

UC Santa Cruz

Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince

Title

Learning from Paradigmatic Information

Permalink

<https://escholarship.org/uc/item/7jt3z39w>

Author

Tesar, Bruce

Publication Date

2006-06-20

Learning From Paradigmatic Information*

Bruce Tesar

Rutgers University, New Brunswick

1. Paradigmatic Information

Paradigmatic information in phonology consists of the surface realizations of different morphemes in combination. It involves knowledge of morpheme identity; knowing that the same morpheme surfaces differently in different morphological contexts implies knowing that the portions of the relevant surface forms are in fact realizations of the same morpheme.

In this paper, I will distinguish two kinds of paradigmatic information. The first is **morphemic alternation**: the surface realizations of the **same morpheme in different morphological environments**. A simple illustration of this is the surface realizations of the verbal prefix *un-*, with the surface realization [ʌŋ] in *unkind* and the surface realization [ʌn] in *unable*.

The second kind of paradigmatic information is morphemic contrast: the surface realizations of **different morphemes in the same morphological environment**. This is illustrated by the surface realizations of the prefixes *un-* and *en-* in the morphological environment of the verbal stem *able*, the first surfacing as [ʌŋ] in *unable*, and the second surfacing as [en] in *enable*.

Paradigmatic information is of interest here because of its role in phonological learning. It stands in addition to purely phonotactic learning, in which each word is treated as isolated and monolithic.¹ Phonotactic learning has been characterized as a stage in which the learner has no awareness of word-internal morphological structure, and no knowledge of shared morphemes across words (Hayes 2004, Prince and Tesar 2004). In other words, phonotactic learning does not make use of paradigmatic information. Phonotactic information, under this characterization, consists of the observed inventory of surface word forms.

The key ideas of this paper concern the role of paradigmatic information in phonological learning. Paradigmatic information is necessary for learning; there are aspects of phonological systems which are not revealed through phonotactic information alone. Both kinds of paradigmatic information, as defined above, are

* I wish to acknowledge the significant contributions to this research project made by John Alderete, Adrian Brasoveanu, Nazarré Merchant, and especially Alan Prince, who after a decade continues to be an outstanding colleague and collaborator. Helpful discussion was provided by the Rutgers Optimality Reading Group. Parts of this paper are 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 author and do not necessarily reflect the views of the National Science Foundation or of the National Institutes of Health. Any errors are the sole responsibility of the author.

¹ This statement of the distinction between paradigmatic and phonotactic information is somewhat artificial, given that words can also have different phonetic surface realizations depending on syntactic context, and that it is likely that learners employ the same kinds of strategies for segmenting both words and morphemes from utterances. A more general statement of phonotactic information might treat utterances as monolithic, rather than words. To keep the discussion manageable, I will stick to morphologically-organized paradigms of single-word utterances in this paper.

valuable for learning. Morphemic contrasts are useful for determining the underlying values for alternating features (features which surface differently in different morphological contexts). Morphemic alternations are useful for determining non-phonotactic ranking information.

2. System for Illustration

2.1 Forms and Constraints

The following linguistic system will be the basis for the examples in this paper. The system is based on Optimality Theory (Prince and Smolensky 2004), and presumes a morphological system in which each word consists of a root and a suffix. Further, each morpheme (root or suffix) consists of a single syllable. Each syllable has two features which can be specified in the input, length and stress. Vowel length has two values: long (+) and short (-). The stress feature has two values: main stress (+) or unstressed (-).

Some example surface word forms are given in (1). Each word consists of a root and a suffix; the word *pá:ka* is morphologically segmented *pá:+ka*.

- (1) *pá:ka páka pá:ka: paká: pa:ká paká*

The six constraints of the system are given in (2). The markedness constraints on stress position are standard alignment constraints (McCarthy and Prince 1993). The markedness constraint against long vowels comes from Rosenthal (Rosenthal 1994). 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.

- (2) The six constraints of the system.

MAINLEFT	put main stress on the initial syllable.
MAINRIGHT	put main stress on the final syllable.
*V:	no long vowels.
WSP	long vowels should be stressed.
FAITHSTRESS	input/output correspondents should be identical in stress.
FAITHLENGTH	input/output correspondents should be identical in length.

2.2 Two Example Languages

Two of the many possible languages of this system are given below. The first, which will be called Language A, is shown in (3), as a paradigm formed by four roots and three suffixes. Each surface word (there are 12) is the result of combining the indicated underlying forms for a root and a suffix to form a linguistic input (via concatenation), and using the ranking to determine the optimal surface form. The paradigm shown uses a complete morpheme inventory. The reason only three suffix underlying forms are given is that the underlying forms */-ka:/* and */-ka/* neutralize in all morphological environments; their behavior is indistinguishable on the surface. A constraint ranking giving rise to this language is shown in (4).

- (3) Language A: underlying and surface forms

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

- (4) WSP >> FAITHSTRESS >> MAINLEFT >> MAINRIGHT >> FAITHLENGTH >> *V:

Language B, another possible language of the same system, is shown in (5), and is the result of the constraint ranking in (6). Only three roots are shown, because root underlying forms /pa:/ and /pa/ neutralize in all morphological environments.

