Using learnability as a filter on factorial typology: a new approach to Anderson and Browne's generalization*

John Alderete, alderete@sfu.ca

Simon Fraser University, March 2008

Abstract. This article investigates the learnability filter (LF) hypothesis, according to which the set of logically possible grammars predicted by linguistic theory is reduced to a proper subset of learnable grammars by external principles of language learning. Anti-faithfulness constraints (Alderete, 2001a) provide a linguistic theory that predicts the existence of circular chain shifts of two segment types, including purely phonological exchanges that are unattested cross-linguistically (Anderson and Browne, 1973). Overt data representing such systems are fed into a standard OT learning model in which learners have IO-antifaithfulness constraints at their disposal. Despite this evidence and these constraints, learners always select grammars in which segmental exchanges are restricted to morphologically defined environments, consistent with typological findings. These results are shown to have implications for the nature of constraints in Optimality Theory, the correct analysis of morpho-phonological alternations, and a host of representational assumptions in phonology and morphology.

1. Introduction

Restrictiveness is often seen as a consequence of the assumptions that constitute a linguistic theory. On this view, typological gaps result from theory-internal assumptions that simply fail to generate a pattern. While it is not always clear which assumptions are theory-internal and which are external, this approach is so pervasive in generative linguistics that it is hard to find a domain of language that has not been explored in this way. An alternative view on typological gaps is that gaps are simply as-yet unattested patterns, and that the reason they are unattested is not a consequence of linguistic theory proper but principles outside of it. This approach allows for models with greater generative capacity, but predicts that certain logically possible systems are rare or unattested because of external factors, like diachronic forces shaping language or cognitive processes that underlie language but are not explicitly formalized in linguistic theory.

* Thanks to Alexei Kochetov, Elliott Moreton, Johan Rooryck, Bruce Tesar, Anne-Michelle Tessier, and two anonymous Lingua reviewers for their comments and questions, as well as the audience members of a linguistics colloquium at the National Taiwan Normal University. This article is supported in part by the Social Sciences and Humanities Research Council of Canada (SSHRC 410-2005-1175). Any errors that remain, despite this help, are the responsibility of the author.

This article does not attempt to argue for one of these approaches over another in any particular domain. Rather, it draws this distinction as a way of making the point that these rather different approaches can in fact combine in certain domains to explain observations that cannot otherwise be explained with principles of linguistic theory alone. In particular, the explanation for certain morpho-phonological systems is argued to depend crucially on the structure of grammars in Optimality Theory (Prince and Smolensky, 1993/2004), as well as an external model of learning these systems based on (Tesar and Smolensky, 2000). The argument therefore extends a proposal articulated in (Boersma, 2003), namely that principles of learnability can reduce typologies predicted by linguistic theory. I call this hypothesis the learnability filters hypothesis.

This approach is applied to a now classic problem, namely the finding that exchange rules are always morphological in nature ((Anderson and Browne, 1973) , (Anderson, 1992), (Moreton, 2004), cf. (Chomsky and Halle, 1968)). The linguistic theory of exchanges requires antifaithfulness constraints (Alderete, 2001a) and an extension of antifaithfulness to input-output correspondence. This new theory of exchanges over-generates in that it allows for certain classes of phonological exchanges that are not known to exist. However, by feeding these unattested patterns in a particular learning model, it is shown that only morphological exchanges are allowed because only morphological analyses of the exchange data can be learned.

The particular learning model employed below is based on principles proposed for learning phonotactic distributions ((Tesar and Smolensky, 2000), (Hayes, 2004), and (Prince and Tesar, 2004)) and a set of extensions of that theory designed for learning alternations ((Tesar et al., 2003), (Alderete et al., 2005)) and non-automatic phonological patterns (Pater, 2005). The argument for the combined approach, based both in linguistic theory and learnability, is that the filtering effect of learnability explains the morphological nature of exchange rules with principles that are motivated on totally independent grounds. The explanation depends on very general principles, like the distinction between markedness and faithfulness constraints, and the fact that learning algorithms like Biased Constraint Demotion (Prince and Tesar, 2004) are sensitive to these distinctions. This approach contrasts with alternatives that attempt to explain the morphological nature of exchanges with purely linguistic principles, like the stipulated non-existence of input-output antifaithfulness assumed in (Alderete, 2001a), as well as some other recent approaches to the problem.

The rest of this article is organized as follows. The next section provides an analysis of the problem of exchange rules by proposing a theory-internal description of phonological exchanges, as opposed to morphological exchanges. Sections 3 and 4 establish the specific algorithms and principles underlying learning that are needed to test the learnability filters (LF) hypothesis, with section 4 demonstrating for the first time how these principles can be applied to the problem of learning a grammar of morphological exchanges from overt data. Section 5 tests the LF hypothesis with the assumed learning model by showing that: (i) overt data consistent with either a morphological or phonological analysis must be learned as morphological, and (ii) even overt data that supports surface phonological generalization must be treated as morphological. Section 6 explores some additional consequences of the learning model, including finding indirect evidence for the LF hypothesis in the mechanisms involved in learning inputs, some conjecture about the type of evidence that might support input-output antifaithfulness, and discussion of other analytical biases in learning automatic alternations. Section 7 summarizes the main argument and then section 8 discusses a number of problems with purely linguistic approaches to exchange rules.

2. Analysis of the problem

To understand the problem posed by exchange rules, it is necessary to review the results of some prior work, beginning with the formal analysis of circular chain shifts in (Moreton, 2004). This work showed that in a version of Optimality Theory characterized by the assumptions below, certain collections of mappings are not computable. That is, if the premises (1a-c) below are true, then no ranking of constraints can characterize a grammar that produces a circular chain shift of any number of sound structures. This is because such a rotation, including simple two-way exchange rules, involve a set of unfaithful mappings, and the complete rotation entailed by these mappings cannot improve on markedness.

(1) The computable functions (CF) argument, summarizing (Moreton, 2004)

a. Homogeneity: inputs and outputs are made of the same representational elements

b. Inclusivity: the candidate set always includes the fully faithful candidate

c. Conservativity: all constraints are either markedness constraints or faithfulness constraints

d. Conclusion: no OT grammar can characterize a function consistent with a circular or infinite chain shift

A remarkable 'theoretical fact' (in the sense of (Smolensky, 2006)), this analysis is useful in that it clarifies a set of alternations that are in a sense outside the scope of simple markedness-faithfulness interactions. Circular chain shifts are in fact well-attested, but cross-linguistic investigation has concluded that they are always morphological in nature, meaning essentially that they are correlated with particular morphological constructions. In the Nilotic language Dholuo, for example, voiced and voiceless segments are exchanged, but only in plurals formed with /-e/ and construct state forms ((Gregersen, 1972), (Okoth-Okombo, 1982), cf. (Tucker and Creider, 1994)).

(2) Morphological voicing polarity in Dholuo

a. Voiceless obstruents to voiced obstruents

Singular	Plural (-e)	Construct State	
gɔt	gɔd-ɛ	gɔd	'hill'
koti	kod-e	kod	'coat'
kɔθ	kɔð-ɛ	kɔð	'rain'

b. Voiced obstruents to voiceless obstruents

kidi	kit-e	kit	'stone'
mɔgo	mɔk-ɛ	mɔk	'flour'
puoðo	puoθ-e	puoθ	'garden'

Similar alternations, crucially tied to morphological constructions, have been found in Arabic, Czech, Biblical Hebrew, Diegueño, Dinka, and Menomini. This generalization, stated originally in (Anderson and Browne, 1973) for exchange rules, and extended to circular chain shifts in (Moreton, 2004), raises the fascinating question, why are circular chain shifts always morphological in nature?

A prerequisite to answering this question, and indeed proposing any kind of analysis of it, is to understand more precisely what it means to say that exchange rules are always morphological, and further, what types of analyses would contradict it. Anderson and Browne's original discussion likens the known examples to ablaut in English verbal morphology and states that they are morpholexical, entailing essentially that the process is restricted to specific morphemes or particular morphological contexts. For Dholuo, because of the existence of loanword adaptations like *kod-e* 'coats', it would appear that the exchange is morpholexical in the sense that it is restricted to particular constructions, i.e., plurals formed with /-e/ (but not other plurals) and the construct state. A pretheoretical characterization such as this is useful in that it can help us relate known facts to specific theories. It is limited, however, in the sense that it does not allow the analyst to rigorously test the generalization. What if all affixes in a language trigger an exchange? It could still be said that the exchange is always correlated with a specific set of morphemes, or observed in a particular set of constructions.

A more productive approach to understanding Anderson and Browne's generalization is therefore to classify analyses of a given phenomenon on the basis of their theory-internal characteristics. These characteristics, and not pretheoretical observations, can be used to correctly classify morpho-phonological alternations, including exchange rules. A step in this direction is made in (Alderete, 2001a) by developing a set of so-called transderivational antifaithfulness constraints for morphologically-triggered phonological alternations. Antifaithfulness theory is an extension of the theory of faithfulness constraints (especially that of (McCarthy and Prince, 1995)), because antifaithfulness constraints are built from independently motivated faithfulness constraints. In particular, an antifaithfulness constraint is the wide-scope negation of the proposition expressed by a corresponding faithfulness constraint. Transderivational antifaithfulness constraints are defined on a base-output correspondence relation in the sense of (Benua, 2000). It is this OO-relation that encodes a base-derivative relationship characteristic of input-output mappings in morphology and underscores the morphological function of the alternation.

The analysis of the Dholuo voicing exchange presented in (Alderete, 2001a) makes clear this morphological function. The exchange is a two-step circular chain motivated by the need to satisfy OO¬IDENT(VOICE), the antifaithfulness constraint requiring a violation of IDENT(VOICE) in certain derived forms (see the appendix for constraint definitions, here and below). For example, the mapping from singulars like *gɔ*[t] to plurals ending in /-e/ is controlled by a particular output-to-output correspondence relation. This specific relation encodes the morphological relationship between the base singular and the plural form *gɔ*[d]-*ɛ*. This same relation provides the formal basis for defining the corresponding segments referred to by OO¬IDENT(VOICE). The root-final alternations satisfying OO¬IDENT(VOICE) are thus a morphological mark in the same sense that the insertion of the suffix /-e/ is: it is contributed in the same base-derivative mapping, which is governed by constraints defined on the same correspondence relation. The role of constraints defined on output-to-output correspondence will be shown below to allow for a theory-internal characterization of 'morphologically motivated' phonological alternations.

The introduction of antifaithfulness constraints is of course a departure from constraint conservativity (1c), which makes the conclusion of the CF argument invalid. To avoid this for phonological systems, (Alderete, 2001a) assumes that antifaithfulness constraints can only be defined on OO- and BR-correspondence. The latter type of constraint is employed in the analysis of well-known non-automatic base alternations in reduplicative constructions (Alderete et al., 1999). The stipulated absence of IO-antifaithfulness, however, does not in any way constitute an explanation of Anderson and Browne's generalization. The rest of this article can be understood as a proposal to remove this stipulation from antifaithfulness theory and rule out non-morphological exchanges with principles of learning.[1] In particular, IO-antifaithfulness is allowed to generate certain phonological exchanges, which are characterized directly below. But since they cannot be learned, grammars producing phonological

[1] It should be noted that (Moreton, 2004) anticipates a role for learnability in detecting non-computable mappings, which is a different approach to ruling out non-morphological exchange rules discussed further in 8.2.

exchanges will never be represented in the minds of language learners. This filtering effect is the result of the following hypothesis, stated more generally.

(3) The learnability filter (LF) hypothesis

> *In normal language learning, certain logically possible grammars are not learnable because the principles that underlie language learning do not allow learners to posit these grammars from overt data and learning experience.*

To test this hypothesis for exchange rules, a conceptual system is required for classifying phonological versus morphological analyses. The terms p-compatible and m-compatible proposed below are not intended as general characterizations of 'phonological' or 'morphological' processes in linguistic typology. Rather, they are mechanisms for classifying analyses of certain morpho-phonological alternations in the numerous linguistic systems explored below. No effort is made, for example, to characterize concatenative morphology, though it is clearly morphological in nature.

(4) Conceptual system for classifying morpho-phonological alternations

a. P-compatible: an analysis of a set of mappings that fully accounts for the data with just the interaction of input-referring constraints and markedness

b. M-compatible: an analysis of a set of mappings that makes crucial use of one or more constraints defined on output-to-output correspondence; 'crucial use' entails that at least one OO-constraint outranks a markedness constraint or an input-referring constraint, and that the data accounted for with this ranking is not accounted for by some other ranking

c. Mp-ambiguous: a set of mappings that can receive both a p- and m-compatible analysis

An analysis of some collection of mappings is p-compatible if the mappings can be accounted for with exclusive use of markedness, IO-faithfulness, and IO-antifaithfulness. The easiest test for p-compatibility is to remove all OO-constraints (i.e., constraints defined on an output-to-output correspondence relation) from a comparative tableau (Prince, 2003) and ask if the data can still be accounted for. As illustrated below, the inclusion of nonconservative constraints in the definition of p-compatible allows for exchanges that are phonological in nature, which, in turn, enables the LF hypothesis to be tested for exchange rules. An analysis is m-compatible if a subset of the data is accounted for with either OO-faithfulness or OO-antifaithfulness. Intuitively, this means that at least one constraint defined on an OO-correspondence relation is responsible for a paradigm uniformity effect (OO-faithfulness) or a morphologically motivated phonological process (OO-antifaithfulness). Without the OO-constraint, it would either not be possible to account for the alternation, or it would be accounted for with some other constraint, a fact that can be easily seen in a comparative tableau. P- and m-compatibility are not mutually exclusive, however, as there are alternations that can be analyzed as phonological, without OO-constraints, and are still consistent with an analysis that does employ OO-constraints. I will refer to such systems as mp-ambiguous systems.

