

# Contrast Analysis Aids the Learning of Phonological Underlying Forms<sup>\*</sup>

John Alderete<sup>1</sup>, Adrian Brasoveanu<sup>2</sup>, Nazarré Merchant<sup>2</sup>, Alan Prince<sup>2</sup>, and Bruce Tesar<sup>2</sup>

<sup>1</sup>Simon Fraser University and <sup>2</sup>Rutgers University

## 1 Introduction

One of the many challenges to be faced in explaining language learning is the interdependence of the phonological mapping and the phonological underlying forms for morphemes (Albright and Hayes 2002; Hale and Reiss 1997; Tesar and Smolensky 2000; Tesar et al. 2003). The learner must attempt to infer both simultaneously, based on the surface forms of a language.

The present paper presents evidence in support of the idea that observations about surface contrast can inform the learner about the content of underlying forms. More specifically, contrasting outputs for morphemes in a given environment can provide information about underlying forms. We present one way of capitalizing on such information in a learning algorithm, and show how contrast information can combine with phonotactic information to aid learning.

## 2 The Learning Situation

### 2.1 Phonotactic Learning

Phonotactic learning is a proposed early stage of learning in which learner treats words as isolated forms, without any attempt to relate different occurrences of morphemes to each other (Hayes 2004; Prince and Tesar 2004). The goal of the phonotactic learner is to find the most restrictive constraint ranking that accommodates all of the attested forms. One way to attempt this (independently proposed by Hayes (2004) and Prince and Tesar (2004)) is for the learner to assume an underlying form for each word that is identical to the surface form of the word, and search for a ranking that maps each attested word to itself, while mapping as few unattested words as possible to themselves. One algorithm that has been proposed for phonotactic learning is the Biased Constraint Demotion (BCD) algorithm (Prince and Tesar 2004).

Phonotactic learning can determine significant properties of the constraint ranking. However, it cannot always determine the complete ranking. Some aspects of phonological mappings are not apparent in the phonotactics alone, and in fact it is possible for several distinct phonological mappings to have identical phonotactic distributions. The inadequacy of phonotactic information alone can be illustrated with lexical stress. Suppose we have a language containing both initially stressed words, like *páka*, and finally stressed words, like *paká*. A phonotactic learner will infer from this a mapping which maps /páka/ → *páka* and /paká/ → *paká*. But what about the input /paka/?<sup>1</sup> A stress must be assigned but the phonotactics don't indicate the default stress position. The phonotactics aren't

---

<sup>\*</sup> The authors thank Jane Grimshaw, John McCarthy, Joe Pater, and the Rutgers Optimality Reading Group for helpful comments and discussion. This material is based upon work supported by the National Science Foundation under Grant No. BCS-0083101, and by NIH NRSA Training Grant No. 1-T32-MH-19975-05. 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 or of the National Institutes of Health. Any errors are the responsibility of the authors.

<sup>1</sup> While such an input isn't identical to any possible surface form, it is part of the rich base, and could easily arise from a combination of morphemes with underlying forms motivated by surface alternations.

(1)       $r1[?,-]$     $r2[?,?]$     $r3[+,-]$     $r4[+,+]$                        $s1[-,-]$     $s2[?,-]$     $s3[?,?]$

### 2.3 Joint Ranking/Lexicon Learning

Phonotactic learning can determine part of the ranking, but not the part requiring alternation information. Initial lexicon construction can determine part of the lexicon, but it cannot set the features that alternate. The rest of the ranking and lexicon must be learned via other information.

One approach to combining ranking and lexical information is to propose different lexical hypotheses (assignments of values to the features of the underlying forms), and test the hypotheses for consistency (Kager 1999). A set of data (including the underlying forms) is inconsistent if there does not exist any ranking of the constraints which will map the hypothesized underlying forms to their surface forms. The surgery learning algorithm (Tesar et al. 2003) uses Biased Constraint Demotion to test lexical hypotheses with inconsistency detection (Tesar 2004a). BCD efficiently determines if a ranking exists for a given lexical hypothesis, without having to evaluate lots of rankings.