- (5) Language B: underlying and surface forms

r1=/pa/	r3=/pá/	r4=/pá:/	
<i>paká</i>	<i>páka</i>	<i>pá:ka</i>	s1=/-ka/
<i>paká:</i>	<i>páka</i>	<i>pá:ka</i>	s2=/-ka:/
<i>paká</i>	<i>paká</i>	<i>paká</i>	s3=/-ká/
<i>paká:</i>	<i>paká:</i>	<i>paká:</i>	s4=/-ká:/

- (6) WSP >> FAITHSTRESS >> MAINRIGHT >> MAINLEFT >> FAITHLENGTH >> *V:

2.3 Same Phonotactics, Different Languages

What is significant about languages A and B is that they have identical phonotactic inventories. The inventories of possible surface forms are shown in (7) and (8). Both languages have lexically variable stress, and long vowels are always stressed on the surface. Yet the languages are definitely different, to the point of having different numbers of contrasting roots and suffixes. Language A has stress defaulting to the left, while language B has stress defaulting to the right, as indicated by the different surface form each language assigns to the input /paka/.

- (7) Language A phonotactic inventory:
- páka paká pá:ka paká:*

- (8) Language B phonotactic inventory:
- páka paká pá:ka paká:*

This “phonotactic ambiguity” is not unique to this pair of languages. In fact, in this system, there are six distinct languages sharing this same phonotactic inventory. Paradigmatic information will be required to learn at least the elements of these languages that differ. Specifically, paradigmatic information will be necessary to determine (a) the relative ranking of MAINLEFT and MAINRIGHT; (b) the ranking of FAITHLENGTH relative to FAITHSTRESS and {MAINLEFT, MAINRIGHT}; and (c) the underlying forms of the morphemes.

3. Goals and Motivations

3.1 The Nature of The Proposal

This paper does **not** lay out an overall architecture for language learning, or anything like a complete theory of learning. It focuses much more narrowly on how paradigmatic information relates to the grammar, and on how paradigmatic

information can and must be used by an algorithmic learner. While the example discussed throughout the paper is presented in a pseudo step-by-step fashion (to illustrate ordered dependencies between different kinds of information), this should not be taken to imply that the operations depicted must be performed all at one time, or even only once. It is not difficult to imagine a learner applying the operations at different times to different paradigms of words as the appropriate words are acquired by the learner. The current work aims to develop principles governing how learners employ paradigmatic information, principles which might be exploited in more than one way in the context of larger learning proposals.

3.2 Information and Computation

Learning algorithms frequently break data into subsets, and perform specific processes over those subsets. The issue of what data subsets to focus on is a significant one in learning, and one that is subject to (at least) two often conflicting concerns. One concern is information: what data subsets are sufficient to provide the key information needed? The other concern is computational efficiency: what data subsets can be processed efficiently by learning algorithms?

Information content of data subsets matters because the data are crucially inter-related. In fact, the essence of paradigmatic information is the inter-relatedness of different words that share morphemes; properly interpreting one such word depends on what the others are like. If the learner is going to focus processing on data subsets, it will want to construct subsets that group together forms that are inter-related in revealing ways. However, computational effort typically increases with the size and complexity of data subsets. If the learner processes a data subset by trying many different possible underlying forms for each of the morphemes in the data subset, then the required computational effort can be expected to increase significantly for data subsets with larger numbers of morphemes (other things being equal).

The learner needs to construct data subsets that realize the best trade-off between these conflicting concerns. The entire collection of all forms heard by the learner is guaranteed to contain all of the relevant information available to the learner, but simultaneously evaluating all morphemes will be too computationally inefficient. A single word can be processed (relatively) efficiently, but contains too little information. Specifically, paradigmatic information cannot be obtained by processing single words in isolation.

3.3 Working Assumptions

Assumptions about faithfulness must be a core part of any research into the nature of underlying forms. Discussion in this paper is restricted to the most basic and fundamental kind of faithfulness constraint, IDENT(feature) constraints (McCarthy and Prince 1995). The faithfulness constraints in (2), FAITHSTRESS and FAITHLENGTH, are both IDENT constraints (the names were chosen to emphasize the fact that they are faithfulness constraints). Understanding how other types of faithfulness constraints interact with paradigmatic information will likely build on an understanding of IDENT constraints. Further, I will not consider any IDENT constraints with covert restrictions. A hypothetical constraint with a covert restriction would be IDENT-LOW(long), which would apply only to input-output correspondents

in which the input correspondent is specified as +low. The fact that the restriction on the constraint is conditioned on a property of the input correspondent (which is not overtly observable) makes the restriction covert.

A consequence of restricting faithfulness to such IDENT constraints is that every grammatical surface form maps to itself. This is a property that has been commonly assumed in work on phonotactic learning (Hayes 2004, Prince and Tesar 2004). It means that if *pá:ka* is a possible surface form of the language, then an identical input form will be mapped unchanged to that surface form by the phonology: /pá:ka/ → *pá:ka*. This is true whether or not the correct input for some actual occurrence of *pá:ka* is that same form. In other words, these working assumptions ban opaque mappings like chain shifts. The interested reader can consult recent work by McCarthy (2004) for some ideas on learning opaque mappings within a single ranking, and work by Bermudez-Otero (2003) for some ideas on learning opaque mappings realized via a sequence of (non-opaque) constraint rankings.

While the example used in the paper features only root + suffix morphological structures, the work generalizes to words with multiple suffixes, as well as prefixes, where the linguistic input for a word is formed by concatenating the underlying forms of the morphemes in the order dictated by the morphology. However, the current work assumes that each segment in a word is affiliated with exactly one morpheme. Future work will need to investigate the implications of analyses with subsegmental morphemes (like floating features) and with processes like coalescence.