In the schematic illustrations below and in later sections, linguistic systems are represented with paradigms, adapted from realistic systems but controlled for a number of factors that would complicate learning unnecessarily. All paradigms are based on two roots /pit/ and /pud/, shown in rows, and the constructions they occur in are given in columns. It is also necessary sometimes to indicate the base-output relationships that exist in order to evaluate OO-correspondence constraints. To focus the application of certain constraints, alternations are limited exclusively to [voice] specification in the root-final consonant. Lastly, the root vowel /i/ is used to indicate an underlying voiceless /t/, and root /u/ to indicate an underlying voiced /d/. This is just a typographic devise, however, to clarify certain problems in learning the lexicon, so /pit/ and /pud/ are effectively minimal pairs in all of the systems below.

Henceforth, unambiguous p-compatible systems will be referred to as 'phonological', and unambiguous m-compatible systems will be referred to as 'morphological', and I will dispense with the scare quotes. In the case of an ambiguous system, the terms phonological and morphological will only be used to characterize specific analyses, i.e., specific rankings of constraints. To illustrate an unambiguously phonological system, consider the context-free exchange of [voice] in every root-final consonant in the system below. While learning the inputs here will be non-trivial (a point discussed below in 6.2), this set of mappings is p-compatible because it can be analyzed by simply ranking IO¬IDENT(VOI) at the top of the hierarchy. No OO-constraints are needed. However, it is not m-compatible. There are no surface alternations, so OO-antifaithfulness cannot be involved, and there is no crucial use of OO-faithfulness that can be shown to be distinct from the necessary use of IO-antifaithfulness for the base forms.

(5) Exchange across-the board

Inputs	*Base*	*Forms derived from simplex base*		
	___#	___+*a*	___+*ka*	___+*kae*
/pit/	pi[d]	pi[d]-a	pi[d]-ka	pi[d]-kae
/pud/	pu[t]	pu[t]-a	pu[t]-ka	pu[t]-kae

The second example illustrates a more subtle p-compatible system, where a voicing exchange is observed in all forms in which the root-final consonant appears in an unstressed syllable. The system depicted below has regular penultimate stress, allowing an exchange in words with suffixes that require the root-final consonant to be in a syllable other than the penultimate syllable. This system can be analyzed with a top-ranked positional faithfulness constraint (Beckman, 1999) that preserves featural identity in stressed syllables, IO-IDENTSTRSYLL(VOICE), and inserting IO¬IDENT(VOICE) in the next position in the hierarchy. Thus, positional faithfulness requires faithful treatment of the root-final in the monosyllabic bases and words with mono-syllabic C-initial suffixes. But, when vowel-initial or disyllabic suffixes push the root-final outside of the penult, an exchange is observed. This ranking is a phonological analysis because no OO-constraints are needed. Of course the system is m-compatible too, because the exchange could be an effect of OO¬IDENT(VOICE) defined for a particular morphological class, namely the class of words containing the suffixes /-a/ and /-kaga/. Therefore it is mp-ambiguous; there are two ways to approach this system.

(6) Phonological elsewhere: exchange in unstressed syllables

Base	*Forms derived from simplex base*		
___#	___+*a*	___+*ka*	___+*kaga*
pít	pí[d]-a	pít-ka	pi[d]-kága
púd	pú[t]-a	púd-ka	pu[t]-kága

Other notions of 'phonologically conditioned' come to mind beyond these elsewhere cases, for example systems in which the exchange is conditioned by the presence of some segmental structure, e.g., an exchange of [voice] before the natural class of obstruents. These cases turn out to be rather complex, and depend in part on the constraints available for ranking. They are treated in some detail in subsection 5.2.

The main goal of this article is to show that Anderson and Browne's generalization, given the conceptual system of classifying analyses above, and the learning model introduced below, can be explained as a filtering effect of learnability. In particular, the typological space generated by antifaithfulness theory includes systems with p-compatible analyses of exchange rules. But a learner in this model literally cannot posit such an analysis, because of certain learning biases and also the structure of the linguistic theory of exchange rules, antifaithfulness. The next section develops the learning model, and sections 5 and 6 investigate the LF hypothesis in linguistic typologies that are designed to test its predictions.

3. Background

3.1 Basic assumptions of the learning model

In the learning model, a grammar is understood as the combination of a lexicon that provides the input representations of the generative component and a constraint ranking that accounts for the mappings of inputs onto outputs, i.e., the 'processes' of a language. Grammar learning therefore involves a learner that, given the overt forms of a language, and a set of well-formedness constraints, searches through a space of possible grammars (i.e., possible lexicon/ranking combinations) and selects a grammar that is consistent with the overt data and meets certain additional criteria or learning-theoretic biases. The learning processes below define the core operations of the assumed model, adapted largely from models of phonotactic learning, (Boersma and Hayes, 2001), (Hayes, 2004) (Prince and Tesar, 2004), and (Tesar and Smolensky, 2000).

(7) Core algorithms in assumed learning model

a. Identity mapping: for overt forms that do not alternate, their surface forms are used as inputs

b. Production-directed parsing (PDP): given a lexicon and a constraint ranking, use this interim grammar to select the optimal output in language production

c. Error-detection: given the output of PDP and the observed correct form, learners detect errors and use them to formulate winner-loser pairs

d. Biased constraint demotion (BCD), (Prince and Tesar, 2004)): given a set of winner-loser pairs, BCD finds the most restrictive grammar by recursively ranking constraints that favor no losers and that instantiate certain learning biases, including markedness>>faithfulness.

Overt forms are used to project inputs under (7a) only when they do not alternate at the surface. When a morpheme has more than one surface allomorph, more sophisticated input learning procedures are required (3.2). Lexicons created in this way can be modified later via surgery learning (Tesar et al., 2003), but they are employed as part of an interim grammar that generates errors in language production. These errors are employed to construct winner-loser pairs that provide the primary data for selecting a constraint ranking with (7d). The dynamic nature of error-generation and grammar learning will be illustrated in 3.3.

An important theoretical assumption in this model is the bias for markedness>>faithfulness (M>>F) in BCD. The inclusion of antifaithfulness constraints in the set of constraints a learner has at her disposal raises the question, how are IO- and OO-antifaithfulness constraints implicated in the M>>F bias? The hypothesis explored here is that IO-antifaithfulness is treated like IO-faithfulness, because it refers to input representations, but OO-antifaithfulness is classed with markedness, consistent with the comparison between OO-faithfulness and markedness made in (McCarthy, 1998) and (Tessier, 2007). It turns out that this hypothesis, which is the null hypothesis in a sense, provides an important basis for favoring morphological analyses in the mp-ambiguous cases examined below in sections 5 and 6.

3.2 Contrast analysis

When there is more than one surface allomorph for a given morpheme, contrast analysis is used to fill in the initial lexicon.

(8) Contrast analysis ((Tesar, 2006), (Alderete et al., 2005), (Merchant and Tesar, 2006))

a. Contrast pair: a pair of words formed from two morphemes in the same morphological environment

b. Faithful contrasting feature (FCF): for a given contrast pair, a feature F that the contrasting morphemes have different values for F on the surface, where each surface realization of F is faithful to the underlying value of its member of the contrast pair

c. Initial lexicon: given a paradigm of surface forms and a lexicon composed of inputs of forms specified for non-alternating features, create contrast pairs and examine them for FCFs; when/if there is only one FCF for contrast pair, specify that feature for the pair in their respective input forms

Contrast analysis is used here to specify the underlying feature [voice] in a root-final segment that may alternate for phonological or morphological reasons in the linguistic systems examined below. These systems do not present special challenges in terms of selecting a unique phonological feature, since only one feature can vary. The systems with exchange rules, however, do pose an interesting problem for filling in the lexicon, because exchange rules allow for the existence of contrast pairs that provide conflicting information about how to specify a feature. Two proposals are developed in section 4 that allow learners to exploit FCFs, even when this kind of conflict arises. One of these proposals involves identifying the base of a base-output correspondence relation as the construction to use first in filling in the lexicon. While the use of constraints defined on output-to-output correspondence is technically inconsistent with conditions necessary to ensure the existence of FCFs, it provides a tractable method for exploiting FCFs when the larger system seems to violate the FCF property globally (i.e., that FCFs must exist), which is a possibility conjectured in (Tesar, 2006).

3.3 Classifying mappings by their correspondence relations

Any approach to exchange rules, and morpho-phonological alternations in general, will have to account for their non-automatic nature, i.e., the fact that they are tied to particular morphological constructions and therefore do not apply across the board. One type of analysis of non-automatic alternations in OT involves the identification of distinct correspondence relations (e.g., (Itô and Mester, 1999), (Benua, 2000)), and recent work on learning lexical stratification has proposed distinct input-to-output correspondence relations as a means of learning the phonologies of different lexical strata (Pater, 2005).[2] The proposal below builds on this approach by extending it to the analysis of non-automatic morpho-phonological alternations based on output-to-output correspondence relations. Since exchanges rules are, by hypothesis, motivated by the satisfaction of OO-antifaithfulness constraints, learning non-automatic

[2] The use of the term 'distinct correspondence relation' as opposed to the more general term 'coindexation of constraints' that is now commonplace in discussions of strata and exceptions is intended to maintain the claims of (Itô and Mester, 1999) and (Benua, 2000), namely that only constraints that are defined on a correspondence relation can be involved in the analysis of distinct strata. If, on the other hand, markedness constraints can be coindexed by form, it is easy to see that CF argument is invalid. Particular markedness relations can be linked to particular phonological types, producing exchanges as the effect of reversals of markedness relations in specific phonological environments.

alternations is fundamentally a problem of correctly classifying base-derivative pairs by their OO-correspondence relations.

The pattern of non-automatic devoicing is given below to illustrate the problem. In this example, assume that devoicing is the result of the antifaithfulness constraint, OO¬MAX(VOICE), which requires a deletion of [voice] in the root of a derived form if its base has a corresponding [voice]. The force of this constraint must be limited to specific base-derivative pairings, i.e., formations with /-kae/. This is achieved by distinguishing OO1- and OO2-correspondence. The problem posed by this example thus reduces to classifying the base-derivative pairs as either OO1- or OO2-correspondence, and ranking the two anti-Max constraints, OO1¬MAX(VOICE) and OO2¬MAX(VOICE), accordingly.

(9) Non-automatic devoicing

Base	*Forms derived from simplex base*	
___#	___+ka_1	___+kae_2
pit	pit-ka	pit-kae
pud	pud-ka	pu[t]-kae

A tractable way of classifying mappings by their correspondence relations is to employ inconsistency detection, a method developed originally in (Tesar, 2004) for resolving structural ambiguity in overt forms, and applied in (Pater, 2005) to the problem of lexical stratification. The problem below applies the same reasoning to non-automatic alternations. In particular, learning starts under the presumption that all base-derivative pairs are related by the same OO-correspondence relation. But the accumulation of errors may lead to inconsistency, which is informative, because it tells the learner that something is wrong with the inputs. One response to inconsistency when learning alternations is to use a protocol called surgery learning (Tesar et al. 2003). Surgery learning responds to inconsistency by holding onto inconsistent winner-loser data, and changing the phonological content of the input. However, in this case, a simple change of phonological structure is insufficient. Because the two roots above are miminal pairs (note: vowel quality is orthographic only) and the phonological structure of the affix is irrelevant, changing the underlying representations for the roots or affixes will not help. The extension of Pater's approach to lexical stratification is to allow for a more radical type of surgery learning, where surgery on an affix's input specification for OO-correspondence can be modified, dubbed here as 'surgery on R'.

(10) Surgery on R

> Given inconsistency, and after surgery on a linguistic form has been attempted and revoked, the input of an affix may be altered to change the OO-correspondence relation that relates its base and the derivative formed with said affix. This change is applied to the last winner-loser pair added to the set of learning evidence.

All affixes are assumed to trigger the same correspondence relation initially, but surgery on R can split them into classes in response to inconsistency. Surgery on R is illustrated immediately below by applying the package of algorithms introduced in this section to non-automatic devoicing in (9). To make the learning setting more realistic, in addition to standard IO- and OO-IDENT constraints for [voice] and OO¬MAX(VOICE), the learner is assumed to have two markedness constraints: *d]$_\sigma$, which bans syllable-final voiced stops, and *d#, which specifically bans voiced stops word-finally (all constraints are defined in the appendix).

Learning of this system begins by building an initial lexicon. Three nonalternating morphemes are found, so they are identity-mapped into the lexicon (11a). The single alternating morpheme, *pud ~ put*, appears in two contrast pairs (11c). Note that the root vowels are just typographic markers of underlying voicing introduced for expository reasons, so the forms in (11c) are phonological minimal pairs that appear in the same morphological context. There must therefore be a faithful contrastive feature, and [voice] is selected as a means of distinguishing inputs /pit/ and /pud/.

(11) Building the initial lexicon

a. Nonalternating morphemes: /pit/, /-ka/, /-kae/
b. Alternating morpheme: *pud ~ put*
c. Contrast pairs: (*pit, pud*), (*pit-ka, pud-ka*)
d. Contrast analysis: [voice] is FCF, hence /pud/

Because of the bias for markedness>>faithfulness in Biased Constraint Demotion, the initial constraint hierarchy posits all markedness constraints at the top (this results from applying BCD to the null set of winner-loser pairs (Prince and Tesar, 2004)). Applying PDP successively on the data in (9), from left-to-right, top-to-bottom, the second form results in an error because with all markedness constraints at the top, a syllable-final contrast in [voice] is neutralized. This is symbolized in (12b) below with the mapping shown for the interim grammar's constraint hierarchy, = ℂℍ. This results in the winner-loser data, or an ERC vector, in (15a) below. Given that the OO-constraints are irrelevant for these base forms, the only consistent grammar is one that ranks IO-IDENT(VOI) at the top of the hierarchy (12d), effectively licensing a voicing contrast.

(12) Grammar learning part I: faithfulness high

a. Initial hierarchy: {*d]$_\sigma$, *d#] } >> { IO-IDENT(VOI), OO¬MAX(VOI), OO-IDENT(VOI)}
b. PDP: *pit, pud* Error! (ℂℍ: /pud/ → [put])
c. ERC vector for *pud ~ put*: (15a) below
d. BCD: IO-IDENT(VOI) >> { *d]$_\sigma$, *d#], OO¬MAX(VOI), OO-IDENT(VOI)}

No further changes to the constraint hierarchy are needed until PDP works on *put-kae*, which, given the ranking in (13a), produces an error, and adds the ERC vector (15b) to the winner-loser data. The learner has no choice at this point but to insert the antifaithfulness constraint OO¬MAX(VOI) at the top of the hierarchy, since it is the only constraint, for the data in (15a-b), that favors no losers.