The surgery approach overcomes the complexity of the number of possible rankings, but it still has to select and evaluate lexical hypotheses. Initial lexicon selection can set non-alternating features, but the resulting space of lexical hypotheses still consists of all possible combinations of values for the unset underlying features. The size of that space grows exponentially in the number of alternating features. Anything that could (accurately) set the values of some of the alternating features in advance would reduce further the size of the lexical search space, thus helping by reducing the space of lexical hypotheses to be tested for consistency. The rest of this paper will describe how contrast information can be used to set underlying values for some alternating features.

## 3 Faithful Contrastive Features

A *contrast pair* is a pair of words formed from two morphemes in the same morphological environment. A simple English example is *bedz* ~ *beds*, which are formed by the roots for the words “bed” and “bet”, each combined with the plural suffix. The morphemes being contrasted in this example are the two roots. The surface forms for the words are non-identical, so something must be different about the inputs for the words. Because the morphological environment (the suffix) is the same for both words, the portion of the input corresponding to the suffix is the same for both words. Therefore, the difference in the inputs for the two words must lie in a difference between the underlying forms for the two roots. In general, for any contrast pair, the differences between the surface forms of the words are consequences of differences between the underlying forms of the differing pair of morphemes in the pair.

One can determine that a contrast exists simply by observing the non-identity of the two surface forms of a contrast pair. But characterizing the nature of the contrast requires several correspondence relations. A surface-surface correspondence must be established between the two surface forms in order to characterize how the two forms differ.<sup>2</sup> Also, an input-output correspondence must exist relating each surface form to its respective input, which contains the underlying forms for the morphemes.

All three correspondence relations will prove necessary to reason about contrast between the underlying forms of different morphemes: an element of the underlying form for one of the contrasting morphemes relates via input-output correspondence to a surface element, then via surface-surface correspondence to a surface element of the other word of the pair, and finally via input-output correspondence to an element of the underlying form of the other contrasting morpheme.

Given such correspondence relations, we can identify conditions under which a distinction between the surface forms of a contrast pair directly indicates something about the underlying forms. A feature in a contrast pair is a *faithful contrastive feature* (FCF) if the contrasting morphemes of the pair have different values for that feature on the surface, and each surface realization of the feature is faithful to the underlying value of the feature for its respective morpheme. Note that we are not (yet) concerned with learning here; the concept of FCF is defined entirely independently of learning issues. This is illustrated in (2). The contrast pair consists of two words, the first combining root1 with

---

<sup>2</sup> The term “surface-surface correspondence” is used to distinguish the correspondence between surface forms used in this paper from the more familiar theory of output-output correspondence (Benua 1997).

suffix1, and the second combining root2 and suffix1. The two roots are the contrasting morphemes. The subscripts on the syllables indicate the input-output correspondence relations. The surface-surface correspondence (which is not indicated in (2)) relates the first syllables of the words (which relates r1 and r2 to each other), and the second syllables of the words (relating the occurrences of the environment morpheme s1 to each other).

$$(2) \quad r1+s1: /pa_1 + ka_2/ \rightarrow p\acute{a}_1ka_2 \qquad r2+s1: /pa_{:1} + ka_2/ \rightarrow p\acute{a}_{:1}ka_2$$

In this contrast pair, the length feature on the syllable of r1 and r2 is an FCF. The surface realization of r1 is short, while the surface realization of r2 is long, so the two contrast on that feature. Further, r1 is underlyingly short, and r2 is underlyingly long, so the surface realization of the feature in each morpheme is faithful to its respective underlying value.

Faithful contrastive features are appealing because the surface values directly reflect the corresponding underlying values. If a learner could identify FCFs in contrast pairs, then they could use them to set underlying values for the features. But that raises the question of how to identify FCFs. Is there even a guarantee that FCFs will exist?

It has been proven that, for systems meeting certain conditions, contrast pairs are guaranteed to have FCFs (Tesar 2004b). We state here some conditions that are jointly sufficient to guarantee the existence of FCFs. These conditions are too restrictive for phonologies in general, but represent the current extent of our understanding. First, correspondences, both input-output and surface-surface, are strictly one-to-one and onto (no insertion or deletion). Second, faithfulness constraints only preserve feature identity under correspondence (all faithfulness constraints are IDENT constraints (McCarthy and Prince 1995)). Third, features are binary.

Any Optimality Theory system meeting those conditions has the Faithful Contrastive Feature Property. For two morphemes in a given morphological environment, if an appropriate surface-surface correspondence can be established between the morphemes and those morphemes surface differently in that environment, then at least one of the differing features of the two morphemes is a faithful contrastive feature. This guarantees that FCFs exist. Next we will show that it is possible to identify and exploit some FCFs.