Paradigmatic information requires knowledge of morphological segmentation. The work in this paper investigates the relationships between paradigmatic information and phonological learning by assuming that the learner already has correct knowledge of morphological segmentation. In reality, morpheme discovery and phonological learning almost certainly happen together. Figuring out how children simultaneously identify morphemes and learn phonologies requires building on an increased understanding of the relationships between morpheme identity, lexical underlying forms, and phonological mappings. The work in this paper contributes to the latter, with the goal of making progress towards the former.

4. Phonotactic Learning and Faithful Mappings

4.1 Phonotactic Ranking Information

While phonotactic information is insufficient to determine the entire grammar, it can determine parts of the grammar. Phonotactic learning here involves treating each word in isolation, as if it were monomorphemic. Proposals for phonotactic learning have used linguistic inputs that are identical to the observed outputs, and then searched for the most restrictive ranking that will map each of the observed forms to itself. Both Biased Constraint Demotion (Prince and Tesar 2004) and Low-Faithfulness Constraint Demotion (Hayes 2004) try to characterize restrictiveness in terms of a general bias towards having markedness constraints dominate faithfulness constraints.

The phonotactic inventory of Language A, repeated below in (9), determines some ranking relations (phonotactic learning cannot determine the underlying forms of morphemes for the simple reason that it has no knowledge of morpheme identity).

The ranking relations shown in (10) and (11) are necessary to map the observed forms to themselves. Note that both involve the domination of markedness constraints; these ranking relations are necessary to preserve marked structures observed in the surface forms. The ranking relations in (12) result from the markedness over faithfulness bias. Note that the constraint WSP is not violated in any surface forms; ranking it above all faithfulness constraints will help to ensure that the WSP is not violated by the outputs assigned to other inputs not currently under consideration.

(9) Language A phonotactic inventory: *páka paká pá:ka paká:*

(10) FAITHSTRESS \gg {MAINLEFT, MAINRIGHT}

(11) FAITHLENGTH \gg *V:

(12) WSP \gg {FAITHSTRESS, FAITHLENGTH}

The ranking relations in (10), (11), and (12) capture the phonotactics of the language. But they do not determine the entire ranking. The phonotactic ranking information is based solely on fully faithful mappings (the output is fully faithful to the input). Unfaithful mappings are needed to determine the rest of the ranking. Alternating morphemes are necessary to motivate unfaithful mappings; because each morpheme must have a single underlying form, a morpheme with multiple surface realizations must be involved in at least one unfaithful mapping.

4.2 Setting Non-Alternating Features

It is worth reviewing why it is challenging to learn underlying forms for alternating morphemes. It is not sufficient to test the different possible underlying forms for a single given morpheme, and see which one works for that morpheme in all observed environments. Leaving aside the issue of the potentially huge number of possible underlying forms for even a single morpheme, a more fundamental shortcoming lies in the fact that the behavior of one underlying form depends upon the underlying forms of the morphemes it combines with. A grammatical mapping assigns a linguistic output (a surface form for a word) to a linguistic input, where the input is formed by combining the underlying forms of all of the morphemes in the word. The learner cannot test the consequences of a hypothesized underlying form for one morpheme without making assumptions about the underlying forms for the other morphemes it combines with. The learner must reason simultaneously about the underlying forms for multiple morphemes.

The above reasoning applies to the full learning of underlying forms. However, it does not imply that nothing can be learned about underlying forms from considering just the surface realizations of a single morpheme. Given the working assumptions of section 3.3, if a feature of a morpheme does not alternate, it can be safely set underlyingly to match its (single) surface realization. The reasoning supporting this conclusion is as follows. If faithfulness to the underlying value of the feature plays a role in determining the surface value of the feature in some morphological context, then the underlying value **must** match the surface value (because the only direct effect faithfulness can have is to force the surface value to be identical to the underlying value). If, on the other hand, faithfulness to the underlying feature value **doesn't** play a role in determining the surface value, then

the underlying value **doesn't matter**, and it is perfectly safe (although not necessary) to set the underlying value to match the surface value.

Recall the surface forms of Language A, repeated here as (13). Consider first the surface realizations of root r1. r1 alternates in stress: it is stressed in the environment of suffix s1, but unstressed in the environments s2 and s3. On the other hand, it does not alternate in vowel length: r1 surfaces –long in every environment. Thus, the learner may safely set the underlying length feature of r1 to –long. The stress feature for r1 alternates on the surface, and thus cannot be set initially; further processing is required. Performing this analysis for every morpheme results in an initial lexicon in which non-alternating features are set to their (single) surface realization, while alternating features are marked as not yet set. This is shown for Language A in (14).

(13) The surface forms of Language A.

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

(14) The initial lexicon for Language A /+ – stress, + – long/, ? = unset

r1 /?,-/	r2 /?,?/	r3 /+,-/	r4 /+,+/
s1 /-,-/	s2 /?,-/	s3 /?,?/	

In the initial lexicon, the symbol ‘?’ denotes a feature that has **not yet** been set by the learner. This symbol only has status with respect to the learner, not with any adult phonology itself. In particular, it does not indicate underspecification, or any separate value of a feature.² An unset feature will ultimately be set by the learner to one of the possible values of the feature. For stress and length in the current system, the possible values are ‘+’ and ‘-’.