(13) Grammar learning part II: OO-mutation constraint high

a. New grammar: IO-IDENT(VOI) >> { *d]$_\sigma$, *d#], OO¬MAX(VOI), OO-IDENT(VOI)}
b. PDP: *pit, pud, pit-ka, pud-ka, pit-kae, pud-kae* Error! (ℂℍ: /pud-kae/ → [pud-kae])
c. ERC vectors for *pud ~ put* and *put-kae ~ pud-kae*: (15a-b) below
d. BCD: OO¬MAX(VOI)>> IO-IDENT(VOI) >> {*d]$_\sigma$, *d#], OO-IDENT(VOI)}

This application of BCD results in a new grammar, (13d/14a), which essentially requires all root-final voiced segments in suffixed forms to be devoiced. This invariably leads to an error with *pud-ka*, which, when added to the winner-loser data in (15), leads to inconsistency. Morphological devoicing is restricted to derivatives with /-kae/, but OO¬MAX(VOI) at the top requires it across-the-board. Because every constraint favors a loser in at least one winner-loser pair, none of them can be inserted at the top by BCD. Importantly, surgery on the featural content of /pud/ will not get the learner out of this problem. Doing so only reverses the W/L marks for IO-IDENT(VOI), but does not resolve inconsistency.

(14) Grammar learning part III: inconsistency

a. New grammar: OO¬MAX(VOI) >> IO-IDENT(VOI) >> {*d]$_\sigma$, *d#], OO-IDENT(VOI)}
b. PDP: *pit, pud, pit-ka, pud-ka* Error! (ℂℍ: /pud-ka/ → [put-ka])
c. ERC vectors for *put ~ pud, put-kae ~ pud-kae,* and *pud-ka ~ put-ka*: (15a-c) below
d. BCD: Inconsistency!
e. Surgery on form: change /pud/ to /put/ (= reverses violations for IO-IDENT(VOI))
f. BCD: Inconsistency!

(15) Cumulative winner-loser data for non-automatic devoicing

Base, Input	Winner	Loser	*d]$_\sigma$	*d#	IO-IDENT	OO¬MAX(VOI)	OO-IDENT
a. *NA*, /pud/	pud	put	L	L	W	e	e
b. pud, /pud-kae/	put-kae	pud-kae	W	e	L	W	L
c. pud, /pud-ka/	pud-ka	put-ka	L	e	W	L	W

At this point, surgery on R is called for, because all other learning options, i.e., changing the ranking and changing the phonological structure of the input, have been tried and failed. Surgery on R requires changing the OO-correspondence relation of the affix of the input of the last winner-loser pair added to the stack (just as surgery on linguistic form does), which is /-ka/. Assuming there are at least two distinct OO-correspondence relations, all derivatives containing /-ka/ will therefore stand in correspondence with their base via OO2-correspondence. We also assume, for concreteness, that UG actually has duplicates of all faithfulness and antifaithfulness constraints, which become relevant now to the problem of selecting a consistent grammar.[3]

The revised OO-correspondence relations are shown in the comparative tableau below with subscripted indexes: derivatives with /-kae/ retain OO1-correspondence, but /-ka/ derivatives are changed to OO2-correspondence. As a result, two previously dormant constraints, i.e., the two constraints defined on OO2-correspondence, become relevant, and the marks previously allotted for *pud-ka ~ put-ka*, are transferred to these constraints, as shown with the arrows and strikeout lines of the prior marks. The striked-out marks are now simply 'e' here and in parallel examples below, since the OO1-constraints are irrelevant to the words formed with /-ka/. Grammar-learning then proceeds by ranking the two OO-constraints that favor no losers, OO1¬MAX(VOI) and OO2-IDENT(VOI), at the top, and then inserting IO-IDENT(VOI). This leads to no further errors. Inconsistency has been resolved by allowing correspondence relations to be reclassified, and the system with a non-automatic alternation has been learned.

[3] An alternative to this assumption, developed in (Pater, 2005), is to assume that constraints are actually created in response to inconsistency detection. I opt for the prior existence of constraints because this appears to be consistent with related assumptions for conjoined constraints, and further makes it possible to reclassify correspondence relations as a simple extension of surgery learning. No crucial results discussed in this article hinge on this decision, but since the duplicated constraints are innate, it seems to make the prediction that the specific number of phonological levels, which could be specified as any natural number other than 0 or 1, is fixed universally.

(16) Grammar learning part IV: surgery on R

a. Surgery on R: OO-relation for /-ka/ derivatives from R_{OO1} to R_{OO2}
b. New set of ERC vectors: (a-c) below
c. BCD: { OO1¬MAX(VOI), OO2-IDENT(VOI)} >> {IO-IDENT(VOI), OO2¬MAX(VOI), OO1-IDENT(VOI)} >> {$*d]_\sigma$, *d#]}
d. PDP: *pit, pud, pit-ka, pud-ka, pit-kae, pud-kae, put-kae* (No errors.)

(17) Cumulative winner-loser data for non-automatic devoicing, with surgery on R shown

	Base	Input	Winner	Loser	$*d]_\sigma$	*d#	IO-IDENT	OO1¬MAX	OO2¬MAX	OO1-IDENT	OO2-IDENT
a.	*NA*	/pud/	pud	put	L	L	W	e	e	e	e
b.	pud_1	/pud-kae_1/	put-kae_1	pud-kae_1	W	e	L	W	e	L	e
c.	pud_2	/pud-$ka_{1\to2}$/	pud-ka_2	put-ka_2	L	e	W	~~L~~	L	~~W~~	W

The protocol below summarizes the learning procedures applied here and below with subsequent learning examples.

(18) Application of the learning principles

a. Learning inputs part I: for all non-alternating forms (i.e., pre-segmented affixes and roots), identity-map surface forms to inputs
b. Learning inputs part II: for remaining forms, use contrast analysis to project alternating features and inputs (modified below for conflicting contrast pairs)
c. Grammar learning: using left-to-right, top-down presentation order, apply PDP/BCD to the entire system until no more errors are produced or until an inconsistent set of winner-loser pairs is produced; if inconsistency is found, apply surgery on linguistic form to the last winner-loser pair added to the list of errors; if surgery on forms is unsuccessful, apply surgery on R; reapply PDP/BCD until no more errors are produced

The learning examples discussed below apply a specific order of processing the data, and this may limit the generality of the results. While it is true that slightly different rankings may be learned in some cases, it is unlikely that different lexicons will be learned as a result of data order, as it is clear that surgery on linguistic forms (the only means of making lexical changes beyond the initial lexicon) has no utility at all in these learning examples. Further, presentation order was altered for the crucial cases below in section 5 involving mp-ambiguity, with no substantive change to the results. It does not appear to be the case, therefore, that presentation order will have an impact on the classification of a system as morphological or phonological.

Finally, this learning protocol is designed to learn non-automatic alternations tied to specific morphemes, and not other important aspects of morphological systems, like identifying the base of a complex form or segmenting words into their component morphemes. While these problems undoubtedly interact with learning alternations, these problems are set aside, with the hope that progress on the problem of learning non-automatic alternations will support future research on these independent problems.

4. Contrast analysis in a system with segmental exchanges

4.1 Introduction

Before the predictions of the LF hypothesis for exchange rules can be explored, the groundwork needs to be laid for learning this particular type of mapping. In particular, it is not fully determined at this point how contrast analysis will deal with the full reversal of a surface contrast exhibited in exchanges. Unlike neutralization, which effectively eliminates contrast pairs, chain shifting alternations may produce conflicting contrast pairs, as illustrated below.

(19) A system with conflicting contrast pairs

Base	*Forms derived from simplex base*	
___#	___+*ka*	___+*kae*
pit	pi[d]-ka	pi[d]-kae
pud	pu[t]-ka	pu[t]-kae

Contrast analysis identifies three contrast pairs in the above data: (*pit, pud*), (*pid-ka, put-ka*), and (*pid-kae, put-kae*). This data shows that there is a contrast between the two roots and that this contrast is realized by [voice], but it enumerates conflicting evidence for how to represent this feature in the underlying representation. The learner could use the simplex bases as its guide to the underlying representations, and try to learn a phonology that produces the root-final changes in square brackets. Or it could use the derived forms to learn the inputs, and instead change the base root-finals.

Clearly some mechanism is needed to process the conflicting information presented by exchange rules. Contrast analysis helps to clarify what that mechanism might be by finding conflicting contrast pairs. With a list of contrast pairs, the problem of processing the conflicting information in systems with exchanges can be modeled as making an intelligent guess as to the correct contrast pair to use in building an initial lexicon. Three plausible approaches to this problem are considered here.

(20) Principles for working with conflicting contrast pairs

a. Base Rules: given a set of contrast pairs that provide conflicting information about the specification of a FCF, and given a base-derivative relation defined on output-to-output correspondence, if one of the contrast pairs is from the construction that the base appears in, use the information in the contrast pair from the base construction

b. Majority rules: given a set of contrast pairs that provide conflicting information about the specification of a FCF, use the set of contrast pairs that suggests the same FCF specification and has the largest number of members

c. Random selection: given a set of contrast pairs that provide conflicting information about the specification of a FCF, select one of them at random and use its information to construct specify the FCF

To illustrate the difference between the first two, consider how they apply to the system above in (19). Base rules predicts that the roots will be projected as /pit/ and /pud/, because this specification of [voice] is exhibited in the base contrast pairs. Majority rules, on the other hand, predicts that /pid/ and /put/ will be the underlying representations for the two roots, because there are two contrast pairs that suggest them and only one that suggests the opposite specification for [voice].

Both of the first two approaches appear to have some inherent plausibility. Major rules is based on the presumption that there is some importance to the total number of pairs that suggest a lexical hypothesis. At least functionally speaking, it may be easier to develop a phonology for a smaller set of forms (i.e., the contrast pairs that are smaller in number) than a larger set of forms. Using the set of contrast pairs that are greater in number to specify inputs means that learners only have to select a phonological grammar for the smaller set of contrast pairs, which may very well be easier to generalize.

The plausibility of base rules stems from some suggestive evidence of the primacy of base forms. In the child language development literature, there are some well-documented cases in which children learn 'semantically primary' forms first, like singulars before plurals, and present tense before past tense, etc. (see (Bybee Hooper, 1979) for references and discussion). Since there is a cross-linguistic tendency for morphological bases to be semantically primary, it would make sense that learners use them first for building a lexicon because of this developmental pattern. The learning problems studied here do not represent developmental stages, but the principle of base rules in essence encapsulates this primacy for base forms.

The learnability test below is designed to distinguish these two hypotheses in learning exchange rules, with the hope of producing a functioning system for learning exchanges. Further details, including the potential interaction between these principles, and validating the guiding assumptions that motivate them, will have to be the focus of future work.

4.2 Test typology I

The approaches to resolving conflicting contrast pairs are explored below in a controlled typological space. Each linguistic system in this typology is fed to the learning model, under different assumptions about resolving conflicting contrast pairs, in order to better understand base rules and majority rules, and further, determine if exchanges are learnable at all in this model.

Since base rules singles out contrasts in the base construction, each system must encode base-derivative relationships, which are modeled here using OO-correspondence along the lines of (Benua, 2000). Majority rules, on the other hand, makes decisions based on the total number of uniform contrast pairs, so the linguistic systems must differ in the total number of contrast pairs for the two roots. Test typology I was built with these requirements in mind, as shown below. Following the data organization in section 3, each system in the test typology contains two paradigms with four members: a simplex CVC base, an inflected form CVC-*a*, and two derived forms CVC-*ka* and CVC-*kae*. The derived forms stand in an output-to-output correspondence relation with the base CVC form. The inflected form CVC-*a*, however, does not stand in correspondence with the base. The /-a/ forms are needed to vary the number of contrast pairs, but they are excluded from base-derivative relationships because doing so would explode the search space for the learner, with no real benefit to the test.

(21) Design of test typology I

	Base: CV[t/d]	CV[t/d]-a	Derived CV[t/d]-ka	Derived CV[t/d]-kae
Contrast	t/d	t/d	t/d	t/d
Exchange	*NA*	*NA*	d/t	d/t
Neutralization	t/t	d/d	t/t	t/t

All of the alternations occur in the root-final [t/d] position, and they all involve the feature [voice]. A contrast in [voice] is preserved in the Contrast and Exchange rows above, and neutralized as shown in the Neutralization row. The neutralizations and exchanges are the result of the free ranking of the following constraints: the markedness constraints from above, *d], *d#, an additional markedness constraint *VtV to motivate intervocalic voicing in CVCV forms, IO-IDENT(VOI), and two duplicates of three OO-constraints, OO-IDENT(VOI), OO¬IDENT(VOI), and OO¬MAX(VOI) (see appendix for constraint definitions). Test typology I was computed using OTSoft (Hayes et al., 2003) to factorially rank these constraints. In particular, there are phonological effects in which markedness constraints neutralize the [voice] constraints by dominating faithfulness, like final devoicing and intervocalic voicing. There are also morphological effects in the derived forms caused by OO-antifaithfulness: a high-ranking OO¬IDENT(VOI) can bring about an exchange of [t/d] with [d/t], and OO¬MAX(VOI) can result in syllable-final neutralization of t/d to t/t by requiring a deletion of a [voice] feature in a base [d].

Given these assumptions, the test typology includes thirty six different systems: 2 patterns for base * 2 patterns for inflected CVC-*a* * 3 patterns for derived CVC-*ka* * 3 patterns for CVC-*kae*. Each system encodes two base-derivative relationships in which exchanges are possible, and the range of contrasts for each system falls between four surface contrasts (fully contrastive) and zero contrasts (neutralization in all constructions). The actual systems are available in a PDF file weblinked to the author's webpage.

