, but as polymorphemic forms that share morphemes. The data for language D, which the learner structured as shown in (5) during phonotactic learning, are now structured by the learner as in Table 10.

Table 10 The fully segmented forms of language D.

Bare Root	Root + Suffix
tat$_1$	tad$_1$-e$_5$
tat$_2$	tat$_2$-e$_5$
dat$_3$	dad$_3$-e$_5$
dat$_4$	dat$_4$-e$_5$

The identity map assumption about underlying forms, used during phonotactic learning, is no longer sustainable here. That is because some of the morphemes alternate. The learner is now aware that morpheme #1 surfaces as *tat* in the bare root context and as *tad* in the suffixed context. A single underlying form cannot be identical to both. However, the learner proposed here still attempts to adhere to the identity map view as much as possible. Specifically, it compares the different surface forms for a morpheme, and attempts to identify those features which alternate. The learner assumes that any features that do not alternate across surface forms are represented underlyingly as they appear in the surface forms. This is a continuing reflection of the restricted identity map hypothesis. Thus, for morpheme #2, which surfaces consistently as *tat* (it does not alternate), the learner presumes the underlying form to be /tat/$_2$, and does not worry further about that morpheme. For morpheme #1, the learner restricts its uncertainty about the underlying form to the voicing on the final consonant; the rest of the underlying form is assumed to be as it appears on the surface.

The learner initially analyzes a paradigm as follows. For each morpheme, it examines all of the surface forms, and establishes a correspondence between them. For the system discussed in this paper, establishing the correspondence is a simple matter. The only thing that can alternate is voicing on consonants, so every surface variant of a morpheme has the same number of segments, and they can be aligned first segment to first segment, etc. In some more complex linguistic systems, in particular systems permitting epenthesis and deletion, establishing the correspondence across surface variants will be more complex.[4] Once the correspondence is established, the learner starts constructing an underlying form for the morpheme by including all feature values that do not vary across the surface variants. For morpheme #1, this part of the underlying form would look something like /taT/, where /T/ does not yet have a voicing feature value specified.

Having fixed the invariant features of the morpheme, the learner then creates several underlying form hypotheses for the morpheme, a different one for each observed feature value that varies across surface forms. For morpheme #1, the underlying form hypotheses are $/tat/_1$ and $/tad/_1$. These hypotheses will be tested using the phonotactic mapping after the learner has finished analyzing the paradigm.

A full analysis of the paradigm yields the set of underlying form hypotheses shown in Table 11.

Table 11 Underlying form hypotheses for the morphemes of the paradigm.

Morpheme	UF Hypotheses
#1	$/tat/_1$, $/tad/_1$
#2	$/tat/_2$
#3	$/dat/_3$, $/dad/_3$
#4	$/dat/_4$
#5	$/\text{-}e/_5$

Morphemes 2, 4, and 5 do not alternate, so the learner presumes the single underlying form constructed for each to be correct. Further work must be done with respect to morphemes 1 and 3, however, to choose from among the multiple hypotheses.

---

[4] The literature on string and structure comparison, going back at least to 1983 (Sankoff and Kruskal, 1983), will be relevant to this point. Note that this kind of cross-form comparison will be necessary for any paradigmatic analysis, and is required for grammatical analyses that employ output-output correspondence constraints.

Note that, for this example, the hypothesized underlying forms are identical to the surface allomorphs for each morpheme. We resist the temptation to simply define the set of underlying form hypotheses as the set of surface allomorphs, because more complex cases may require underlying forms that are not identical to any single surface variant (Schane, 1974).

5.2 Using phonotactics to test underlying forms

The learner now uses the ranking constructed during phonotactic learning to evaluate to underlying form hypotheses. This is where the interaction between the ranking and lexicon is directly invoked. For each alternating morpheme, the learner constructs linguistic inputs for each of the words containing the morpheme. The words containing morpheme #1 are *tat* and *tade*. To test underlying form /tat/$_1$, the learner constructs two linguistic inputs: /tat/ for *tat*, and /tat+e/ for *tade*. The learner also constructs linguistic inputs for the two words using the other underlying form hypothesis for morpheme #1: /tad/ for *tat*, and /tad+e/ for *tade*.