Setting non-alternating features is useful, as it limits the effective lexical search space for subsequent learning to only the features that alternate. But it will not give the learner the unfaithful mappings needed to determine non-phonotactic ranking relations. To achieve unfaithful mappings, the learner needs to set the underlying values of some alternating features. If a feature alternates, it must be unfaithfully mapped in some environment.

5. Inconsistency Detection and Contrast

5.1 Morphemic Alternations Are Inadequate for Setting Underlying Forms

One learning technique that can be applied to the evaluation of hypothesized underlying forms for morphemes is inconsistency detection (Tesar 2004a). Applied to the evaluation of underlying forms, inconsistency detection involves constructing a hypothesis of underlying forms for a set of morphemes, and then searching for a

² Note that, in a linguistic system in which underspecification of some feature can play a role in the adult phonology, one of the possible values of the feature is a value labeled ‘underspecified’; in applying this approach to such a system, one might imagine the learner at various points changing a feature listed as unset ‘?’ by setting it to the value ‘underspecified’.

constraint ranking that assigns the correct surface form to each combination of the morphemes. The constraint ranking search can be efficiently performed using Multirecursive Constraint Demotion. If no available ranking works, then the combination of underlying forms constituting the hypothesis must be incorrect (Kager 1999, Tesar et al. 2003). The available rankings can be constrained by other, already learned, ranking information.

Underlying forms can sometimes be set with inconsistency detection using only phonotactic ranking information (Pater 2000, Tesar and Prince to appear). But that won't work when the relevant ranking information isn't phonotactically apparent. Such is the case with the stress feature for *r1*. Paradigmatic information must be utilized in order to set that feature.

But what kind of paradigmatic information? In the case of the stress feature of *r1*, given what the learner knows thus far, applying inconsistency detection to morphemic alternation information for *r1* will not be sufficient. The forms containing *r1* are shown in (15). Applying inconsistency detection requires considering different possible underlying forms for both *r1* and the suffixes. However, we can show that inconsistency detection will fail to set the underlying stress value for *r1* by showing that there exist consistent combinations of underlying forms for *r1* and the suffixes with *r1* both underlyingly +long and –long. These are shown in (16) and (17).

- (15) *r1s1: páka* *r1s2: paká* *r1s3: paká:*
- (16) *r1 = /pa/* works if FAITHSTRESS \gg MAINLEFT \gg MAINRIGHT
r1s1: /paka/ \rightarrow páka
r1s2: /paká/ \rightarrow paká
r1s3: /paká:/ \rightarrow paká:
- (17) *r1 = /pá/* works if FAITHSTRESS \gg MAINRIGHT \gg MAINLEFT
r1s1: /páka/ \rightarrow páka
r1s2: /páká/ \rightarrow paká
r1s3: /páká:/ \rightarrow paká:

In (16), *r1* is underlyingly –stress, and default stress in the ranking is initial. Stress is thus initial by default in *r1s1*, and underlying stress on the suffix is faithfully preserved in *r1s2* and *r1s3*. In (17), *r1* is underlyingly +stress, and default stress in the ranking is final. The underlying stress on *r1* is faithfully preserved in *r1s1*, while the default final stress decides the conflict between the underlying stresses in *r1s2* and *r1s3*. Based only on the alternation information for *r1*, both values for the stress feature of *r1* have a consistent solution. The two solutions involve different rankings, but the differences in the rankings are precisely the kind of non-phonotactic ranking information the learner does not yet have.

5.2 Morphemic Contrasts Can Set Underlying Forms

The other kind of paradigmatic information, morphemic contrast, proves to be more useful for setting underlying forms. A **contrast pair** is a pair of surface-distinct words differing in only one morpheme (Alderete et al. 2005, Tesar 2004b). A contrast pair for *r1* can be formed by combining the words *r1s2* and *r3s2*. This pair could be described as *r1* and *r3* appearing in the morphological environment *s2*. The output forms are shown in (18).

(18) r1s2: *paká* r3s2: *páka*

By definition, the two words of a contrast pair have non-identical surface forms. The difference between the surface forms cannot be solely a consequence of the underlying forms of the environment morphemes, because those morphemes must have the same underlying form in both words. In (18), environment morpheme s2 must have a single underlying form. Therefore, the surface contrast between the two words must be due to a difference in the underlying forms of the morphemes that differ between the two words of the pair; in the present example, r1 and r3.

Inconsistency detection can be applied to a contrast pair by separately evaluating the possible lexical hypotheses for the pair (Merchant and Tesar to appear). Each lexical hypothesis is a distinct set of values for the unset features of the morphemes in the pair. For the pair r1s2 and r3s2, the current underlying feature assignments are repeated in (19). There are two unset features, the stress features for r1 and s2. Thus, there are four possible lexical hypotheses (four combinations of the two values for each of the two features). Inconsistency detection will evaluate which lexical hypotheses are consistent with some constraint ranking. Any features that have the same value in all of the lexical hypotheses can be safely set to that value, as the correct lexical hypothesis must be among the consistent ones.