4.3 Learning results and discussion

The overt forms of individual systems from test typology I were input to the learning model introduced in section 3. The constraints at the disposal of the learner in this model were all the constraints used in constructing the factorial typology, plus the IO-antifaithfulness constraint, IO¬IDENT(VOI). The inclusion of this constraint makes possible certain linguistic analyses that are explored later in 6.2, but since test typology I does not contain any mp-ambiguous systems, the discussion below focuses on how precisely the learner produces a consistent morphological analysis of the given data.

The learning model was successful at producing a consistent grammar for all of the thirty six systems. In particular, surgery on R and the mechanisms proposed above for dealing with conflicting contrast pairs allowed the learner to produce a lexicon and constraint ranking for each system that correctly accounted for the given data. The errors needed to learn these systems ranged between 0 and 6, though some systems could have produced more errors if ranking decisions were made that are not fully determined by the model. Inconsistency was detected in twenty four of the thirty six systems, and it was always correlated with one or more non-automatic alternations in the system. The non-automatic alternations were always accounted for by ranking output-to-output constraints high relative to markedness constraints. Finally, surgery on phonological form (i.e., changing the value of [voice] in the input) never resolved inconsistency, but surgery on R always did.

As for conflicting contrast pairs, the two approaches, base rules and majority rules, can be distinguished by how conclusive they are in building the initial lexicon. In test typology I, paradigms produced conflicting contrast pairs whenever the base supported a [t/d] contrast and there was a morphological exchange of [voice], a combination of attributes that is observed in ten systems. As shown below, base rules is the only single principle that resolved all ten of these conflicts in setting up the initial lexicon.

(22) Information from contrast pairs

No conflict (so [voice] value is unambiguous)	26
Conflict resolved by base rules only	4
Conflict resolved by majority rules only	0
Conflict resolved by base rules and majority rules	6

This result is not direct evidence for base rules, because, as shown below, it is possible to resolve conflicting contrast pairs by a random selection procedure, and still converge on a consistent grammar. However, it does show that building a lexicon based on the principle of base rules is fully determinant, which may mean that the base rules approach is more efficient in more complex learning problems.

The rest of the discussion below focuses on specific systems to illustrate learning theoretical results of different types. The first system we look at exhibits a non-automatic exchange, to illustrate that the learning model can successfully learn such a system with surgery on R.

(23) Illustration: non-automatic exchange with /-kae/, four contrast pairs

Mini-inflectional paradigm		*Forms derived from simplex base*	
___#	___+*a*	___+*ka*	___+*kae*
pit	pit-a	pit-ka	pi[d]-kae
pud	pud-a	pud-ka	pu[t]-kae

The first step of building the lexicon involves distinguishing alternating and non-alternating forms, and finding all contrast pairs where a contrast is exhibited in the alternating morphemes, in this case the roots. Contrast analysis enumerates four pairs, since the roots differ in only the root-final segments in all four constructions (recall the root vowel is just a typographic marker). Since both base rules and majority rules point in the direction of underlying /pit/ and /pud/, these forms are projected as inputs.

(24) Building initial lexicon

a. Non-alternating forms: /-a/, /-ka/, /-kae/
b. Alternating morphemes: *pit ~ pid, pud ~ put*
c. Contrast pairs: (*pit, pud*), (*pit-a, pud-a*), (*pit-ka, pud-ka*), (*pid-kae, put-kae*)
d. Contrast analysis: [voice] is FCF, inputs /pit, pud/ by base rules and majority rules

Production-directed parsing processes the first two forms and produces an error for *pud*, because with *d] and *d# in the top stratum, word-final devoicing results from the learner's production grammar. BCD thus builds a new grammar with these constraints below IO-faithfulness (25d), and a similar run of PDP results in the demotion of *VtV (25e-g).

(25) Grammar learning part I: faithfulness high

a. Initial grammar: {*d], *d#, *VtV} >> all other correspondence constraints (by BCD)
b. PDP: *pit, pud* Error! (ℂℍ: /pud/ → [put])
c. ERC vector for *pud ~ put*: (28a)
d. BCD: { *VtV, IO-IDENT(VOI)} >> {*d], *d#} >> all other constraints
e. PDP: *pit, pud, pit-a* Error! (ℂℍ: /pit-a/ → [pid-a])
f. ERC vectors for *pud ~ put, pit-a ~ pid-a*: see (28a-b)
g. BCD: IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints

The new interim grammar has IO-faithfulness at the top of the hierarchy (25g/26a), putting it above markedness. This grammar correctly accounts for derivatives with /-ka/, but generates an error when it gets to the first /-kae/ derivative because the current grammar preserves the input (26b). Given this winner-loser data (28a-c), BCD must first rank an antifaithfulness constraint, because the winner-loser pair *pid-kae ~ pit-kae* adds information that makes it impossible to find a consistent ranking with just markedness and faithfulness constraints. However, by inserting OO1¬IDENT(VOI) at the top of the hierarchy, a consistent ranking is found (26d).

(26) Grammar learning part II: emergence of antifaithfulness

a. New grammar: IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints
b. PDP: *pit, pud, pit-a, pud-a, pit-ka, pud-ka, pid-kae* Error! (ℂℍ: /pit-kae/ → [pit-kae])
c. ERC vectors for *pud ~ put, pit-a ~ pid-a, pid-kae ~ pit-kae*: see (28a-c)
d. BCD: OO1¬IDENT(VOI) >> IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints

The new interim grammar also produces an error when it gets to the /-ka/ derivatives, because OO1¬IDENT(VOI) applies to these forms as well (27b). The accumulated set of ERC vectors, i.e., all the rows in (28), provides the basis for BCD to conclude that the system as currently analyzed is inconsistent, and no ranking of constraints is possible. (The inconsistent ERCs are the first crossed-out marks for the OO-constraints in (28), prior to surgery on R.) Surgery on the form of the last winner-loser pair is of no avail, since this would just invert the W/L values for IO¬IDENT(VOI) and IO-IDENT(VOI) without resolving the inconsistency. Surgery on R is then called, with the effect of changing the OO-correspondence relation for /-ka/ derivatives (the affix of last winner-loser pair added to the ERC data), shown in the subscripts for the mapping in (28d). This has the effect of making all OO1- correspondence constraints irrelevant for /-ka/ derivatives, and instead all constraints sensitive to them are defined on OO2- correspondence constraints, as shown with the rightward arrows. This re-classification of the correspondence relations enables BCD to find a consistent grammar (27g) that correctly accounts for the data and that generates no more errors.

(27) Grammar learning part III: inconsistency detection and surgery on R

a. New grammar: OO1¬IDENT(VOI) >> IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints
b. PDP: *pit, pud, pit-a, pud-a, pit-ka* Error! (ℂℍ: /pit-ka/ → [pid-ka])
c. ERC vectors for *pud ~ put, pit-a ~ pid-a, pid-kae ~ pit-kae, pit-ka ~ pid-ka*: (28a-d)
d. BCD: Inconsistency!
e. Surgery on form: no avail
f. Surgery on R: OO-relation for /-ka/ derivatives from R_{OO1} to R_{OO2}
g. BCD: {OO1¬IDENT(VOI), OO2-IDENT(VOI)} >> IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints
h. PDP: *pit, pud, pit-a, pud-a, pit-ka, pud-ka, pid-kae, put-kae* (No more errors.)

(28) Winner-Loser data for non-automatic exchange

Input, Base	Winner	Loser	*d]$_\sigma$	*d]#	*VtV	IO-IDENT	IO¬IDENT	OO1-IDENT	OO2-IDENT	OO1¬IDENT	OO2¬IDENT	OO1¬MAX	OO2¬MAX
a. /pud/, NA	pud	put	L	L	e	W	L	e	e	e	e	e	e
b. /pit-a/, NA	pita	pida	e	e	L	W	L	e	e	e	e	e	e
c. /pud-kae/, pud	putkae$_1$	pudkae$_1$	W	e	e	L	W	L	e	W	e	W	e
d. /pud-ka/, pud; $R_{1\rightarrow 2}$	pudka$_2$	putka$_2$	L	e	e	W	L	~~W~~ →	W	~~L~~ →	L	~~L~~ →	L

In this last example, base rules and majority rules pointed in the same direction for building the initial lexicon. The next case, based on the system in (19), contains conflicting contrast pairs and therefore suggests two possible lexicons.

(29) Illustration: exchange in all derived forms

Mini-inflectional paradigm		*Forms derived from simplex base*	
___#	*___+a*	*___+ka*	*___+kae*
pit	pi[d]-a	pi[d]-ka	pi[d]-kae
pud	pud-a	pu[t]-ka	pu[t]-kae

The fact that an exchange of the base root-final is observed in both derivatives leads to three contrast pairs (30c). Using base rules to project inputs leads to /pit, pud/, contrary to the prediction of majority rules. By taking this route, the learner generates the evidence in (32) to ultimately select the grammar in (31).

(30) Building the initial lexicon (base rules version)

a. Non-alternating forms: /-a/, /-ka/, /-kae/
b. Alternating morphemes: *pit ~ pid, pud ~ put*
c. Contrast pairs: (*pit, pud*), (*pid-ka, put-ka*), (*pid-kae, put-kae*)
d. Contrast analysis: [voice] is FCF, hence /pit, pud/ by base rules (not majority rules)

(31) Learned grammar, if base rules resolves conflicting contrast pairs

a. Lexicon: /pit, pud/, /-a/, /-ka/, /-kae)
b. Constraint hierarchy: OO1¬IDENT(VOI) >> IO-IDENT(VOI) >> {*d], *d#, *VtV} >> all other constraints

(32) ERC vectors for base-rules version of system in (29)

Input, Base	Winner	Loser	$*d]_\sigma$	*d]#	*VtV	IOIDENT	IO¬IDENT	OO1-IDENT	OO2-IDENT	OO1¬IDENT	OO2¬IDENT	OO1¬MAX	OO2¬MAX
a. /pud/, NA	pud	put	L	L	e	W	L	e	e	e	e	e	e
b. /pit-a/, NA	pita	pida	e	e	L	W	L	e	e	e	e	e	e
c. /pit-ka/, pit	pid-ka	pit-ka	L	e	e	L	W	L	e	W	e	e	e

This outcome is different, however, from the grammar that results if majority rules is used as the basis for the initial lexicon. Concretely, if the roots in the derived forms are used, i.e., /pid, put/, the resulting constraint hierarchy is rather different, including requiring a role for IO-antifaithfulness to account for the opposite [voice] specifications in the base forms.

(33) Learned grammar, if majority rules resolves conflicting contrast pairs

a. Lexicon: /pid, put/, /-a/, /-ka/, /-kae)
b. Constraint hierarchy: { *VtV, OO1¬IDENT(VOI)} >> IO¬IDENT(VOI) >> {*d], *d#} >> all other constraints

The ultimate analysis of the exchange in (33) is not phonological in the sense defined in section 2, because of the crucial role of OO1¬IDENT(VOI). The grammar is highly unintuitive, however, because IO-antifaithfulness mutates the root-final consonant in the base, and OO-antifaithfulness changes that specification back again in the derived forms. One wonders if new words entering a language structured like this would be treated in the same way. Of course these problems could be due to the existence of IO-antifaithfulness, which is a hypothesis like any other and could be rejected, but the fact that unituitive analyses such as these are never learned with base rules is instructive.

5. Testing the LF hypothesis

Now that a functional model exists for learning exchanges, this section explores the predictions of the learnability filter hypothesis by conducting two kinds of tests. The first test (5.1) involves a new test typology of exchange rules based on the classification of morpho-phonological alternations outlined in section 2. In particular, test typology II contains mp-ambiguous systems that can either have a morphological or phonological analysis. The LF hypothesis predicts that certain logically possible analyses can be eliminated in learning, favoring either a morphological or phonological analysis of exchanges. The specific prediction to be tested here is if all overt systems that exhibit an exchange will be analyzed as morphological through learning, consistent with Anderson and Browne's generalization. The second test (5.2) involves feeding the learning model linguistic systems that appear to be governed by a surface phonological generalization, to see how the model will classify the system. The grammars that are learned are then studied and compared with other logically possible grammars, again testing the prediction that exchanges can only be treated as morphological through learning.

5.1 MP-ambiguous systems

The discussion below focuses on the learning of mp-ambiguous systems, modeled after the 'phonological elsewhere' case from section 2, repeated below. In this system, stress generally falls on the penultimate syllable, if there is one, or the final syllable in monosyllabic words. Factoring in stress allows for the following phonological generalization: root-final consonants mutate in unstressed syllables.

(34) Exchange of t/d in unstressed syllables (uniform penultimate stress)

Base	*Forms derived from simplex base*		
___#	___+*a*	___+*ka*	___+*kaga*
pít	pí[d]-a	pít-ka	pi[d]-kága
púd	pú[t]-a	púd-ka	pu[t]-kága

Section 2 showed that systems like the one above can be treated as phonological, with a combination of IO-antifaithfulness and positional faithfulness, or morphological, by employing OO-antifaithfulness for a specific class of suffixes. Is it possible to systematically exclude the phonological analysis, given the assumptions of the learning model?

Test typology II was designed to answer this question. It was constructed by allowing the free ranking of all of the constraints employed in creating test typology I, together with positional faithfulness for stressed syllables, and then testing all mp-ambiguous systems. The forms of each system have roughly the same paradigm structure as test typology I, except that forms that contain the vowel-initial suffix /-a/ stand in an OO-correspondence relation. In test typology II, it is necessary to consider morphological analyses of exchanges in this environment in order to establish mp-ambiguities. The suffix /-kaga/ is also included to provide a construction in which the root-final appears in the coda position of an unstressed syllable.