Pater (2000) suggests that the learner construct additional winner-loser pairs for each underlying form hypothesis at this point, testing each for consistency with the winner-loser pairs accumulated during phonotactic learning. We suggest that the learner need not immediately construct winner-loser pairs; the learner can simply test the inputs for each underlying form hypothesis, and see if the phonotactic ranking produces the correct surface forms. If the phonotactic ranking is in fact a sufficient mapping for the alternation, this will be adequate to test the underlying form hypotheses. If the phonotactic ranking is insufficient, and requires further refinement, then constructing further winner-loser pairs will in fact be required; this is discussed further in section 6.

Recall from section 4.4 the ranking constructed for Language D during phonotactic learning, repeated as (9).

(9) NoSFV $\gg$ IDVoi $\gg$ {NoVoi, IVV}

The learner tests the /tat/ hypothesis for morpheme #1 by using the phonotactic ranking to obtain the optimal output for each of the two inputs. Table 12 shows the results of testing the underlying form hypothesis /tat/$_1$ for morpheme #1. While it succeeds in the bare root environment, in the suffixed environment the underlying form, combined with the phonotactic ranking, yields the predicted output *tate*, which does not match the attested output *tade*.

Table 12 The underlying form /tat/$_1$ fails in the suffixed environment.

Input	Ranking Output	Observed Output	Result
/tat/	tat	tat	succeeds
/tat+e/	tate	tade	fails

Table 13 shows the results of testing the underlying form hypothesis /tad/$_1$ for morpheme #1. This hypothesis succeeds in both environments. Thus, the learner can successfully choose /tad/$_1$ as the correct underlying form for morpheme #1.

Table 13 The underlying form /tad/$_1$ succeeds in both environments.

Input	Ranking Output	Observed Output	Result
/tad/	tat	tat	succeeds
/tad+e/	tade	tade	succeeds

Similar reasoning will allow the learner to correctly select /dad/$_3$ as the underlying form for morpheme #3. Observe that the learner has learned the correct grammar efficiently. The learner constructed two winner-loser pairs during phonotactic learning, which can be counted as evaluating two possible grammars. During alternation learning, the learner evaluated two possible underlying forms for each of the two alternating morphemes, which can be counted as evaluating four additional grammars, for a total of 6 evaluated grammars, well below the 384 total possible grammars.

This learner succeeds in solving the mutual dependence of mapping and lexicon for this system by focusing on phonotactics first. The identity map hypothesis does not give the learner the correct lexicon, but gives it a good enough set of starting inputs to allow the learner to recover the phonotactics. In the case of language D, surface forms like *tate* demonstrate that intervocalic voicing is not active in the language. That information, encapsulated in the support, is critical to alternation learning. When the learner considers hypotheses for the underlying form of a morpheme like #1, the choice can be made on the basis of the processes necessary for the analysis using each underlying form. The choice of /tat/ as the underlying form for the {tat, tade} alternation required an intervocalic voicing analysis; that analysis is inconsistent with the phonotactics. The choice of /tad/ required a syllable-final devoicing analysis; that analysis is consistent with the phonotactics. The form *tate*, while not part of the {tat, tade} alternation, plays a crucial role in solving the analysis, via its effect on phonotactic learning.

6 A complication: Multiple unfaithful mappings

6.1 The problem: Tied faithfulness constraints

One simplifying property of the system for voicing alternations described above is that there is just one way of being unfaithful in order to satisfy markedness constraints. If preserving the input faithfully would result in a marked output, the only way the system can construct a more harmonic competitor is to change the voicing specification of one or more consonants. Adapting a familiar usage, we will refer to such unfaithful mappings as 'repair strategies', even though nothing is literally repaired. Therefore, if the phonotactics indicate that a certain structure shouldn't be present in the output, the learner knows in advance how to fix it: change the voicing of a consonant in the relevant structure.