(19) r1 /?stress, -long/ r3 /+stress, -long/ s2 /?stress, -long/

(20) Inconsistency detection applied to the four lexical hypotheses of r1s2 and r3s2.

r1	s2	Mappings	Consistent?
-stress	-stress	r1s2: /pa+ka/ → <i>paká</i> r3s2: /pá+ka/ → <i>páka</i>	Yes
-stress	+stress	r1s2: /pa+ká/ → <i>paká</i> r3s2: /pá+ká/ → <i>páka</i>	Yes
+stress	-stress	r1s2: /pá+ka/ → <i>paká</i> r3s2: /pá+ka/ → <i>páka</i>	No
+stress	+stress	r1s2: /pá+ká/ → <i>paká</i> r3s2: /pá+ká/ → <i>páka</i>	No

Of the four lexical hypotheses, only two are consistent. The other two sets of mappings cannot be achieved by any available ranking of the constraints. The benefit for the learner comes from the observation that the stress feature for r1 has the same value in every consistent lexical hypothesis: -stress. Thus, based on this contrast pair, r1 may be set to -stress. Note that applying inconsistency detection to the contrast pair r1s2 with r3s2 accomplishes what couldn't be accomplished with any single form in isolation, and what couldn't be accomplished with the morphemic alternation information for r1.

The stress feature for s2 cannot be set on the basis of this contrast pair: it has conflicting values among the consistent lexical hypotheses. Other information will be needed to set that feature. This is not surprising: s2 is the environment morpheme, and thus its underlying form cannot alone account for the surface contrast in the pair. The key information provided by a contrast pair concerns the morphemes being contrasted.

6. The Roles of Morphemic Contrast and Morphemic Alternation

6.1 Learning Non-Phonotactic Ranking Information

The contrast pair of r1s2 with r3s2 succeeded in setting the underlying value for the stress feature of r1. This is the underlying value of an alternating feature. It was faithfully preserved in the form r1s2 that was employed in the contrast pair. But now the learner can look for environments in which the feature is does not surface faithfully, environments where r1 surfaces as stressed. Such environments can provide evidence concerning unfaithful mappings of the feature. Such environments can be identified by examining the different surface realizations of a morpheme (here, r1) to see where the newly set alternating feature is not faithfully preserved. In other words, the learner identifies such unfaithful mapping environments by consulting morphemic alternation information.

The form r1s1 surfaces with r1 stressed, and thus is not faithful to the underlying –stress feature value of r1. The learner now has complete underlying forms set for both r1 and s1: r1 is /–stress, –length/, and s1 is /–stress, –length/. The learner now knows both the linguistic input and the output for r1s1: /paka/ → *páka*. Thus, the learner can now obtain further ranking information from this unfaithful mapping, using Multirecursive Constraint Demotion. The result of this is shown in (21). The winner, with initial stress, differs from the loser, with final stress. Both violate FAITHSTRESS to the same extent, but conflict on the stress alignment constraints. Expressed as an elementary ranking condition (Prince 2003), this pair shows that MAINLEFT, which prefers the winner, must dominate MAINRIGHT, which prefers the loser.

(21) A winner-loser pair for r1s1, showing that MAINLEFT ≫ MAINRIGHT.

r1 /pa/ s1 /-ka/	WSP	MAINLEFT	MAINRIGHT	*V:	FAITHSTRESS	FAITHLENGTH
☞ <i>páka</i>			*		*	
<i>paká</i>		*			*	
ERC		W	L			

The learner has now obtained crucial non-phonotactic ranking information: MAINLEFT ≫ MAINRIGHT. The learner was able to obtain this information by first setting the underlying value of an alternating feature, using morphemic contrast, and then examining an unfaithful mapping of the feature, using morphemic alternation information.

6.2 Setting Features and Learning Rankings

The learner, now armed with additional ranking information, can use that additional information to set further underlying feature values. The additional ranking information further restricts the set of available rankings. This makes it easier to detect inconsistency when evaluating lexical hypotheses, because there are fewer possible rankings available to possibly support a given lexical hypothesis.

In the running example, suppose the learner next considers the surface form r2s3, which surfaces as *paká*. The learner's current lexical representations for r2 and s3 are given in (22). Together, these morphemes have two unset features, yielding

four lexical hypotheses. These four hypotheses can be evaluated using inconsistency detection, with the results as shown in (23).

(22) r2 /?stress, +long/ s3 /?stress, +long/

(23) Inconsistency detection applied to r2s3.

r2	s3	Mappings	Consistent?
–stress	–stress	r2s3: /pa:+ka:/ → <i>paká:</i>	No
–stress	+stress	r2s3: /pa:+ká:/ → <i>paká:</i>	Yes
+stress	–stress	r2s3: /pá:+ka:/ → <i>paká:</i>	No
+stress	+stress	r2s3: /pá:+ká:/ → <i>paká:</i>	No

The learner is able to set the stress features of both r2 and s3, based on this one form. The learner is able to do so because of the non-phonotactic ranking information previously obtained; each of the three inconsistent lexical hypotheses runs afoul of the ranking relation MAINLEFT ≫ MAINRIGHT. Two important points should be noted here. First, inconsistency detection applied to r2s3 could not have set the features without the non-phonotactic ranking information. Second, no contrast pair contrasting s3 with another suffix could set the stress feature of s3 without the non-phonotactic ranking information. There is an ordered dependence between these different non-phonotactic aspects of the grammar.

The learner was able to set the stress feature for s3 using r2s3 because the stress feature is crucially faithfully realized in the form r2s3: the suffix surfaces as stressed. The learner can now use morphemic alternation information to find a form in which stress for s3 is not faithfully mapped. The word r3s3 is such a form, surfacing as *páka* with s3 unstressed. Because the learner now has fully set underlying forms for both r3 and s3, it can learn from the unfaithful mapping /páka:/ → *páka*. This form pits faithfulness to the length of the suffix (FAITHLENGTH), which can only be satisfied if the suffix surfaces as stressed (due to the high-ranking WSP), against the default stress alignment to the left (MAINLEFT). The surface form indicates that default stress placement gets priority, revealing additional non-phonotactic ranking information: MAINLEFT ≫ FAITHLENGTH. The winner-loser pair shown in (24) requires that either MAINLEFT or *V: dominate FAITHLENGTH. Phonotactic ranking information already requires that FAITHLENGTH dominate *V:, therefore MAINLEFT must dominate FAITHLENGTH. With this additional ranking information, the learner now has all the ranking information necessary to determine the language; it has the complete ranking.