The factorial typology of the constraints above, applied to these linguistic forms, produces fifty seven distinct systems. We focus, however, on a subset of these systems, namely the eighteen systems that exhibit a [voice] contrast in the simplex base, because only these systems support a morphological analysis of exchanges in suffixed structures, and the LF hypothesis requires this analysis for a system to be mp-ambiguous. Of these eighteen systems, eleven are analyzed as unambiguously morphological, i.e., the exchange is caused by OO-antifaithfulness, and no other analysis is possible. Four more systems are unambiguously phonological in nature: they do not contain exchanges and the observed neutralizations must be handled with markedness-faithfulness interactions. The remaining three systems are mp-ambiguous. That is, there is a viable phonological analysis that does not involve OO-correspondence constraints at all, and there is also a morphological analysis driven, at least in part, by OO-antifaithfulness. These systems are depicted schematically below, indicating the root-final for the *pit/pud* paradigms. (The actual forms of these systems can be viewed in the systems chart available from the author's webpage, numbered as shown below.)

(35) Mp-ambiguous systems

	___#	___+*a*	___+*ka*	___+*kaga*
a. (II.1)	t/d	d/t	t/d	d/t
b. (II.2)	t/d	d/t	t/d	t/t
c. (II.7)	t/d	d/d	t/d	d/t

All of these systems preserve the root-final of inputs /pit, pud/ in stressed positions (i.e., the simplex base and before /-ka/), but neutralize or exchange this contrast in unstressed positions. It turns out that in all three cases, the model predicts a grammar in which the exchange is morphologically motivated, and the degree of neutralization has no impact on this result. This result follows from the grouping of IO-antifaithfulness with IO-faithfulness in Biased Constraint Demotion, and OO-antifaithfulness with markedness, as explained in 2.1. Given a choice between ranking an OO-antifaithfulness constraint, or an IO-antifaithfulness constraint, the learner opts for the former, effectively preferring the morphological analysis. This bias is illustrated below with some selected examples, starting with the mp-ambiguous system in (34).

In this system, both roots alternate, so the contrast pairs in (36c) below will be used as a basis for projecting [voice] in inputs. By base rules, inputs /pit, pud/ are posited, which has non-trivial consequences for grammar learning.

(36) Building initial lexicon

a. Non-alternating morphemes: /-a/, /-ka/, /-kága/
b. Alternating morphemes: *pit ~ pid, pud ~ put*
c. Contrast pairs: (*pít, púd*), (*píd-a, pút-a*), (*pít-ka, púd-ka*), (*pid-kága, put-kága*)
d. Contrast analysis: [voice] is FCF, inputs /pit, pud/ by base rules

Initial grammar learning contends with the phonotactic facts, including the absence of final devoicing and intervocalic voicing, as shown below.

(37) Grammar learning part I: phonotactic learning

a. Initial grammar: {*d], *d#, *VtV} >> all other constraints
b. PDP: *pit, pud* Error! (ℂℍ: /pud/ → [put])
c. ERC vector for *pud ~ put*: (39a)
d. BCD: { *VtV, IO-IDENT} >> {*d], *d#} >> all other constraints
e. PDP: *pit, pud, pid-a, put-a* Error! (ℂℍ: /pud-a/ → [put-a])
f. ERC vectors for *pud ~ put, put-a ~ pud-a*: see (39a-b)
g. BCD: {IO-IDENTSTRSYLL, OO1¬IDENT}] >> {*d], *d#, *VtV} >> all other constraints

After PDP has produced the error in (37e), the learner has no choice but to elevate OO1¬IDENT in the ranking: it is the only way to get a consistent grammar for the /pud-a/ → [put-a] mapping, since IO-antifaithfulness is tied up with the loser mark for surface [put]. However, since there is no specific ordering between IO-IDENTSTRSYLL and OO1¬IDENT, the error below in (38b) can be produced. Because insertion of IO-IDENTSTRSYLL at the top of the hierarchy frees up both antifaithfulness constraints, IO¬IDENT and OO1¬IDENT end up in competition for second position in the constraint hierarchy. Here the bias of BCD against ranking input-referring constraints is crucial, and OO1¬IDENT is inserted next in the hierarchy. Since PDP produces no further errors, the ultimate grammar is one that treats the exchange as morphological, despite the availability of the phonological analysis.

(38) Grammar learning part II: morphological over phonological

a. Interim grammar: {IO-IDENTSTRSYLL, OO1¬IDENT}] >> {*d], *d#, *VtV} >> all other constraints
b. PDP: *pit, pud, pid-a, put-a, pit-ka* Error! (CH: /pit-ka/ → [pid-ka])
c. ERC vector for *pud ~ put, put-a ~ pud-a, pit-ka ~ pid-ka*: see (39a-c)
d. BCD: IO-IDENTSTRSYLL >> OO1¬IDENT >> {*d], *d#, *VtV} >> all other constraints
e. PDP: *pit, pud, pid-a, put-a, pit-ka, pud-ka, pid-kaga, put-kaga* End search

(39) Winner-loser data for #II.1, using base forms for inputs

Input, Base	Winner	Loser	$*d]_\sigma$	*d]#	*VtV	IO-IDENT	IO¬IDENT	IO-IDSTSYLL	OO1-IDENT	OO2-IDENT	OO1¬IDENT	OO2¬IDENT
a. /pud/, NA	púd	pút	L	L	e	W	L	W	e	e	e	e
b. /pud-a/, pud	púta	púda	e	e	L	L	W	e	L	e	W	e
c. /pit-ka/, pit	pítka	pídka	W	e	e	W	L	W	W	e	L	e

An interesting fact of this system is that it contains a mapping that cannot be rectified by inserting a markedness constraint or an IO-faithfulness constraint into a dominating position in the hierarchy. The ERC vector for the mapping /pud-a/ → [put-a] shows that the markedness constraints have no way to motivate devoicing in intervocalic positions, and, since the input /pud-a/ is posited, the winner cannot be favored by IO-faithfulness. The only way to motivate this mapping is to insert an antifaithfulness constraint above the constraints that favor losers, and BCD is decisive because OO1¬IDENT does not refer to inputs. It turns out that all of the mp-ambiguous systems lead to the generation of errors that present precisely this kind of situation. In II.2 (35b), this same mapping leads to the morphological analysis, and in II.7 (35c), the mapping /pit-kaga/ → [pid-kaga] again is neither motivated by markedness (since coda voicing is never favored by markedness), nor faithfulness (since this is an unfaithful mapping). The undominated constraints in these systems (positional faithfulness and certain markedness constraints) again set IO- and OO-antifaithfulness in direct competition in accounting for this data, requiring BCD to decide in favor of the morphological analysis. In sum, unfaithful mappings that cannot be motivated by markedness must be accounted for with antifaithfulness, capturing Moreton's (2004) original insight into this type of problem. BCD has a bias against ranking constraints that refer to inputs, and this bias always favors the morphological analysis.

5.2 Handling of surface phonological generalizations

Thus far, the model has been tested by feeding it linguistic systems that could either be phonological or morphological. How will the learner respond to exchanges that appear at the surface to be conditioned by purely phonological factors? Consider the system below.

(40) Surface phonological exchanges

Base	Pre-obstruent (mutation)	Elsewhere (no change)
pit	pid-ta, pid-si, pid-ke	pit-ra, pit-le, pit-e
pud	put-ta, put-si, put-ke	pud-ra, pud-le, pud-e

Two types of analyses of this data are logically possible. The learner can treat the system as a phonological exchange, provided she has the necessary constraints to do so. This finding would falsify the LF hypothesis. Or, the learner can treat an apparently phonological system as morphological by sorting the morphological constructions into a set of correspondence classes that accidently correlate with the observed phonological patterns. The systems explored below show that if the system is mp-ambiguous, BCD will select the morphological approach. On the other hand, if the learner is handed a system that is not mp-ambiguous, it is forced by the constraints themselves to classify the pattern as morphological, because the assumed set of universal well-formedness constraints cannot compute a phonological exchange.

As syllabification has non-trivial consequences for the treatment of this type of system, two typologically plausible syllabifications will be considered: one in which obstruent + sonorant clusters are allowed, and another without any onset clusters. In the first case, shown below with representative examples, it turns out that the system is mp-ambiguous, with the inclusion of onset faithfulness, a constraint that assigns a privileged faithfulness status for onset positions ((Lombardi, 1999), (Beckman, 1999)). Rather parallel to the 'phonological elsewhere' case above, the root-final exchange could be treated as the muted effect of IO-antifaithfulness, i.e., mutate in all contexts but onset positions, because pre-obstruent consonants must be dumped in the coda for syllable-related reasons. Or, all obstruent-initial suffixes could be assigned to an OO-correspondence class that predicts a morphological mutation in just this environment.

(41) Pre-obstruent mutation, with consonant clusters

Base	Pre-obstruent (mutation)	Elsewhere (no change)
pi.ta	pi[d].ta	pi.tra
pu.da	pu[t].ta	pu.dra

Given mp-ambiguity, the model again opts for the morphological analysis because of BCD. This result is shown in the comparative tableau below, which summarizes the learning steps as an accumulation of winner-loser data (i.e., error generation is parallel to above illustrations). When BCD is applied to the full set of winner-loser pairs, it first ranks onset faithfulness highest in the hierarchy, since it alone favors no losers. The remaining winner-loser pair (42b), forces a decision between IO- or OO-antifaithfulness, and just like with the stress-based generalizations above, it inserts OO1¬IDENT next because it does not refer to inputs and is thus classed with markedness constraints.

(42) Winner-loser data for pre-obstruent mutation, with onset clusters

Input, Base	Winner	Loser	*d]	*VtV	IO-IDENT	IO-ONSIDENT	IO¬IDENT	OO1-IDENT	OO2-IDENT	OO1¬IDENT	OO2¬IDENT
a. /pit-a/, NA	pi.ta	pi.da	e	L	W	W	L	e	e	e	e
b. /pit-ta/, pit-a	pid.ta	pit.ta	L	e	L	e	W	L	e	W	e
c. /pit-ra/, pit-a	pi.tra	pi.dra	e	e	W	W	L	W	e	L	e

The primary difference between this system, and the one depicted below, without onset clusters, is that the syllabification of the root-finals as codas in all suffixed environments eliminates the phonological analysis. This result follows from the nature of the constraints themselves, and not from specific assumptions in the learning model. As discussed in section 2, while antifaithfulness constraints produce exchanges, they do not have the ability to specify the phonological environment in their scope, parallel to an SPE rule package. Concretely, there are no constraints in *CON* of the type, IO¬IDENT /__C[-SON], that specifically require a faithfulness reversal before a specific phonological class. Because of this restriction on antifaithfulness, it is impossible to analyze the exchanges without correspondence classes, as those created by surgery on R in (44).[4]

(43) Pre-obstruent mutation, without consonant clusters

Base	Pre-obstruent (mutation)	Elsewhere (no change)
pi.ta	pi[d].ta	pit.ra
pu.da	pu[t].ta	pud.ra

[4] As careful readers of the antifaithfulness literature will know, some approaches to edge and positional privilege effects have built reference to edges and prominent positions into the definition of an antifaithfulness constraint (see (Alderete, 2001a) for some conjecture). If these types of positional privilege are indeed necessary in the output-to-output domain, then to fully relax the restriction on the domain of correspondence for antifaithfulness, these types of effects will have to be examined in the input-output domain. However, the principal effects of such constraints, e.g., mutate the stressed syllable, seem rather more plausible than antifaithfulness constraints with an *ad hoc* context.

(44) Winner-loser data for pre-obstruent mutation, without onset clusters

Input, Base	Winner	Loser	*d]	*VtV	IO-IDENT	IO-ONSIDENT	IO¬IDENT	OO1-IDENT	OO2-IDENT	OO1¬IDENT	OO2¬IDENT
a. /pit-a/, NA	pi.ta	pi.da	e	L	W	W	L	e	e	e	e
b. /pit-ta$_1$/, pit-a	pid.ta$_1$	pit.ta$_1$	L	e	L	e	W	L	e	W	e
c. /pit-ra$_{1\rightarrow 2}$/, pit-a	pit.ra$_2$	pid.ra$_2$	W	e	W	e	L	~~W~~ →	W	~~L~~ →	L

The inconsistency of this data is shown by the winner-loser marks prior to surgery on R: no ranking of constraints can account for both pre-obstruent mutations and the lack of mutation elsewhere. Surgery therefore introduces an arbitrary correspondence class that regulates the effects of two classes of OO-correspondence constraints, and allows for a consistent grammar to be selected, shown below. Any additional suffixes encountered that fail to trigger a mutation will then be classed as derivatives regulated by OO2-correspondence. Any additional mutating suffixes will not trigger special learning, as they are already accounted for by their initial classification as OO1-correspondence and the learned grammar below.

(45) Learned grammar for pre-obstruent mutation, without onset clusters

{OO2-IDENT, OO1¬IDENT} >> IO-ONSIDENT >> {*d], *VtV} >> all other constraints

In summary, linguistic systems that appear to have a phonological exchange, based on the natural class of a neighboring segment, cannot be analyzed as such in this mode because of the context-free nature of antifaithfulness constraints. Once again, the only possible analysis of the exchange data is a morphological one, consistent with Anderson & Browne's generalization.

One might object to this approach to the putatively phonological cases on the grounds that it does not appear to make predictions in the spirit of Anderson and Browne's original statement, namely that all attested exchange rules are morpholexical. The fact that the approach developed here actually predicts systems that appear at the surface to be consistent with an automatic phonological generalization appears to be a weakening of the proposal, and one that should not be entertained in the interest of predictiveness. My response to this is that the learnability theory developed here is indeed predictive and falsifiable. The learnability theory makes predictions about the analysis of overt data in the mind of the learner, and not surface facts directly. The cases studied above attest to the falsifiability of the hypothesis—the overt data of certain exchanges could very easily, in different theories, be treated as phonological. Furthermore, the predicted morphological analysis provides the cognitive basis for additional predictions of an empirical nature. For example, the contention that all exchange rules must be represented in the mind as a set of morphological relationships, formalized as output-to-output correspondence, provides the theoretical framework for an extension of those relationships to new words entering the language, or the existence of a morphological class that can resist historical sound changes.