More complex linguistic systems typically have multiple repair strategies. A consequence of this is that phonotactics alone cannot determine the complete constraint ranking for the language. Further adjustment of the ranking will have to take place during alternation learning, in order for the learner to obtain the correct grammar. We argue here that this sort of more complex relationship between mapping and lexicon does not diminish the value of phonotactic learning for the learning of alternations.

We will illustrate the proposal using the same data as before, from language D. However we modify the linguistic system giving rise to the data by adding a second repair strategy to GEN, that of deleting segments (formally, allowing input segments to have no output correspondent). Introducing this further freedom into GEN requires also introducing a constraint to regulate it. Here, the relevant constraint is Max, which is violated by any input segment that does not have an output correspondent. We now have a system for voicing alternations in which potential marked structures, such as syllable-final voiced obstruents, can be avoided either by changing the voicing specification of a consonant or by deleting a segment entirely.[5]

This change to the linguistic system has a visible impact even on phonotactic learning. The support constructed by phonotactic learning with the new system is shown in Table 14, and the resulting ranking is given in (10).

---

[5] While the plausibility of using deletion to avoid voicing markedness, as opposed to other kinds such as place markedness, is questionable (Lombardi, 1995), it is convenient for the purposes of this illustration, allowing us to introduce a second repair strategy with a minimum amount of alteration of the earlier linguistic system.

Table 14 The winner-loser pairs from phonotactic learning.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi	Max
/dat/	dat ~ at	L				W
/dat/	dat ~ tat	L			W	
/tate/	tate ~ tade	W		L	W	
/tate/	tate ~ tae			L		W

(10) NoSFV $\gg$ {IDVoi, Max} $\gg$ {NoVoi, IVV}

Notice that the phonotactic ranking has IDVoi and Max tied in the second stratum. These two faithfulness constraints correspond to the two repair strategies: altering voicing feature values violates IDVoi, while deleting a segment violates Max. One of these constraints will be violated when necessary to ensure satisfaction of the top-ranked NoSFV. But which one? A full determine ranking would have one of these faithfulness constraints dominate the other, with the lower (dominated) one indicating the preferred repair strategy. But the repair strategy is not directly indicated by phonotactics. The learner can observe that voiced segments never occur syllable-finally. They cannot directly observe how a potentially offending input is altered. However, alternations can indicate that. Thus, while phonotactic learning has done much of the work in establishing the ranking, it will be up to alternation learning to determine the relative ranking of IDVoi and Max.

6.2 Refining the ranking

Following the description of the learner in section 5, once the learner becomes morphologically aware, it will construct underlying form hypotheses for each of the alternating morphemes, and test those hypotheses using the phonotactic ranking. In the new system, the phonotactic ranking cannot decide the matter as it stands. Using morpheme #1, {tat_1, tad_1-e}, to illustrate again, neither of the underlying form hypotheses correctly generates all of the surface forms. The forms using the hypothesis $/\text{tat}/_1$ for morpheme #1 behave as they did previously, with the hypothesis failing in the suffix environment. The hypothesis $/\text{tad}/_1$ now fails in the bare root environment, because, under the phonotactic ranking, the outputs *ta* and *tat* tie for optimality; this is summarized in Table 15. We treat a tie as a failure because the input has not been assigned a unique output. The uniqueness assumption is motivated by standard restrictiveness concerns (non-uniqueness could always be inferred from positive evidence). This is precisely the repair strategy conflict: to match the attested surface form, the learner needs a ranking enforcing devoicing rather than deletion, and in the phonotactic ranking, the two are tied.

Table 15 The hypothesis /tad/$_1$ fails in the bare root environment.

Input	Ranking Output	Observed Output	Result
/tad/	tat, ta	tat	fails
/tad+e/	tade	tade	succeeds

When the learner fails to match an observed surface form, as happens for both hypotheses here, the learner can treat the mismatch in the same way that it treated a mismatch during phonotactic learning: as an error in the sense of error-driven learning, to be corrected by further learning. The same MRCD procedure used during phonotactic learning can be applied here to further modify the constraint ranking, via the accumulation of further winner-loser pairs. The learner will separately attempt to refine the ranking to support each underlying form hypothesis. The attempt that succeeds will indicate to the learner both the correct underlying form and the correct ranking.