(24) A winner-loser pair for r3s3.

r3 /pá/ s3 /-ká:/	WSP	MAINLEFT	MAINRIGHT	*V:	FAITHSTRESS	FAITHLENGTH
☞ <i>páka</i>			*		*	*
<i>paká:</i>		*		*	*	
ERC		W	L	W		L

An overview of the key steps in the learning of non-phonotactic information in the example is shown in (25).

(25) Overview of learning steps in the example.

Forms	Mapping	Info Learned	Info Type
r1s2 with r3s2	faithful	r1 is -stress	underlying feature
r1s1	unfaithful	MAINLEFT \gg MAINRIGHT	ranking
r2s3	faithful	s3 is +stress	underlying feature
r3s3	unfaithful	MAINLEFT \gg FAITHLENGTH	ranking

In (25), the second column indicates whether the information being learned is based on a faithful or an unfaithful mapping of a crucial feature. In the first row, learning is based on the pair of forms r1s2 and r3s2. The crucial feature is the stress feature of r1, which is faithfully realized in the surface form r1s2. Note that it is the underlying value of the stress feature for r1 that is learned. In the second row, learning is based on the form r1s1. The crucial feature is again the stress feature of r1, which is not faithfully realized in the surface form r1s1. Note that it is a ranking relation between markedness constraints on the surface realization of stress that is learned. Observe more generally in (25) that underlying feature values are set based on faithful mappings of those features, while ranking information is set based on unfaithful mappings. This is not an accident, but follows from the nature of faithful and unfaithful mappings.

6.3 Unfaithful Mappings and Morphemic Alternation

If a feature is unfaithfully mapped in a given environment, then the surface feature value is not a consequence of faithfulness to the underlying feature value (the surface feature value would be the same if the underlying feature value were changed to match the surface feature value). Such a surface form does not provide useful information about the relevant underlying feature value, because multiple underlying values are consistent with the surface form. However, a surface form which unfaithfully maps a feature whose underlying value is already known is a potential source of ranking information. This is because the surface form may attest to domination of the relevant faithfulness constraint by other constraints.

This observation has implications for the role of morphemic alternation information. The potential inability of morphemic alternations to determine the underlying value of an alternating feature was illustrated in section 5.1 above. More generally, a morpheme feature alternation could be due to:

- faithfulness to the + value of the feature in one environment;
- faithfulness to the – value of the feature in another environment;
- faithfulness in no environments (fully predictable everywhere).

Because of this, morphemic alternation information isn't that useful for setting underlying feature values.

On the other hand, if the underlying value of a feature is already known, then morphemic alternation information is great for identifying unfaithful mapping environments. Examining all of the surface realizations of the relevant morpheme is the logical way to find those environments in which the feature is unfaithfully realized. Thus, morphemic alternation information is useful in the process of learning non-phonotactic ranking information, as it can identify the unfaithful mapping environments that are crucial for determining that ranking information.

6.4 Faithful Mappings and Morphemic Contrast

It was just argued that a surface form unfaithfully realizing an alternating feature is the wrong place to look for information indicating the underlying value of the feature. It follows that the right place will be a surface form that faithfully realizes the alternating feature. Note that it does not follow that just **any** surface form faithfully realizing the feature will conclusively indicate the underlying value; a feature can be faithfully realized on the surface, but for reasons having nothing to do with the underlying value of the feature. The useful surface forms will be those that faithfully realize the feature as a consequence (in part) of faithfulness to the underlying value of the feature. To put it another way, the evidence that an underlying feature must have a particular value will involve surface forms that would surface differently if that feature had a different value underlyingly.

When two words surface differently, the words must differ in at least one feature underlyingly (otherwise, their linguistic inputs would be identical, and they would necessarily surface identically). Further, faithfulness to (at least) one differentiating feature must play a role in determining the surface form for at least one of the two words. It is not sufficient for the two words to have non-identical inputs; at least one of the differences between the inputs must play a causal role in the determination of the surface forms, and underlying features play a causal role only by being faithfully mapped by faithfulness constraints.

Morphemic contrast is the differing of the surface realizations of two morphemes in the same morphological environment. Thus, morphemic contrast information assures the faithful realization of a feature distinguishing the two underlying forms. Choosing two words that differ in only a single morpheme restricts, within the words, the possible location of the underlyingly differing features to the contrasting morphemes, as the environment morphemes must have the same underlying forms in both words. The contrast between the surface realizations of the two words must be a consequence of the preservation by faithfulness constraints of at least one feature that differs underlyingly between the two morphemes. Thus, morphemic contrast is useful for setting underlying feature values, because it ensures the presence of features that are faithfully preserved, while constraining their location.

7. Discussion

7.1 Interdependence in Learning

The two different kinds of paradigmatic information have distinct roles to play in learning. Morphemic contrasts implicate key faithful mappings, which are useful for setting underlying feature values. Morphemic alternations are great for identifying unfaithful mappings, which are useful for determining non-phonotactic ranking information.