A final point is that this objection is not an objection to the antifaithfulness analysis, but an objection to the characterization of the problem in terms of the analysis of exchanges rather than in terms of surface facts. If the goal is to try to predict a morphological analysis, rather than a phonological analysis, then we are stuck with the predictions of the morphological analysis. However, if one considers the various formulations of morpholexical rules in contemporary morphology, there is little basis to single out the approach taken here. Stratal OT (Kiparsky, 1998/to appear) also allows accidental phonological generalization to emerge in sets of morphological collocations. Likewise, there is no way to preclude a set of schemas in word-based morphology (à la (Bybee, 1988)) describing distinct morphological patterns that all contain a common phonological structure. Indeed, it would seem unwise to develop a principled means of excluding such accidental patterns, as non-automatic alternations often exhibit phonological patterns as a consequence of their history. For example, loans aside, the /t/ → [d] mutation in Dholuo appears to be restricted to CVC bases, a fact that (Okoth-Okombo, 1982) analyzes historically as rule-inversion of a prior rule of utterance-final devoicing. In seems therefore that any theory of morpholexical rules will predict phonological patterns like that in (40).

6. Further consequences

6.1 Automaticity and process classification

All of the linguistic systems studied above that required surgery on R might be classified as 'non-automatic', meaning essentially that they do not apply to all contexts they potentially could. In these systems, the non-automatic pattern always led the learner to select a morphological analysis by ranking OO-constraints defined on a particular correspondence relation in a dominating position. This correlation raises the question, is there something special about the automaticity of an alternation and whether or not it is classified as morphological or phonological? Are all automatic alternations classified as phonological, for example, in the model? As shown below, however, automaticity is not the only factor in process classification. Instead, it seems that certain learning biases in the model are decisive, as well as the availability of morphological or phonological analyses built into the well-formedness constraints themselves.

An automatic process is some process P, given a coherent context K, where P applies in K in all forms that contain K. Under this definition, German-style devoicing is an automatic phonological process: obstruents devoice syllable-finally across-the-board, as implied by the square brackets below.

(46) German-style devoicing: automatic and unambiguous

Base	*Forms derived from simplex base*	
___#	___+*a*	___+*ka*
pit	pit-a	pit-ka
pu[t]	pud-a	pu[t]-ka

The learning model classifies this pattern as phonological, however, not because of a surface apparent truism, but because of the unambiguous nature of the constraints. Concretely, contrast analysis requires that the roots /pit, pud/ be projected because of the contrast pair [pit-a/pud-a]. Since this is the correct lexicon, the rest of the grammar is learned with simple phonotactic learning. On the basis of the errors generated by production-driven parsing, shown below, the learner must rank IO-IDENT above *VtV, as shown in (48).

(47) Winner-loser data for German-style devoicing

Input, Base	Winne	Loser	*d]$_\sigma$	*d]#	*VtV	IO-Ident	IO¬Ident	OO1-Ident	OO2-Ident	OO1¬Ident	OO2¬Ident	OO1¬Max	OO2¬Max
a. /pit-a/, pit	pit-a	pid-a	e	e	L	W	L	W	e	L	e	e	e
b. /pud-a/, put	pud-a	put-a	e	e	W	W	L	L	e	L	e	e	e

(48) Learned ranking for German-style devoicing

{*d]$_\sigma$, *d]#} >> IO-Ident >> *VtV >> all other constraints

While the ranking bias for markedness over faithfulness in BCD has no small role in generating errors, the point to be emphasized is that there simply is no morphological analysis of this pattern, given the constraints available for ranking. This unambiguous nature of German-style devoicing stems from the lack of an overt contrast in the base. Because the simplex CVC bases do not have a voicing contrast, the only role for the OO-constraints is for a consistent change in the derived classes (e.g., insertion or deletion of [voice]), which is not observed. However, by switching the morphological relationships in the case above such that the base form contains the vowel-initial suffix, the system is rendered mp-ambiguous. In addition to the phonological analysis of automatic devoicing, there is a possible morphological analysis involving satisfaction of the OO-constraint OO¬Max, which requires devoicing of root-finals in all derived forms.

(49) German': automatic but mp-ambiguous

Base	*Forms derived from simplex base*	
___+*a*	___#	___+*ka*
pit-a	pit	pit-ka
pud-a	pu[t]	pu[t]-ka

The remarkable theoretical fact of this system is that it could be learned as either a phonological or morphological system, depending on hypothesized preferences for inserting markedness and OO-faithfulness constraints high in an interim grammar. Thus, if BCD actually distinguished the two, favoring markedness constraints over OO-faithfulness, then the learner would only have evidence like that in (50a), leading invariably to the phonological analysis in (51). If, on the other hand, BCD asserts a bias for OO-faithfulness over markedness, as argued for explicitly in (Tessier, 2007), then OO-Ident >> *VtV in some early stage of learning will result in errors like that in (50b), which in turn allows for a role for OO1¬Max in the final constraint ranking (52).

(50) Winner-loser data for automatic/ambiguous devoicing; (b) is optional

Input, Base	Winner	Loser	*d]$_\sigma$	*d]#	*VtV	IO-IDENT	IO¬IDENT	OO1-IDENT	OO2-IDENT	OO1¬MAX	OO2¬MAX
a. /pit-a/, NA	pit-a	pid-a	e	e	L	W	L	e	e	e	e
b. /pud-ka/, pud-a	put-ka	pud-ka	W	e	e	L	W	L	e	W	e

(51) Phonological approach to automatic/ambiguous devoicing

{*d]$_\sigma$, *d]#} >> IO-IDENT >> *VtV >> all other constraints

(52) Morphological approach to automatic/ambiguous devoicing

{OO1¬MAX} >> {IO-IDENT, *d]$_\sigma$, *d]#} >> *VtV >> all other constraints

That the analysis in (52) is in fact morphological, according to the definition in (4), is shown by the position of OO1¬MAX in the hierarchy, and that the phonological markedness constraints, *d]$_\sigma$ and *d]#, in (52) do not account for the winner-loser data in (50b). While it is not clear at this time why a learner would prefer the morphological analysis with a dominant OO-antifaithfulness over the phonological analysis, one possibility is that the imperative for OO-faithfulness constraints above markedness argued for in (Tessier, 2007) is in fact more general, and applies to all OO-correspondence constraints, including OO-antifaithfulness. Given such a bias, the learner is forced to posit the ranking in (52) over (51) in accounting for the winner-loser data. The interesting finding here is therefore that the preference for morphological over phonological analyses is independent of whether or not the overt data reflects an automatic alternation. The same question arises in systems with automatic alternations.

6.2 Across-the-board mutations

The analysis of morpho-phonological exchanges above has shown that it is so far harmless to include IO-antifaithfulness in the set of constraints a learner has at her disposal. The principles guiding learning, and the constraints themselves, preclude phonological exchanges. One type of phonological exchange not discussed yet, however, is a complete across-the-board exchange in which all surface forms reverse the lexical specification of some feature, as shown below for [voice]. As pointed out in section 2, this is an unambiguously phonological exchange, because the only ranking of constraints that can account for this collection of mappings is one in which IO-IDENT(VOI) is top-ranked. How can it be ruled out systematically?

(53) Exchange across-the board

Inputs	*Base*	*Forms derived from simplex base*		
	___#	___+*a*	___+*ka*	___+*kae*
/pit/	pi[d]	pi[d]-a	pi[d]-ka	pi[d]-kae
/pud/	pu[t]	pu[t]-a	pu[t]-ka	pu[t]-kae

One observation about this type of system is that it is completely ruled out by a different principle of phonotactic learning, identity mapping. The grammar of this system requires a lexicon with a specification of [voice] for the root-final that is the opposite of all of the surface forms. When surface forms do not alternate, however, identity mapping predicts that the surface forms will be used to build the lexicon, contrary to the assumptions needed for the above system. If the surface [voice] specifications are used as the guide for the lexicon, then IO-antifaithfulness will not favor any winner, and it will therefore be banished to the lower ranks of the constraint hierarchy.

What evidence, then, might exist for a role of IO-antifaithfulness in formal grammars? The theoretical motivation for including it is that it simplifies antifaithfulness theory by removing the *ad hoc* stipulation that antifaithfulness constraints are only defined on BR- and OO-correspondence relations. As for empirical evidence, one situation that might require a role for IO-antifaithfulness is if learners had evidence for a particular set of lexical representations, and an independent set of evidence for widespread reversal of the lexical representation. In other words, if the inputs were effectively learned 'ahead of time' or on the basis of separate evidence, then across-the-board reversals could activate IO-antifaithfulness. Evidence for this separation of learning the lexicon from learning the grammar could therefore support the contention that IO-antifaithfulness is part of the cognitive capacity of learners in special conditions.

While taking learning the lexicon out of the equation is somewhat artificial, this is essentially the nature of learning secret languages and ludlings. Secret languages are often based on the lexicon of a source language, so learners do not have to learn a new lexicon. Moreover, the overt data from the secret language provides independent evidence for mutations of source words. Indeed, it appears that antifaithfulness constraints provide a useful set of tools for characterizing the unfaithful mappings observed in secret language phonology: ¬MAX for segment or feature deleting operations, ¬DEP(FEATURE) for infixation or replacement of consonants or vowels with a fixed segment, ¬LINEARITY for metathesizing operations, etc. A secret language containing a segmental exchange rule, like the Dholuo voicing exchange, does not appear to be attested yet, but the surrogate drum-signaling language of the Jabo (Herzog, 1964) exhibits an exchange of two tonal types that accidently has the right characteristics. The actual facts do not show the intention to exchange tone on the part of the drummer; the tonal exchange appears to simply be an artifact of the medium used. Interestingly, however, this case shows that such an exchange is cognitively plausible, because listeners of the drum-signalling system were able to overcome the tonal reversals and interpret the mutated code, a point emphasized in (Wolfe, 1972).

However, to constitute evidence for input-output antifaithfulness requires that the antifaithfulness constraints responsible for secret language mutations work on non-surface representations. Otherwise, the relevant constraints could be formalized as OO-antifaithfulness, where the input is an actual surface word of the source language, and the output is the secret language form (see (Itô et al., 1996) for discussion of this problem for a Japanese argot). One finding relevant to this question is the need for a secret language module to be incorporated 'deeper' in the phonological system than just surface words. Thus, (Mohanan,

1982) argues that secret language forms are generated between the lexical and postlexical phonologies, and (Bagemihl, 1988) studies the derivational checkpoints for over fifty alternate phonological systems, concluding also that at least some secret language systems take the output of the lexical phonology as their input. An OT interpretation consistent with these conclusions would be to allow for a mapping from the output of stem level phonology of (Kiparsky, 1998/to appear) to a secret language module. Though such a mapping is not from an underlying representation, it is also crucially not from surface words, so it appears to be distinct from OO-antifaithfulness.

Another possible environment for across-the-board IO-antifaithfulness effects is sentence phonology. The domain of the sentence has long been construed as separate from the word domain. This separation is predicted in Lexical Phonology with the distinction between lexical and post-lexical levels of phonology ((Kiparsky, 1983), (Mohanan, 1982)), and also by the existence of post-syntactic models of phonological phrasing ((Selkirk, 1984), (Truckenbrodt, 1995)). If the separation of these two domains is such that the faithfulness properties of a word can be distinguished from its faithfulness properties in a larger sentence, then learners could effectively learn the inputs of words separately from the way those inputs are modified in larger sentences. Thus, if sentence-level phonology exhibits evidence for an across-the-board mutation, then learners could activate IO-antifaithfulness at this level, with the emergence of otherwise unattested phonological mappings. While space limitations do not allow a complete analysis, the separation of the evidence for the lexicon and evidence for a ranking appears to provide a fresh look at a classic problem in Chinese phonology, namely the circular chain-shifting phonology of Taiwanese ((Cheng, 1968), (Chen, 2000), (Moreton, 2004))

To sketch the pattern briefly, Taiwanese exhibits a shift of tone in all words of a sentence, except those words that appear at the end of a tone group, a phonological unit above the word equatable with the phonological phrase ((Lin, 1994), cf. (Chen, 1987)). The example below parses a sentence into tone groups, illustrating that the lexical tones do not change tone-group finally (N.C. = no change), but one step tone mutations occur everywhere else. Using the Chao tone transcription system, tones in syllables ending in a voiceless stop chain as follows: 54 → 32 → 54, and elsewhere: 23/55 → 33 → 31 → 53 → 55.

(54) Across-the-board tone mutation in Taiwanese (adapted from (Cheng, 1968))

Lexical tones:	ti33	tsin55	ku53 #	i53-	tsiŋ23 #	tai23-	lam23 #	si33	ts'it32	e23	hu-53	sia23 #
Tone groups:	{31	33	N.C. }{	55	N.C. }{	33	N.C. }{	53	54	33	55	N.C. }
	at	very	long	ago		Tainan		be	a	CL	prefecture capital	

‘Very long ago, Tainan was a prefectural capital’

The across-the-board nature of the tone circle provides precisely the kind of evidence that can activate IO-antifaithfulness, rather like the p-compatible analyses of the phonological elsewhere mutations discussed in section 5. The absence of any tonal changes in tone-group final position can be treated as a positional faithfulness effect (or a phonetically grounded constraint in longer syllables (Hsieh, 2005)). Elsewhere, the observed mutations can be modeled as the unbridled effect of IO-antifaithfulness constraints for tone pitch and register, an analysis that essentially combines the antifaithfulness constraints for tone pitch and tone register employed independently in (Mortensen, 2002) for two different languages, Jingpho (Tibeto-Burman) and A-Hmao (Western Hmongic). This appeal to IO-antifaithfulness side-steps a significant obstacle that arises in a transderivational antifaithfulness approach, namely the problem of identifying bases for words in larger sentences. Words are lexically specified in the input for tone, and IO-antifaithfulness requires a mutation in each non-phrase-final word. The specific details of the analysis require some additional assumptions, including local conjunction and the distinction between MAX and DEP for tone. But it is argued in (Alderete, 2008) that each of these assumptions are independently motivated, so the antifaithfulness analysis constitutes an attractive alternative to analyses that appeal to morphological suppletion and 'paradigmatic replacement' ((Tsay and Myers, 1996), (Moreton, 2004)), or contrast preservation analyses that require second order meta-constraints on entire phonological systems ((Yip, 2002), (Hsieh, 2005), (Barrie, 2006)).