## 4 Contrast Analysis

One way to know that a given feature is an FCF is if it is the only feature in the contrast pair that is possibly an FCF. If a given feature is the only possible FCF, and we are guaranteed that there must be at least one, then it must be an FCF. This can be done with respect to surface data by observing what features differ between the surface realizations of the contrasting morphemes in the contrast pair. Recall the pair of forms in (2), which came from the data in Table 1, and consider the surface forms alone, as the learner would. The surface forms are shown in (3), along with the underlying forms for the morphemes as set in the initial lexicon, (1).

$$(3) \quad \begin{array}{lll} r1+s1: p\acute{a}ka & r2+s1: p\acute{a}:ka \\ r1: [?, -] & r2: [?, ?] & s1: [-, -] \end{array} \quad [+/-stress, +/-long]$$

The surface realizations of the two roots differ in only one feature: the length of the vowel. To be an FCF, a feature must differ in its surface realizations in the contrasting morphemes. Therefore, vowel length of the corresponding syllables must be an FCF for this contrast pair. The learner can correctly infer from this that r1 is underlyingly –long, and r2 is underlyingly +long.

The previous example is something like a featural minimal pair: the surface realizations of the contrasting morphemes differ in only one feature, so the underlying forms must contrast in that feature. However, there are cases of contrast pairs in which the contrasting morphemes differ on more than one feature, but a unique FCF can still be identified. That is the case when all but one of the features on which the morphemes differ on the surface cannot be FCFs for other reasons. The “other reasons” relate to the other half of the definition of FCF: the surface feature values must be faithful to their

underlying values. If all but one of the differing features are known to be unfaithful for one of the morphemes, then the one possibly faithful feature must be an FCF (and therefore must be faithful).

This can be illustrated by continuing the previous example. The contrast pair in (2) causes the learner to set  $r_2$  to be +long underlyingly.  $r_2$  is still unset underlyingly for stress. The learner now considers a different contrast pair:  $r_2+s_3$  and  $r_4+s_3$ . The contrasting morphemes are  $r_2$  (from the previous pair) and  $r_4$ , and the environment is suffix  $s_3$ .

- (4)       $r_2+s_3$ : *paká:*                       $r_4+s_3$ : *pá:ka*  
              $r_2$  [?,+]                       $r_4$  [+,+]                       $s_3$ [?,?]                      [+/-stress, +/-long]

In this pair,  $r_2$  and  $r_4$  differ on the surface in both stress and length. However, we know from the previous contrast pair that  $r_2$  is underlyingly long (it surfaced as long in the environment of suffix  $s_1$ , and contrasted in length with root  $r_1$  in that environment). Therefore, the length feature of  $r_2$  cannot be an FCF, because the surface realization of length, –long, is not faithful to the underlying value of length for  $r_2$ , +long. This leaves only stress in the roots as a possible FCF. From this, the learner is able to set the underlying value of stress for  $r_2$  to –stress.

This approach, setting the underlying values of alternating features when they are identifiably unique FCFs of contrast pairs, can be realized as a procedure, here labeled *Contrast Analysis*. Contrast Analysis (CA) accepts as input a paradigm of analyzed surface forms and an initial lexicon in which the underlying values of non-alternating features have been set to the values matching their surface realizations, and the alternating features are labeled as unset. CA then constructs contrast pairs and repeatedly examines them until no pairs result in any further changes to the lexicon of underlying forms. For each contrast pair examination, if a unique FCF can be determined for the pair, then the underlying values of the determined FCF are set to match their surface values in the pair.

Contrast Analysis can be combined with the other learning procedures described in section 2. The learner first performs phonotactic learning, and then morphological analysis. The learner then performs initial lexicon construction, setting the underlying values of non-alternating features. The learner then performs Contrast Analysis, setting underlying values for alternating features when possible. Finally, the learner uses the surgery learning algorithm to set any remaining underlying features while simultaneously determining the final ranking.