The learner proceeds by making several copies of the support, one for each underlying form hypothesis (/tat/$_1$ and /tad/$_1$). At first, the support contains only phonotactic winner-loser pairs. The learner will be constructing new winner-loser pairs, with morphologically analyzed candidates, and adding them to the phonotactic winner-loser pairs, at each step applying BCD in the usual way to derive a new (refined) ranking. Then, for each underlying form hypothesis, MRCD will be applied to the surface forms of the alternation until either a ranking is found giving the correct surface forms, or inconsistency is reached.

A set of winner-loser pairs is inconsistent if there is no ranking that can simultaneously satisfy all of them. One beneficial side-effect of BCD (and all variants of RCD) is that it will quickly determine if a set of winner-loser pairs is inconsistent without any special effort (Tesar, 1997b). When inconsistency is detected, the learner knows that further altering the ranking will do no good, suggesting that there must be something wrong with the lexicon (Kager, 1999; Tesar, Alderete, Horwood, Merchant, Nishitani, and Prince, to appear).

Consider first the underlying form hypothesis /tat/$_1$. In the suffixed environment, the observed surface form is *tade*, while the hypothesis currently generates *tate*. The learner constructs a new winner-loser pair in the usual way, adopting the observed form as the winner, and the form generated by the learner as the loser. This winner loser pair is added to the pairs constructed during phonotactic learning, yielding the support shown in Table 16.[6]

---

[6] This creates a situation where the support simultaneously contains pairs with morphologically unanalyzed words and pairs with morphologically analyzed words. This does not create a problem here, because a morphologically unanalyzed word is a possible morpheme. In other situations it may be necessary to change existing pairs when new

Table 16 Inconsistent winner-loser pairs for /tat/$_1$.

	Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi	Max
(a)	/dat/	dat ~ at	L				W
(b)	/dat/	dat ~ tat	L			W	
(c)	/tate/	tate ~ tae			L		W
(d)	/tate/	tate ~ tade	W		L	W	
(e)	/tat/$_1$, /-e/$_5$	tad-e ~ tat-e	L		W	L	

This support is inconsistent. This can be seen by observing that the last two rows make precisely opposite requirements of the ranking. The inconsistency will be detected by the learner when it applies BCD to build a ranking. The learner will first rank NoSFV, which accounts for no winner-loser pairs. The learner next will place Max in the ranking, which accounts for winner-loser pairs (a) and (c). That leaves three constraints, each of which prefers the loser in at least one of the remaining winner-loser pairs, in particular pairs (d) and (e). The learner has more constraints to rank, but cannot place any of them into the ranking without contradiction. This tells the learner that the support is inconsistent, meaning that the problem lies with the underlying forms being used.

The learner now will consider the underlying form hypothesis /tad/$_1$. In the bare root environment, the observed form is *tat*, while the hypothesis currently generates a tie between *tat* and *ta*. A tie is not sufficient for the learner, so it selects the generated form, *ta*, that does not match the observed one, and constructs a new winner-loser pair. The new pair is then added to the support from phonotactic learning, yielding the support shown in Table 17. This support is consistent; applying BCD yields the constraint hierarchy in (11). Because this support is consistent, the learner adopts the underlying form hypothesis, /tad/$_1$, into the lexicon, and also adopts the support and the corresponding constraint hierarchy.

Table 17 The hypothesis /tad/$_1$ yields a consistent support.

Lexicon	Winner ~ Loser	NoVoi	NoSFV	IVV	IDVoi	Max
/dat/	dat ~ at	L				W
/dat/	dat ~ tat	L			W	
/tate/	tate ~ tae			L		W
/tate/	tate ~ tade	W		L	W	
/tad/$_1$	tat ~ ta				L	W

---

underlying form hypotheses are made. This is discussed further in (Tesar, Alderete, Horwood, Merchant, Nishitani, and Prince, to appear).