7. Summary of results

The above discussion can be summarized as a set of principles that prevent the learning of phonological exchange rules in normal language learning. The results above therefore support the learnability filter hypothesis, as it is applied to exchange rules, but they also indicate a role for linguistic theory with the theory of well-formedness constraints.

(55) Principles precluding phonological exchanges

Identity mapping (external)	Prevents across-the-board phonological exchanges in normal learning in which lexicon learning occurs in tandem with learning rankings
Biased Constraint Demotion (external)	Identification of OO-antifaithfulness with markedness and IO-antifaithfulness with faithfulness allows BCD to favor morphological solutions in mp-ambiguous systems
Context-free nature of antifaithfulness constraints (internal)	Absence of phonological contexts in antifaithfulness constraints prevents constraints that specifically encourage exchanges in phonological environments
Constraint classes (internal)	Distinctions between markedness and faithfulness, and input-referring and non-input-referring constraints, are crucial to the formulation of learning biases

These results support two conclusions. First, the extent to which this combination of principles explains Anderson & Browne's generalization supports the underlying contention of the LF hypothesis, based on Boersma's (2003) original statement, namely that learnability considerations can in fact account for gaps in linguistic typology. These results therefore provide additional motivation for a need to consider the role of learnabilility in hypothesis testing tied to particular linguistic theories. Second, the need to appeal to both principles of learning, and aspects of the well-formedness constraints themselves, shows that explanation in linguistics may require examination of the interaction of these different domains. The solution presented here is not purely theory internal, nor external, but arrived at by a combination of both types of systems. The discussion below continues this argument by considering some alternative approaches that attempt to account for Anderson & Browne's generalization with purely theory-internal linguistic assumptions.

8. Implications and discussion of plausible alternatives

What are the implications of the above results for the linguistic analysis of morpho-phonological alternations? The discussion below addresses this question by comparing analyses of the voicing exchange in Dholuo (see section 2) that do not involve nonconservative constraints, analyses like the antifaithfulness approach above that do, and an analysis that employs *ad hoc* language particular constraints with context-sensitivity.

8.1 Are nonconservative constraints necessary?

There are a set of morpho-phonological alternations characterized by the CF argument (see section 2), including circular chain shifts, symmetric metathesis, unconditional augmentation, and other 'unweildy' alternations, which are not computable with conservative constraints alone. I will refer to them as IMECC alternations (for Inconsistent Mappings with Exclusive use of Conservative Constraints; see (Moreton, 2004) 151 ff. for formal characterizations). One approach to IMECC mappings is to try to account for putatively IMECC mappings with purely conservative constraints and analytical assumptions that make the CF argument irrelevant. The question of how to link IMECC-hood with morphologically defined environments is typically answered by proposing novel input representations for derived forms, as illustrated below with two clear examples of this approach.

In an effort to reduce alternations like the Dholuo voicing exchange to purely combinatorial processes, (Stonham, 1994) treats the Dholuo facts as inverse-marking by a featural affix, an autosegmental approach of some interest (see (Akinlabi, 1996), (Zoll, 1996), and the discussion immediately below). In Stonham's analysis, the direction of morphological mappings are arranged as shown below, where both characterized plurals and singulars can be derived through affixation of a floating [+voice] feature that instantiates inverse marking. The decision as to which singulars and plurals are uncharacterized, and therefore receive inverse-marking to derive number, is determined entirely by the observed alternations.

(56) Characterized number through inverse-marking (Stonham, 1994)

a. Uncharacterized singular		Characterized plural	
gɔt	→	gɔt + [+voice]$_{\text{Inver}}$ + e = gɔd-ɛ	'hills'
b. Uncharacterized plural		Characterized singular	
kit-e	→	kit + [+voice]$_{\text{Inver}}$ = kid	'stone'

The analysis is notable in that it accounts for the segmental exchange with purely additive morphology. The primary objection to this analysis, however, is that the assumed arrangement of the morphological relationships is arbitrary, since there is no independent evidence for this use of inverse-marking in Dholuo. Furthermore, the voicing exchange is observed in construct state forms, which cannot be straightforwardly accounted for with inverse-marking. Finally, it seems that such an analysis would require an additional assumption, namely a morpheme structure constraint prohibiting root-final voiced obstruents. Such a constraint would be necessary to preclude CVD-V$_{\text{SG}}$ ~ CVD-e$_{\text{PL}}$ paradigms, because characterized plurals /CVD + [+voice]$_{\text{Inver}}$ + e$_{\text{PL}}$/ would incorrectly come out as *CVD-e$_{\text{PL}}$, with redundant inverse marking. This leads to a duplication problem, because it duplicates the effects of an otherwise regular process of final devoicing.

A more recent analysis in (Wolf, 2005) combines featural affixation and suppletive allomorphy, which, together with new constraints on feature realization, correctly account for the Dholuo facts. In particular, two suppletive allomorphs, [+voice] and [-voice], are available to mark derived forms, and they are disjunctively specified in the input in a set of cohort affixes so that GEN can create input-output mappings for each allomorph. A constraint requiring the realization of floating features, MAXFLOAT, and a constraint requiring that this realization be nonvacuous, NOVACDOC, effectively produce both parts of the exchange. The tableau below illustrates the suppletive analysis for the *t* → *d* alternation, but the same ranking also predicts *d* → *t* in stems with final voiced obstruents.

(57) Dholuo voicing polarity as Suppletive Floaters

Root Plural suppletion set Input: baT[-voi]$_1$ + {[-voi]$_2$, [+voi]$_3$} *Inputs* *Outputs*	MAXFLOATER	NOVACDOC	IDENT(VOI)
a. /bat$_1$ +[-voi]$_2$/ bat$_1$	*!		
b. /bat$_1$ +[-voi]$_2$/ bat$_{12}$		*!	
c. /bat$_1$ +[+voi]$_3$/ bat$_1$	*!		
d. /bat$_1$ +[+voi]$_3$/ → bad$_3$			*

The suppletive floaters analysis is successful in ways that inverse-marking analysis is not, especially in its attempt to use independently motivated mechanisms of floating features and suppletion. The analysis is also notable in not requiring nonconservative constraints—the new constraints MAXFLOATER and NOVACDOC are arguably faithfulness constraints (because they require reference to inputs via correspondence relations). So, if suppletion of two opposite floating features can be shown to be limited to morphology (i.e., affixation), it will have successfully accounted for Anderson and Browne's generalization, as it relates to exchange rules at least.

There are several consequences of this approach for linguistic analysis that cannot be investigated here, but it can be clearly contrasted with other approaches that employ nonconservative constraints on learnability grounds. In particular, the problem of actually learning a system as complex as this, which involves projecting more than one floating feature in the input, and constructing a suppletion set, will be very difficult to implement in OT learning models for alternations. In particular, one of the fundamental assumptions in contrast analysis is that the effect of a morphological environment on a neighboring morpheme is uniform; otherwise, the learner cannot infer that surface contrasts are due to faithful contrasting features in the contrasting morphemes. But the inclusion of the two opposing feature specifications, fused in a disjunctive cohort set of affixes, makes this inference invalid.

Moreover, one of the most difficult problems in learning is positing covert linguistic structure from overt evidence ((Tesar, 2004), (Alderete and Tesar, 2002)). The suppletive floaters analysis requires two kinds of covert structure: multiple floating features and the creation of a suppletion set that establishes the needed disjunction among the floaters. It may be possible to develop algorithms for learning floating features and suppletive allomorphy, but the combined use of them is likely to significantly complicate learning morpho-phonological alternations in ways that are not desired. At the very least, these new algorithms will require a reconceptualization of current OT learning models, which is in contrast to the antifaithfulness analysis given above. (This point should not preclude, however, research investigating these new algorithms and learning models.) Considerations such as these speak in favor of a more simple linguistic analysis, and more powerful, possibily nonconservative, constraints, for motivating morpho-phonological alternations. Two basic ways of implementing such an approach are discussed below.

8.2 *Ad hoc* constraints

If nonconservative constraints are necessary for IMECC alternations, what is the nature of these constraints, and how is their application tied to morphologically defined environments? The approach proposed in sections 2-6 assumes that the nonconservative constraints are antifaithfulness constraints, and that the reason why they only motivate morpho-phonological alternations is because purely phonological circular chain shifts are unlearnable, given standard assumptions in OT learning models introduced above. The effects of nonconservative constraints can only arise, in normal language acquisition anyway, in the morphological domain.

One rather different approach, sketched but never explored in (Moreton, 2004), is that nonconservative constraints are *ad hoc* constraints that can specify language particular IMECC mappings. Linking nonconservativity and morphological environment could then come from the assumption that the *ad hoc* constraints are not UG constraints and are constructed in response to inconsistency. The winner-loser pairs from the IMECC alternations can be used to construct nonconservative constraints inductively, rather than simply activating innate antifaithfulness constraints, as proposed above in sections 2-6. Concretely, the winner-loser pairs (58a-b) below, perhaps after considerable attention on the part of the learner (for example, after being pulled from cache à la (Tessier, 2007)), can support the construction of a constraint like one below in the rightmost column. If the learner only has the first three conservative constraints, the winner-loser data is inconsistent; the creation of an alpha-switching constraint, tied to a particular environment, allows for a ranking consistent with the facts.

(58) Inconsistency detection leads to construction of language particular constraints

Input	Winner	Loser	*d	*VtV	IDENT(VOI)	αvoi→ -αvoi/__]-e]$_{Nplur}$
a. /pit+e $_{Npl}$/	pid-e	pit-e	L	W	L	W
b. /pud+e $_{Npl}$/	put-e	pud-e	W	L	L	W
c. /pud/	pud	put	e	e	W	e

Language particular constraints such as these are not unprecedented (see e.g., (Hayes, 1999)), and it may be possible to establish a mechanism that would avoid the existence of such constraints negating results in OT that derive from the assumption that all constraints are universal. For example, the formation of such constraints could be reserved for later stages of language development, in response to learning problems like the one illustrated above. Also, while Moreton (2004) does not propose a mechanism for predicting that *ad hoc* constraints are always linked to a particular morphological construction, such a mechanism would not be especially hard to formalize: the creation of an *ad hoc* constraint could be guided by morphological categories rather than phonological ones.

Without explicit formalization, it is impossible to directly compare this approach with the antifaithfulness approach in terms of performance in learning. At a higher level, however, the assumption that nonconservative constraints are not given *a priori* entails a rather different conception of the learning problem: the UG plus *ad hoc* constraints approach requires the learner to both learn the nonconservative constraints to be ranked, in tandem with learning the lexicon and the ranking of all constraints. In contrast, the constraints are universal in the analysis proposed here and so they do not have to be learned.

(59) Different conceptions of learning morpho-phonology

	UG Constraints Only	UG plus *Ad hoc* Constraints
Given:	UG constraints (all universal) Overt forms of the language	UG constraints (conservative) Overt forms of the language
Target function:	Lexicon + ordering of constraints	Lexicon + nonconservative *ad hoc* constraints + ordering of all constraints

Clearly the search space for the learner is much larger on the latter approach, but, as shown by recent research in phonological learning by Bruce Tesar and his colleagues, the size of the search space is much less important than the algorithms involved in navigating it. Sections 4-6 have illustrated how learners can grapple directly with winner-loser pairs to select consistent grammars with *a priori* antifaithfulness constraints and known learning algorithms. However, algorithms that can identify and isolate the relevant winner-loser data, and construct from it the relevant *ad hoc* constraints, are not as yet known. One idea that could be explored in this light is that antifaithfulness constraints are formulated inductively as *ad hoc* constraints in response to IMECC inconsistency. Indeed, the format of antifaithfulness constraints will reduce the learner's search space considerably, because they limit the target constraints to simple negations of existing faithfulness constraints. Further, they may also help in isolating the winner-loser pairs relevant for constructing the target constraint: as negations of faithfulness constraints, they can be fashioned to favor winners in all cases that the faithfulness constraint favors the loser, as shown in the inconsistent winner-loser pairs in (58a-b) above.

8.3 Nonhomogeneity and predicting nonconservativity

Another important issue for any approach that maintains the assumptions of the CF argument has to do with the role of nonhomogeneity in the analysis. Homogeneity, the assumption that inputs and outputs are made of the same representational elements, is not proposed in (Moreton, 2004) as an over-arching constraint on human language. Rather, it is an assumption necessary for proving the CF argument. Input-output mappings also include nonhomogeneous structure, like morphological labels in inputs, that is crucial to distinguishing conservative phonology and IMECC alternations. Conservative phonological processes can only improve on markedness, so, with some coherent theory of markedness, they define a set of natural phonological processes that are at the core of phonology. Nonconservative phonology may include natural phonological processes, but the existence of nonhomogeneous structure can 'let in' IMECC alternations, because they contain structure that conservative constraints can refer to, and, in some cases, render putatively IMECC alternations tractable with conservative constraints. This is the case in the suppletive floaters analysis sketched in 8.1, where the suppletion set {[-voice], [+voice]} is only specified in the input, and reference to this structure by conservative constraints produces the observed exchange. The conjecture that goes hand-in-hand with the CF argument is therefore that some morphological structure positively correlated with IMECC alternations is nonhomogeneous. However, it must be said that this assumption has yet to be established. The discussion below fleshes out some further issues that any analysis of IMECC alternations will have to address in order to maintain the assumptions of the CF argument.

Nonhomogeneous structure is present only at one level of representation, i.e., input structure that is erased in outputs, or output structure that is inserted by GEN. The structures below are postulated to be nonhomogeneous in (Moreton, 2004).