## 5 Simulation Results

The overall learning algorithm described at the end of the previous section was applied to all of the languages generated by the stress/length linguistic system. The system has 6 constraints, and yields a total of 24 distinct languages. The learning algorithm succeeds in learning all 24 languages, a meaningful fact about the algorithm. Of particular interest is how the algorithm succeeds, and what contributions toward success are made by the different components of the algorithm. An examination of the behavior of the algorithm on each of the languages reveals that each part of the algorithm has differing importance, depending upon properties of the language being learned. The parts of the algorithm combine to ensure that all of the languages are learned.

The 24 languages can be classified as shown in Table 2. The columns categorize the stress systems, while the rows categorize the patterns of vowel length. The middle column contains the 14 languages in which surface stress is sensitive to underlying specification of stress. The other columns contain languages with completely predictable stress, in pairs with left/right default. Languages with Initial/Final stress always have stress on the initial/final syllable, while QS languages are quantity sensitive, so that stress can be pulled away from the default position by a heavy syllable. The language depicted in Table 1 is a lexical stress language in which default stress overrides length.

**Table 2 Typology of the 24 languages**

	Initial Stress	QS Def Left	Lexical Stress	QS Def Right	Final Stress
<b>Length overrides default stress</b>		2 languages	10 languages	2 languages	
<b>Default stress overrides length</b>	2 languages		2 languages		2 languages
<b>No long vowels</b>	1 language		2 languages		1 language

From this initial classification of the languages, we can identify four larger groups that the learning algorithm relates to differently. The first group consists of languages with no alternating morphemes; this group contains the purely initial stress and purely final stress languages, for a total of 6 of the 24 languages. For these languages, phonotactic learning determines the correct constraint ranking. Initial lexicon construction determines the entire lexicon, as there are no alternating features. There is no necessary role for contrast analysis and surgery in the learning of these languages.

The second group contains the languages with quantity sensitive but predictable stress, 4 of the 24 languages. For these languages, phonotactic learning determines the complete ranking. Initial lexicon construction determines much of the lexicon, but some of the features alternate. Contrast Analysis is unable to set any of the alternating features for these languages. The reason is that the alternating morphemes alternate on stress, due to quantity sensitivity. Surface stress is purely predictable in these languages, so underlying stress is not contrastive, and therefore contrast analysis cannot determine the underlying values of the (alternating) stress features. In fact, the underlying values of these features do not matter. The surgery learning procedure assigns default values to the features, completing the lexicon, and thus the learning of the language.

The third group contains 12 of the 24 languages. All of them are lexical stress languages, the ones with no long vowels and the ones in which length overrides default stress. For these 12 languages, phonotactic learning cannot determine the complete ranking. In particular, it cannot determine the default stress assignment. Initial lexicon construction sets some of the underlying features, but many of the features alternate on the surface. Contrast analysis is able to set the values for all of the alternating features, completing the lexicon. For some of the languages, contrast analysis sets over half of the features of the lexicon. Surgery then easily applies BCD to determine the complete ranking from the determined lexicon.

The fourth and final group contains the languages with lexical stress in which default stress overrides length. This group contains 2 of the 24 languages. The two languages are mirror images of the same pattern, differing only in whether default stress is initial or final. Phonotactic learning cannot determine the complete ranking, once again being unable to determine default stress assignment. 9 of the 14 features in the lexicon do not alternate, and are set via initial lexicon construction. Contrast analysis sets 5 features, but cannot set the sixth. The surgery learning algorithm then sets the single unset feature and determines the rest of the ranking.

The language in Table 1 is an example of this last class of languages. A closer examination of learning with this language reveals the kinds of interactions that can elude the additive effects of phonotactic learning and contrast analysis alone. The result of initial lexicon construction for this language was shown in (1). The lexicon after contrast analysis is shown in (5).

(5)     r1[−,−]   r2[−,+]   r3[+,−]   r4[+,+]            s1[−,−]   s2[+,−]   s3[?,+]

The only feature that hasn't been set is the stress feature for suffix s3. That feature cannot be set by contrast analysis because the underlying form for s3 already contrasts with both s1 and s2 in length. In this language length only surfaces on stressed syllables, and suffixes are only ever stressed on the surface if they are stressed underlyingly. Thus, s3 must be +stress; otherwise, it's length would never surface, and it would be indistinguishable from s1, [−,−] (this is why there is no separate suffix in this language with underlying form [−,+]). In any environment in which s3 and another suffix contrast in stress, the suffixes will also contrast in length (s3 will surface as +long, while the other suffix will

surface as –long). Stress will thus never be a unique FCF, because length will also be an FCF. To put it another way, contrast analysis has no way of knowing a priori if the difference in stress is caused by the difference in length (e.g., via quantity sensitivity) or by underlyingly specified stress. The ranking is needed to determine that s3 must be +stress, and that the ability of s3's +long feature is dependent on its +stress feature.