(11) NoSFV $\gg$ Max $\gg$ IDVoi $\gg$ {NoVoi, IVV}

Observe that the new hierarchy has Max dominating IDVoi. Now, when enforcement of NoSFV forces the optimal candidate to be unfaithful to the underlying form, as with /tad/$_1$, the unfaithfulness will take the form of devoicing, in violation of IDVoi, rather than deletion, which would violate the higher-ranked Max. The learner has succeeded in solving the more complex mutual dependence between mapping and lexicon, correctly selecting the syllable-final devoicing analysis. The learner's algorithm effectively rejected an intervocalic-voicing analysis as inconsistent with the phonotactics, and rejected a deletion-based strategy for avoiding syllable-final voiced obstruents as inconsistent with the alternation data.

7 Discussion

7.1 Mutual dependence

In the example systems examined here, phonotactic learning is able to determine the mapping up to the point of choosing among repair strategies. The inability to do the latter is reflected in unresolved ties between faithfulness constraints in the hierarchy. The learned phonotactic information, in particular the support and the corresponding ranking, allow the learner to analyze alternations, testing different possible underlying forms for consistency. The alternations provide information about repair strategies, allowing the ranking to be further refined.

The resolution of the mutual dependence has a back-and-forth quality, using a hypothesized lexicon of underlying forms to determine a better hypothesis mapping, and vice-versa. The initial identity map assumption about underlying forms during phonotactic learning allows the learner to learn a lot about the ranking. It cannot learn everything about the ranking at this stage, because that requires the correct underlying forms. But it can learn enough to determine the phonotactics. Put another way, the learner uses an initial guess about the lexicon to gain insight into the mapping. That ranking information allows the learner to evaluate hypotheses for underlying forms for morphemes, once the learner begins morphologically analyzing words. If a single underlying form hypothesis for a morpheme yields all of the correct surface forms, using the phonotactic ranking, then that hypothesis is selected as correct. In other words, the learner uses the phonotactic mapping to gain insight into the underlying forms. When multiple repair strategies are possible, the learner must learn more about the ranking from alternations. The learner evaluates the underlying form hypotheses by constructing winner-loser pairs for each, and testing them for consistency with the phonotactic mapping. The winner-loser pairs for the consistent underlying form allow the learner to refine the ranking to reflect information about repair strategies. Here, the learner is using the more detailed, morphologically aware information it has constructed about underlying forms to gain further insight into the mapping.

7.2 The continuing nature of phonotactic learning

The characterization of learning given in the examples above has a kind of discontinuity in it, at the point where the learner becomes morphologically aware. At that point, phonotactic learning ends, the learner analyzes words morphologically, and then alternation learning begins. We expect the reality of human language learning to be more fluid. The actual morphological analysis of words can be expected to interact with alternation learning. Further, it is at least plausible that the learner proceed with alternation learning on some morphologically analyzed paradigms of words, while other sections of the lexicon remain morphologically unanalyzed. This could give rise to rather interesting intermediate stages, in which the learner possesses a support containing winner-loser pairs from both morphologically analyzed and unanalyzed words.

The enforcement of phonotactic restrictiveness is the responsibility of Biased Constraint Demotion (BCD). Restrictiveness is enforced in part by the fact that, every time the support is changed, BCD is applied to the entire support to reconstruct an entire ranking. Whenever BCD is invoked, it acts to find the most restrictive ranking consistent with the current support. This is true even during alternation learning. In section 6.2, when a new winner-loser pair was added to the support to test an underlying form hypothesis, BCD was invoked to construct the new constraint ranking. Further, if the learner were to encounter data that required loosening phonotactic restrictions later on, after alternation learning had begun, the use of BCD would accommodate this.[7] Thus, in a certain sense, phonotactic learning is on-going, throughout language learning. It is the treating of words as isolates, and the assumption of surface-identical underlying forms, that fades away as morphological analysis takes place.