Algorithmically, the use of faithful and unfaithful mappings feed each other, as shown in the example running throughout this paper. Once the alternating features of a morpheme have been set (using environments where those features are necessarily faithfully preserved), the underlying form for that morpheme can be used to learn non-phonotactic ranking information via a morphological environment in which some of the alternating features are unfaithfully mapped. In the example, this

happened when the contrast pair of r_1s_2 with r_3s_2 was used to set the stress feature for r_1 (shown in (20)), and then the underlying form for r_1 , combined with the underlying form for s_1 , was used to learn that $\text{MAINLEFT} \gg \text{MAINRIGHT}$ (shown in (21)). The non-phonotactic ranking information thus obtained can be crucial for the setting of further features. In the example, this happened when the ranking information $\text{MAINLEFT} \gg \text{MAINRIGHT}$ made it possible for inconsistency detection to be used to set the underlying value of the stress feature for s_3 (shown in (23)).

The algorithmic interleaving of the setting of underlying feature values and determination of ranking information further highlights the interdependence of the ranking and the lexicon in learning. The two must be learned together. The constraint ranking relates the behaviors of the morphemes to each other. The determination of the ranking information $\text{MAINLEFT} \gg \text{MAINRIGHT}$ did not involve the morpheme s_3 (it involved r_1 , r_3 , s_2 , and s_1), yet the ranking information was crucial for determining the underlying form for s_3 (via the form r_2s_3). Morpheme behaviors can interact independently of their joint appearance in individual words; the common thread connecting the behaviors of all of the morphemes is the constraint ranking.

7.2 Contrast Pairs: Virtues and Possible Limitations

This paper builds an argument that morphemic contrast information is best suited for revealing the underlying values of features that alternate. The particular method used to illustrate the use of such information involves applying inconsistency detection to contrast pairs. In this method, the contrast pair data structure is the key data subset that is used to gain access to non-phonotactic information. I here go on to suggest that contrast pairs strike the right balance between information content and computational efficiency: it is the right kind of unit for a learner to be processing over.

A contrast pair is not just an arbitrary set of more than one word in the language. It is a pair of words related in a particular way. By design, a contrast pair implicates a difference in the underlying forms of the two morphemes that differ between the words of the pair. Simultaneously, the activity of a faithfulness constraint sensitive to a difference between the underlying forms of the two morphemes is also implicated. The two words of a contrast pair are related so as to reveal desired non-phonotactic information about underlying feature values.

The structure of a contrast pair also makes it computationally appealing. Because the inconsistency detection approach illustrated in this paper involves searching many local lexica (combinations of values for unset features), computational benefits result from minimizing the number of unset features being simultaneously considered. In a contrast pair, all of the environment morphemes are shared between the two words of the pair. The number of morphemes contained in a contrast pair is only one more than the number of morphemes in either of the words of the pair by itself. Given that processing words in isolation is insufficient, one additional morpheme is the minimal increase in the number of morphemes being simultaneously processed. Other things being equal, minimizing the number of morphemes can be expected to minimize the number of unset features being considered in combination.

The contrast pair strikes a balance between information content and computational efficiency because the key elements of the structure serve both

concerns simultaneously. Restricting the number of morphemes involved minimizes the number of underlying features that can possibly be responsible for the surface contrast while simultaneously minimizing the number of unset underlying features defining the search space of possible local lexica.

It is not currently known whether contrast pairs are sufficient, in the general case, to provide access to key non-phonotactic underlying feature values. It is a distinct possibility that there are OT grammars meeting the basic conditions described above for which contrast pairs are insufficient. One way this might imaginably arise is if, in a contrast pair, the wrong value of an unset feature appears possible because the learner is still considering as possible a certain alternative ranking, a ranking that can be ruled out if an additional word is added to the pair for simultaneous consideration. It would have to be the case that the additional ranking information is implied not by the third word in isolation (otherwise, the current method could handle it), but by the interaction between the words of the pair and the third word.

It is not the case that no information is ever shared between different contrast pairs. It is the case that the only information shared between contrast pairs is “non-contingent” knowledge about underlying feature values. A contrast pair changes the knowledge state of the learner when it allows the learner to set the underlying value of a feature. That newly set feature can then influence the processing of other contrast pairs, when the morpheme containing the set feature is contained in the latter contrast pair, or when the newly set feature permits the determination of additional ranking information. The kind of contrast pair processing described in this paper does not share contingent information. One example of such contingent information would be the information that several underlying feature values must be related to each other in restricted ways but does not pin down the exact value of any of the features. Another example of such contingent information is the ranking relations required for each consistent combination of underlying features in a contrast pair. For the kind of contrast pair processing described here to be adequate, it must be possible for the learner to succeed without explicitly capitalizing on the kinds of contingent information just described.

7.3 Issues for the Application of Inconsistency Detection to Contrast Pairs

The use of unfaithful mappings to learn non-phonotactic ranking information, as described in this paper, crucially depends upon the learner setting all of the features in key morphemes, so that the underlying form for the morphemes is completely known. It is at this point, when the underlying forms for all of the morphemes of a word are fully determined, that the learner can learn more about the ranking, because the input and output are fully known. But, in complex cases, it may be less likely that many of the underlying forms for morphemes will be fully set at the right point. The key danger here is the possibility of fully predictable features for morphemes, features whose underlying value never plays an important role in determining the surface realization of the morpheme in any environment. Such features, when they alternate, will never be set by inconsistency detection, because all values of the feature are consistent with some ranking (indeed, all values are equally consistent with the correct ranking).