## 6 Discussion

The most interesting property of the learning simulations is the apparent trade-off between phonotactic learning and contrast analysis. In the first two groups, phonotactic learning was able to learn the complete ranking, prior to any consideration of alternations, and contrast analysis did not make any contributions to learning the lexicon. In the third and fourth groups, where phonotactic learning could not determine the entire ranking, contrast analysis made very significant contributions to learning the lexicon. In the third group contrast analysis was able to set the entire lexicon, make the learning of the ranking rather simple. In the fourth group, contrast analysis set most of the lexicon, leaving a much smaller search space of still-unset features to be worked out in conjunction with the learning of the ranking.

The simulations described here are significantly limited, however, and the complete success of the learning algorithm must be viewed in light of those limitations. The test system only had two features per morpheme, with limited interaction between the two features. Having more features, interacting in more complex ways, would have the potential to make learning more difficult. Particularly problematic for the contrast analysis procedure presented in this paper are interactions that inherently involve several features at once. This is illustrated by the language discussed at the end of section 5. The underlying feature that could not be set by contrast analysis was one that only differed with other forms on the surface when another, related feature also differed on the surface.

Another artificial property of the simulations is that all possible surface forms (within the monosyllabic morpheme restriction) were included in the dataset presented to the learner. In actual language learning, the data provided to the learner will only be a subset of the possible grammatical forms of the language. The significance of this is that the underlying form of a given morpheme might be settable by contrast analysis given the right word to contrast with, but the key contrasting word, while a possible form of the grammar, is not actually available in the dataset to the learner. Thus, some things settable in principle by contrast analysis might not be settable in practice with a given data set.

Despite these limitations, there are reasons to believe that contrast analysis could be significantly beneficial to a learner even in more complex cases. If contrast analysis sets even some features, it reduces the lexical space, defined in terms of the number of unset features, that the learner must search in combination with learning the ranking. Furthermore, even if a number of features don't get set by contrast analysis due to the lack of convenient contrastive forms in the data, if even one dense lexical neighborhood exists in the actual lexicon of the language, it might contain enough near minimally contrasting morphemes that contrast analysis could set the underlying forms of the morphemes in that neighborhood. If the forms of that lexical neighborhood were sufficient to determine the constraint ranking, the learner could then use that knowledge of the ranking to complete the learning of the other underlying forms of the language.

More generally, the results reported here support the claim that contrast information of the sort provided by contrast pairs is useful in language learning. The contrast analysis procedure presented here is very simple; it doesn't even make reference to the nature of the constraints, let alone what might be known about the ranking via phonotactic learning. Yet it was able to make a significant contribution in these cases. More sophisticated procedures for exploiting contrast information might be even more effective, and could overcome some of the limitations just described for the simple contrast analysis procedure.

## Appendix

The test system used in the constraints uses purely monosyllabic roots and suffixes. Each morpheme has two underlying features: stress and length. The six constraints are given in (6).

(6) The constraints of the system

MAINLEFT	put main stress on the initial syllable.
MAINRIGHT	put main stress on the final syllable.
*V:	no long vowels.
WEIGHTTOSTRESS	long vowels should be stressed.
IDENT(stress)	output syllables should be identical to their input correspondents in stress.
IDENT(length)	output syllables should be identical to their input correspondents in length.

The markedness constraints on stress position are standard alignment constraints (McCarthy and Prince 1993). The markedness constraint against long vowels comes from Rosenthal (1994see also references therein). The markedness constraint linking stress to weight is an Optimality Theoretic version of the weight-to-stress principle (Prince 1990). The two faithfulness constraints are standard IDENT constraints of the correspondence theory of faithfulness (McCarthy and Prince 1995). For this system, stress is culminative: GEN requires that each word have exactly one (main) stress on the surface. Length is not culminative: GEN permits candidates with zero, one, or two long vowels on the surface.