7.3 Complications

The ultimate viability of this account of human language learning will depend in part on its ability to scale up to more complex linguistic analyses. There are several key properties of analyses that have the potential to greatly increase the amount of computational effort exerted by the learner.

As mentioned in section 5.1, we wish to allow the learner the opportunity to consider underlying forms that do not exactly correspond to any single surface allomorph. However, if a morpheme has numerous feature specifications that vary across allomorphs, and the learner separately considers all attested values of each feature, there could be an exponential explosion in the number of underlying form hypotheses (the product of the number of attested values of each feature). This

---

[7] We are indebted to Janet Pierrehumbert for valuable discussion of this point.

can happen even with a single feature type which occurs in multiple positions within a morpheme. For instance, if we enhanced the system of voicing alternations used above to include prefixes, it would be possible (using inter-vocalic voicing) to get paradigms like {tat, tad-e, o-dat}, where the root alternates in the voicing of both the initial and the final consonants. Creating a separate underlying form hypothesis for each combination of attested feature values would give 2 * 2 = 4 possible hypotheses for the one morpheme: /tat/, /dat/, /tad/, and /dad/.

One might approach this by treating the learning of each positional occurrence of each feature within a morpheme independently. Rather than constructing a single underlying form immediately, the learner employs a separate underlying form for each environment, corresponding to each allomorph, and then fixes the value of a single corresponding feature as identical across the forms. If the learner could separately learn each alternating feature in this way, it would avoid the combinatorial explosion of underlying form hypotheses. This simple approach might not be feasible in cases where the features in the different positions can interact. Further complications arise when different feature types interact.

Another challenge is the existence of words with more than one alternating morpheme. While the system of voicing alternations above conveniently had only alternating roots, more complex systems contain words with interacting, alternating morphemes. In such a case, the underlying forms of several morphemes can be mutually dependent, requiring the learner to simultaneously work on the underlying forms of several morphemes, rather than being able to approach them one at a time. A related challenge is posed by systems including alternations that are less phonotactically transparent. In the latter case, the benefits of phonotactic learning are diminished; the phonotactic ranking provides little insight into the processes driving the alternations, and alternation learning must do more of the work. For some recent work examining the learning of underlying forms in a system exhibiting these challenging properties, specifically a system allowing both predictable and lexical stress grammars, see (Tesar, Alderete, Horwood, Merchant, Nishitani, and Prince, to appear).

Acknowledgments

The authors wish to acknowledge John Alderete, Adrian Brasoveanu, Jessica Hughes, Nazarré Merchant, Joe Pater, Janet Pierrehumbert, and the participants of the Rutgers Optimality Reading Group for valuable discussions on this material. This material is based upon work supported by the National Science Foundation under Grant No. BCS-0083101 to Bruce Tesar and Alan Prince. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Any errors are the sole responsibility of the authors.

References

Alderete, John, and Tesar, Bruce. 2002. Learning covert phonological interaction: An analysis of the problem posed by the interaction of stress and epenthesis. Technical Report RuCCS-TR-72, Rutgers Center for Cognitive Science, Rutgers University. ROA-543.

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

Baker, C. Lee. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10: 533-581.

Demuth, Katherine. 1995. Markedness and the development of prosodic structure. *Proceedings of the North East Linguistics Society 25*, ed. by Jill Beckman: 13-25. Amherst, MA: GLSA, University of Massachusetts. ROA-50.

Gnanadesikan, Amalia. 1995. Markedness and faithfulness constraints in child phonology. Ms., University of Massachusetts, Amherst, MA. ROA-67.

Goldsmith, John. 1991. Phonology as an intelligent system. *Bridges Between Psychology and Linguistics: A Swarthmore Festschrift for Lila Gleitman*, ed. by Donna Jo Napoli and Judy Kegl: 247-267. Hillsdale, NY: Lawrence Erlbaum Associates.

Hayes, Bruce. 1999. Phonological acquisition in Optimality Theory: The early stages. Ms., UCLA. ROA-327.