The example used in this paper had full realization of the possible forms of the language: every phonologically possible word was actually attested in the data. Actual human languages do not fully utilize the set of possible forms permitted by their phonologies. This raises the possibility that key elements of the underlying form for a morpheme will be unsettable by contrast pair processing, not due to limitations of the method in principle, but because the crucially contrasting possible form isn't used by the language. A more subtle limitation could come if the relevant morphemes exist in the language, but a particular key combination of morphemes, while a valid form of the language, is infrequent enough that the learner won't reliably have an attested surface form to refer to. The lack of certain contrasting forms will not necessarily deny success to inconsistency detection with contrast pairs. If enough key contrasts exist in the data to determine the constraint ranking for the language, and that constraint ranking is sufficient to determine all important underlying feature values, then contrast pair processing will still succeed.

This paper does not discuss in any depth the selection of contrast pairs for processing. In a language with a large number of morphemes, there will be a large range of contrast pairs that could possibly be formed, and some will be much more informative than others. An intuitive heuristic is to look for pairs in which the contrasting morphemes are quite similar in their surface realizations, except for certain key features which have not been set for at least one of the morphemes. Such a heuristic, like most of the discussion in this paper, presupposes a surface-to-surface correspondence between the morphemes being contrasted. In the example used in this paper, all morphemes consist of a single syllable, so when two morphemes contrast, it is obvious that the single vowels of the two morphemes correspond to each other. With more complex systems, determining whether the surface realizations of two morphemes in a given environment "differ with respect to feature" will require that a surface-to-surface correspondence be established between the two surface realizations, such that corresponding segments surfacing with different values for a feature can be identified. Justifying such surface-to-surface correspondences may be non-trivial.

8. Conclusions

This paper argues that morphemic contrasts are the key to learning underlying values for alternating features in the absence of key ranking information. In particular, contrast pairs can be processed using inconsistency detection to set underlying feature values because they focus on necessarily faithful mappings of features. Morphemic alternations allow the identification of unfaithful mappings of features that have already been set, and those unfaithful mappings provide the basis for learning non-phonotactic ranking information. Non-phonotactic learning will alternate between the setting of underlying feature values and the determination of ranking information because of the dependencies between them.

References

- Alderete, John, Brasoveanu, Adrian, Merchant, Nazarré, Prince, Alan, and Tesar, Bruce. 2005. Contrast analysis aids the learning of phonological underlying forms. In *Proceedings of the Twenty-Fourth West Coast Conference on Formal*

- Linguistics*, ed. by John Alderete, Chung-hye Han, and Alexei Kochetov, 34-42. Cascadilla Press.
- Bermudez-Otero, Ricardo. 2003. The acquisition of phonological opacity. In *Variation within Optimality Theory: Proceedings of the Stockholm workshop on 'Variation within Optimality Theory'*, ed. by Jennifer Spenader, Anders Eriksson, and Osten Dahl, 25-36. Stockholm: Dept. of Linguistics, Stockholm University. ROA-593 (expanded version).
- Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory: The early stages. In *Constraints in Phonological Acquisition*, ed. by 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*, ed. by 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*, ed. by Jill Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk, 249-384. Amherst, MA: GLSA, University of Massachusetts.
- McCarthy, John J. 2004. Taking a free ride in morphophonemic learning. Ms., University of Massachusetts Amherst. ROA-683.
- Merchant, Nazarré, and Tesar, Bruce. to appear. Learning underlying forms by searching restricted lexical subspaces. *Proceedings of the Forty-First Conference of the Chicago Linguistics Society*.
- Pater, Joe. 2000. Unpublished course handout from Ling 751, University of Massachusetts, Amherst.
- Prince, Alan. 1990. Quantitative consequences of rhythmic organization. In *CLS26-II: Papers from the Parasession on the Syllable in Phonetics and Phonology*, ed. by Karen Deaton, Manuela Noske, and Michael Ziolkowski, 355-398.
- Prince, Alan. 2003. Arguing Optimality. In *UMOP 26: Papers in Optimality Theory II*, ed. by Angela Carpenter, Andries Coetzee, and Paul de Lacy, 269-304. GLSA. ROA-562.
- Prince, Alan, and Smolensky, Paul. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell.
- Prince, Alan, and Tesar, Bruce. 2004. Learning phonotactic distributions. In *Constraints in Phonological Acquisition*, ed. by 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, 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*, ed. by G. Garding and M. Tsujimura, 477-490. Somerville, MA: Cascadilla Press. ROA-619.
- Tesar, Bruce. 2004a. Using inconsistency detection to overcome structural ambiguity [Spring]. *Linguistic Inquiry* 35:219-253.
- Tesar, Bruce. 2004b. Contrast analysis in phonological learning. Ms., Linguistics Dept., Rutgers University. ROA-695.

Tesar, Bruce, and Prince, Alan. to appear. Using phonotactics to learn phonological alternations. In *Proceedings of the Thirty-Ninth Conference of the Chicago Linguistics Society, vol. II: The Panels*, ed. by Johnathon E. Cihlar, Amy L. Franklin, David W. Kaiser, and Irene Kimbara. ROA-620.

Department of Linguistics
18 Seminary Place
Rutgers University
New Brunswick, NJ 08901-1184
tesar@rutgers.edu
<http://ruccs.rutgers.edu/~tesar>