The six constraints create a search space of  $6! = 720$  possible rankings. However, there are only 24 distinct possible languages, given the assumptions about the inputs and GEN. The languages vary significantly in size as measured in the number of words in the language. The smallest languages have only one word (no contrasts) while the largest languages have 16 words (full contrast in stress and length). The number of words in each language is a function of the number of surface-distinguishable morphemes in the language, which also determines the size of the lexicon for that language. The largest languages, with the largest lexica, have 8 morphemes (4 roots and 4 suffixes). Because each morpheme has two underlying features, the number of possible lexica in this worst case is  $2^{16} = 65,536$  lexica. The number of possible grammars, given the knowledge that there are 4 roots and 4 suffixes, is  $720 \times 65,536 = 47,185,920$  grammars (combinations of rankings and lexica).

The example language used throughout this paper, with the surface forms in Table 1, are generated by the grammar defined by the ranking in (7).

(7) WEIGHTTOSTRESS  $\gg$  IDENT(stress)  $\gg$  MAINLEFT  $\gg$  MAINRIGHT  $\gg$  IDENT(length)  $\gg$  \*V:

The mappings yielding the surface forms can be constructed from Table 3, which shows underlying forms for the roots and suffixes of the language.

**Table 3 The underlying and surface forms for the example language.**

r1 = /pa/	r2 = /pa:/	r3 = /pá/	r4 = / pá:/	
páka	pá:ka	páka	pá:ka	s1 = /-ka/
paká	paká	páka	pá:ka	s2 = /-ká/
paká:	paká:	páka	pá:ka	s3 = /-ká:/

The unshown suffix underlying form, /-ka:/, does not correspond to a separate morpheme because it neutralizes globally with s1 = /-ka/, for the reasons discussed at the end of section 5.

## References

Albright, Adam, and Hayes, Bruce. 2002. Modeling English past tense intuitions with minimal generalization. In *Proceedings of the Sixth Meeting of the ACL Special Interest Group in Computational Phonology*, ed. Michael Maxwell, 58-69. Association for Computational Linguistics.



- Benua, Laura. 1997. Transderivational Identity: Phonological Relations between Words. PhD. dissertation, University of Massachusetts, Amherst.
- Hale, Mark, and Reiss, Charles. 1997. Grammar Optimization: The simultaneous acquisition of constraint ranking and a lexicon. Ms., Concordia University, Montreal. ROA-231.
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in Phonological Acquisition*, eds. René Kager, Joe Pater and Wim Zonneveld, 158-203. Cambridge: Cambridge University Press.
- Kager, Rene. 1999. *Optimality Theory*. Cambridge: Cambridge University Press.
- McCarthy, John, and Prince, Alan. 1993. Generalized alignment. In *Yearbook of Morphology*, eds. Geert Booij and Jaap Van Marle, 79-154. Dordrecht: Kluwer.
- McCarthy, John, and Prince, Alan. 1995. Faithfulness and Reduplicative Identity. In *University of Massachusetts Occasional Papers 18: Papers in Optimality Theory*, eds. Jill Beckman, Laura Walsh Dickey and Suzanne Urbanczyk, 249-384. Amherst, MA: GLSA, University of Massachusetts.
- Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In *CLS26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, eds. Karen Deaton, Manuela Noske and Michael Ziolkowski, 355-398.
- Prince, Alan, and Tesar, Bruce. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, eds. René Kager, Joe Pater and Wim Zonneveld, 245-291. Cambridge: Cambridge University Press.
- Rosenthal, Sam. 1994. Vowel/glide alternation in a theory of constraint interaction. PhD. dissertation, University of Massachusetts, Amherst.
- Tesar, Bruce. 2004a. Using inconsistency detection to overcome structural ambiguity. *Linguistic Inquiry* 35:219-253.
- Tesar, Bruce. 2004b. Contrast analysis in phonological learning. Ms., Linguistics Dept., Rutgers University. ROA-695.
- Tesar, Bruce, and Smolensky, Paul. 2000. *Learnability in Optimality Theory*. Cambridge, MA: MIT Press.
- Tesar, Bruce, Alderete, John, Horwood, Graham, Merchant, Nazarré, Nishitani, Koichi, and Prince, Alan. 2003. Surgery in language learning. In *Proceedings of the Twenty-Second West Coast Conference on Formal Linguistics*, eds. G. Garding and M. Tsujimura, 477-490. Somerville, MA: Cascadia Press. ROA-619.