Ito, Junko, and Mester, Armin. 1997. Sympathy Theory and German truncations. *University of Maryland Work Papers in Linguistics 5: Selected Phonology Papers from the Hopkins Optimality Theory Workshop 1997 / University of Maryland Mayfest 1997*, ed. by Viola Miglio and Bruce Moren: 117-139. ROA-211.

Kager, Rene. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.

Kiparsky, Paul. 1980. Vowel harmony. Ms., MIT.

Kisseberth, Charles. 1970. On the functional unity of phonological rules. *Linguistic Inquiry* 1: 291-306.

Lombardi, Linda. 1995. Why place and voice are different: Constraint interactions and feature faithfulness in Optimality Theory. Ms., University of Maryland, College Park. ROA-105.

Lombardi, Linda. 1999. Positional Faithfulness and Voicing Assimilation in Optimality Theory. *Natural Language & Linguistic Theory* 17(2): 267-302.

Pater, Joe. 2000. Unpublished course handout from Ling 751, University of Massachusetts, Amherst.

Prince, Alan. 2000. Comparative tableaux. Ms., Rutgers University, New Brunswick. ROA-376.

Prince, Alan, and Smolensky, Paul. 1993. Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University, New Brunswick, and the University of Colorado, Boulder. ROA-537.

Prince, Alan, and Tesar, Bruce. 1999. Learning phonotactic distributions. Technical Report RuCCS-TR-54, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick. ROA-353.

Sankoff, David, and Kruskal, Joseph (Eds.). 1983. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA.

Schane, Sanford. 1974. How abstract is abstract? *Papers from the Parasession on Natural Phonology*: 297-317. Chicago Linguistics Society.

Sherer, Tim. 1994. *Prosodic phonotactics.* Unpublished PhD. thesis. University of Massachusetts, Amherst.

Smolensky, Paul. 1996. The initial state and "richness of the base" in Optimality Theory. Technical Report JHU-CogSci-96-4, The Johns Hopkins University, Baltimore, MD. ROA-154.

Sommerstein, Alan H. 1974. On phonotactically motivated rules. *Journal of Linguistics* 19: 71-94.

Tesar, Bruce. 1997a. Multi-Recursive Constraint Demotion. Ms., Rutgers University, New Brunswick. ROA-197.

Tesar, Bruce. 1997b. Using the mutual inconsistency of structural descriptions to overcome ambiguity in language learning. *Proceedings of the North East Linguistic Society 28* Vol. 1, ed. by Pius N. Tamanji and Kiyomi Kusumoto: 469-483. Amherst, MA: GLSA, University of Massachusetts.

Tesar, Bruce. 1998. Error-driven learning in Optimality Theory via the efficient computation of optimal forms. *Is the Best Good Enough? Optimality and Competition in Syntax*, ed. by Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha Jo McGinnis and David Pesetsky: 421-435. Cambridge, MA: MIT Press.

Tesar, Bruce. 2000. Using inconsistency detection to overcome structural ambiguity in language learning. Technical Report RuCCS-TR-58, Rutgers Center for Cognitive Science, Rutgers University, New Brunswick, NJ. ROA-426.

Tesar, Bruce, Alderete, John, Horwood, Graham, Merchant, Nazarre, Nishitani, Koichi, and Prince, Alan. to appear. Surgery in language learning. *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics*, ed. by G. Garding and M. Tsujimura. Somerville, MA: Cascadilla Press.

Tesar, Bruce, and Smolensky, Paul. 1994. The learnability of Optimality Theory. *Proceedings of the Thirteenth West Coast Conference on Formal Linguistics*, ed. by Raul Aranovich, William Byrne, Susanne Preuss and Martha Senturia: 122-137. CSLI.

Tesar, Bruce, and Smolensky, Paul. 2000. *Learnability in Optimality Theory.* Cambridge, MA: MIT Press.

van Oostendorp, Marc. 1995. *Vowel quality and phonological projection.* Unpublished PhD. thesis. Tilburg University. ROA-84.