(60) Nonhomogeneous structure (Moreton, p. 147)

a. Input, not Output	b. Output, not Input
Phonologically empty morphemes (e.g., RED morpheme)	Phonological phrase boundaries
Morphological constituency (e.g., root vs. affix for faith domains)	
Syntactic constituency (for phrasal phonology)	
Part of speech labels, morphological features	
Diacritic features (e.g., level 1 vs. level 2)	

If some structure is nonhomogeneous, it is pointless to assess faithfulness for that structure, since it cannot be satisfied as a matter of principle. Furthermore, GEN can never change nonhomogeneous structure from one level to the next. It can only insert nonhomogenous output structure or delete input structure. The explanation of Anderson and Browne's generalization in the CF argument therefore depends on the contention that the morphological attributes that accompany IMECC alternations are indeed nonhomogeneous.

One issue directly relevant to nonhomogeneous morphological structure is the nature of morphology-phonology interactionism. One of the most basic results of Lexical Phonology and Morphology is the necessity to compute morphological processes in tandem with phonological processes (see (Hargus, 1993) for a useful summary). It is often the case that morphological generalizations depend crucially on derived phonological structure, like prosodically governed suppletion. In Optimality Theory, many of these problems have been taken as evidence for the simultaneity of morphology and phonology. For example, (Kager, 1997a) analyzes the syllable-based generalizations governing affix allomorphy in Estonian as the direct interaction of constraints on prosodic well-formedness and the alignment of morphological stem structure with prosody. Another, rather dramatic, example is the case of prosodic closure of morphological units in Australian languages like Diyari and Dyirbal ((McCarthy and Prince, 1994), (Kager, 1997b)). The correct foot parse in these cases depends crucially on the position of root, stem, and affix edges, a kind of interactionism that also requires simultaneous reference to morphological constituents and prosodic structure in the output. An analysis of these cases becomes terribly difficult if morphological constituent structure, and, in many cases, more specific morphological features (e.g., [+plural], [-animate], 3rd conjugation), are erased in output candidates, which suggests that they are in fact homogeneous structures.

It is also the case that certain output structures classified by Moreton as nonhomogeneous are relevant for inputs, and presumably specified in inputs too. Aoyagi prefixes in Japanese (Poser, 1990) are a striking case because they are affixes that cause the following stem to begin a minor phrase (i.e., a phonological phrase or higher), yet this information must be specified in the lexicon in order to distinguish Aoyagi prefixes from other prefixes. In other words, it seems that information about output phonological phrases must be specified in inputs.

Finally, it seems that at least some putatively nonhomogeneous structure can be changed by GEN. One type of analysis that supports this contention is the assumption that morphemes can in effect change affiliation from one morphological category to another. For example, to account for certain differences in the application of cooccurrence restrictions in Akkadian and Ponapean, (Lubowicz, to appear) assumes that certain affixes can be 'absorbed' by a root as infixes and therefore be subject to different phonotactic requirements. Another similar idea follows from the superimposition of correspondence relations argued for in (Benua, 2000), and also (Revithiadou, 1999) in a different context, where the computation of which correspondence relation to use in assessing faithfulness, when more than one is available, is determined by GEN. If correspondence relations and morphological category status can be changed in these ways, then it is hard to defend the assumption that they are only present at one level of analysis.[5]

It has not been shown here that all of these apparent exceptions to Moreton's classifications cannot be understood as nonhomogeneous structure. For example, perhaps Aoyagi prefixes in Japanese are best understood as having subcategorization requirements that indirectly entail minor phrases in outputs but do not directly represent them. However, the larger picture emerging from this discussion is that it is not at all clear that the morphological structures that positively correlate with IMECC alternations in the CF argument are in fact nonhomogeneous. When compared to the explanation of Anderson and Browne's generalization under the LF hypothesis, these potential problems are completely side-stepped. If the morphological nature of IMECC alternations follows from principles of learning, the (non)homogeneity of linguistic elements is irrelevant to the explanation and linguists are free to play with representational assumptions about both phonological and morphological structure. This analytical freedom is also retained if language particular antifaithfulness constraints are learned by induction, as conjectured in 8.2.

8.4 The role of Universal Grammar

One of the key questions in the analysis of morpho-phonological alternations, including the analysis of IMECC alternations like exchange rules, is just how predictive linguistic theories are. Discussions in contemporary morphology, with specific reference to circular chain shifts, go back and forth on the need for so-called process-based morphology (PBM), i.e., morphology that is marked in a sense by phonological processes ((Anderson, 1992), (Spencer, 1998), (Stiebels and Wunderlich, 2000), (Stonham, 1994)). This debate rages on in contemporary phonology too, with serious questions about the unrestricted use of nonconservative constraints in OT (Prince, 2007). As is common, the discussions typically revolve around finding a balance between descriptive adequacy and theoretical restrictiveness.

The debate as it applies to antifaithfulness is by no means settled, but the present work allows one to see more clearly the benefits of employing universal constraints in the analysis of IMECC alternations. If it is true, as demonstrated in sections 4-6 for controlled linguistic typologies, that the effects of IO-antifaithfulness can be attenuated in learning, then this theory of morpho-phonology successfully addresses at least some concerns about the apparent unrestricted nature of the theory. Furthermore, the problem of learning non-automatic IMECC alternations using antifaithfulness has been shown to be very tractable in contemporary learning models, something that distinguishes its analysis from plausible alternatives. Finally, because of its commitment to universal constraints, a theory of constraints with antifaithfulness makes predictions that are not possible in other theories.

[5] Though it must be said that the standard definition of a candidate in OT entails that correspondence relations are nonhomogeneous. If a candidate is constituted by an input, an output, and a correspondence relation that indicates corresponding elements, then the latter cannot be part of the output form.

One way of clarifying the role of universal constraints is by examining the inventory of morpho-phonological operations predicted by theories of PBM. One of the common objections to so-called item-and-process theories of PBM is that there is no analysis of the range of possible mappings that PBM can allow. In part in response to this, (Zwicky, 1988) proposes that UG somehow define an inventory of possible morphological operations, e.g., affixation, segmental mutations, subtraction, metathesis, feature shift, etc., in essence, an inventory of morphological processes akin to the list of natural phonological processes enumerated in Natural Generative Phonology (Stampe, 1979). In OT, natural phonological processes are conservative phonology, predicted essentially by markedness-faithfulness interactions. Antifaithfulness theory proposes that the set of mappings in PBM, an entirely different beast, is predicted through the negation of faithfulness constraints. Thus, the faithfulness/antifaithfulness homology predicts a very specific set of processes, rather close to Zwicky's original vision for PBM.

(61) Faithfulness/antifaithfulness homology

There is a faithfulness constraint for every antifaithfulness constraint, and vice versa.

(62) Antifaithfulness effects

¬DEP(X)	Obligatory insertion of features for morphological reasons: Athabaskan valence morphology (Shaw, 1991); insertion of [labial] to distinguish verbs with certain objects from other verbs in Chaha (Rose, 1997)
¬MAX(X)	Obligatory deletion of a segment or feature: Subtractive morphology (Martin, 1988); accent-deleting affixes, AKA dominance effects (Alderete 2001a); pre-shortening suffixes in Slovak (Rubach, 1993)
¬NOSHIFT/ SPREAD(X)	Obligatory migration of feature: Tone shift with masculine and diminutive suffixes in Iñapari (Parker 1999); [+nasal] spread as a marker of first singular in Terena (Bendor-Samuel 1960); bi-directional stress shift for marking inflectional distinctions in Russian nouns (Spencer, 1998)
¬LINEARITY	Obligatory metathesis: Metathesis as a marker of 'actual' in Saanich verbs (Montler 1986); vowel metathesis in verbs and definite adjectives in Latvian (Halle 1987)
¬IDENT(F)	Obligatory exchange/rotation of feature: Length inversion as a marker of plural verbs in Diegueño (Walker 1970, Langdon 1970); voicing exchange in plural nouns in Luo (Gregerson 1972, Okoth-Okombo 1982); chaining tone sandhi processes (Cheng, 1968)

There are other, perhaps less accepted, faithfulness constraints that also predict morpho-phonological alternations, like infixation from the negation of CONTIGUITY constraints, but the predicted antifaithfulness effects above actually have a nice goodness of fit with the traditional domain of nonconcatenative morphology, excluding perhaps reduplication. Furthermore, carefully argued analyses have been worked out for virtually all of these process types, including (Bat-El, 2002) and (Horwood, 2001) on subtractive morphology, (Alderete, 2001b) and (Frazier, 2006) on feature deletion in accentual systems, (Alderete, 2001a) for segmental exchanges, (Mortensen, 2002), (Lubowicz, 2003), and (Alderete, 2008) on more complex chains, and (Alderete, 2001a) and (Crosswhite et al., 2003) for accentual shifts and spreading of tone. Alternatives to these analyses have been investigated, as for example (Oostendorp, 2005) on the Limburg Dutch tone mutation, but there has yet to be such a clear characterization of the range of operations available in PBM as that predicted by the faithfulness/antifaithfulness homology. Thus, the implication of universal constraints for linguistic analysis is clear: it provides both the descriptive capacity for well-attested morpho-phonological alternations, as well as an explicit characterization of the range of possible processes in PBM.

The differences between natural phonological processes, on the one hand, and process-based morphology on the other, also has a natural explanation in learnability theory. The absence of phonological exchange rules is not due to formal limits imposed on constraints (contra (Alderete, 2001a)). It is the result of the assumptions of linguistic theory and how those assumptions are grappled with in learning. The generality of the assumptions appealed to in the learning model, including Biased Constraint Demotion and identity mapping, supports a general explanation for at least one difference between the two possible inventories.

9. Conclusion

The learnability filter hypothesis proposes to limit the range of possible morpho-phonological alternations via principles of language learning that work in tandem with the theory of constraints in Optimality Theory. Universal conservative and nonconservative constraints are allowed to freely interact, and as a result, linguistic theory proper may admit certain types of alternations that are not in fact attested in synchronic grammars. The above results show, however, that the formal mechanisms for learning language effectively eliminate some of these alternations from the set of learnable systems. Extending Boersma's (2003) original observation, restrictiveness is not just a consequence of linguistic theory, therefore, but may also come from external factors like the mechanisms of normal language learning. These findings for segmental exchange rules thus fit within larger theories of morphology and phonology that specifically espouse more powerful linguistic theories, and, at the same time, recognize external factors that combine with internal factors to explain restrictiveness (Anderson, 1992). Finally, while this investigation has been focused on filtering out some of the consequences of antifaithfulness constraints, the LF hypothesis may be investigated in other domains to identify other mismatches between the predictions of linguistic theory and the domain of the learnable.

Appendix

The IDENT constraints are based on the constraint formalizations in (McCarthy and Prince, 1995), with the positional restrictions from (Lombardi, 1999) and (Beckman, 1999) and applications to output-to-output correspondence from (Benua, 2000). MAX(VOI) is from (Lombardi, 2001), with the addition that it is defined on OO-correspondence. The antifaithfulness constraints are simply wide-scope negations of these constraints (Alderete, 2001a), given here in nontechnical terms.

Markedness

*d]$_\sigma$: no voiced stops syllable-finally

*d#: no voiced stops word-finally

*VtV: no intervocalic voiceless stops

Input-output faithfulness

IO-IDENT(VOI): corresponding segments agree in the feature [voice]

IO-IDENTSTRSYLL(VOI): for corresponding segments, S_I and S_O, where S_O appears in a stressed syllable in the output, S_I and S_O agree in [voice]

IO-IDENTONSET(VOI): for corresponding segments, S_I and S_O, where S_O appears in an onset in the output, S_I and S_O agree in [voice]

Input-output antifaithfulness

IO¬IDENT(VOI): at least one pair of corresponding segments must not agree in the feature [voice]

Output-output faithfulness

OO-MAX(VOI): an autosegment [voice] in the base ($output_1$) has a corresponding [voice] in the output ($output_2$)

OO-IDENT(VOI): corresponding segments in base ($output_1$) and output ($output_2$) must agree in [voice]

Output-output antifaithfulness

OO¬MAX(VOI): it is not the case that every autosegment [voice] in the base ($output_1$) has a correspondent in the output ($output_2$)

OO¬IDENT(VOI): at least one pair of corresponding segments in base ($output_1$) and output ($output_2$) must not agree in [voice]

10. References

Akinlabi, Akinbiyi. 1996. Featural affixation. *Journal of Linguistics* 32:239-289.

Alderete, John, Beckman, Jill, Benua, Laura, Gnanadesikan, Amalia, McCarthy, John, and Urbanczyk, Suzanne. 1999. Reduplication with fixed segmentism. *Linguistic Inquiry* 30:327-364.

Alderete, John. 2001a. *Morphologically governed accent in Optimality Theory*: Outstanding dissertations in linguistics. New York: Routledge.

Alderete, John. 2001b. Dominance effects as Transderivational Anti-Faithfulness. *Phonology* 18:201-253.

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 no. RuCCS-TR-72. Piscataway, NJ: Rutgers Center for Cognitive Science.

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 24th West Coast Conference on Formal Linguistics*, eds. John Alderete, Chung-hye Han and Alexei Kochetov, 34-42. Somerville, MA: Cascadilla Proceedings Project.

Alderete, John. 2008. Antifaithfulness and local conjunction: an analysis of the Taiwanese tone sandhi without paradigmatic replacement or linguistic systems as candidates. Ms. Simon Fraser University.

Anderson, Stephen R., and Browne, Wayles. 1973. On keeping exchange rules in Czech. *Papers in Linguistics* VI:445-482.

Anderson, Stephen R. 1992. *A-morphous Morphology*. Cambridge: Cambridge University Press.

Bagemihl, Bruce. 1988. Alternate phonologies and morphologies, University of British Columbia: Doctoral dissertation.

Barrie, Michael. 2006. Tone circles and contrast preservation. *Linguistic Inquiry* 37:131-141.

Bat-El, Outi. 2002. True truncation in colloquial Hebrew imperatives. *Language* 78:651-683.

Beckman, Jill N. 1999. *Positional faithfulness: An optimality theoretic treatment of phonological asymmetries*: Outstanding dissertations in linguistics. New York: Garland.

Benua, Laura. 2000. *Phonological relations between words*: Outstanding dissertations in linguistics. New York: Garland.

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

Boersma, Paul. 2003. Review of Bruce Tesar and Paul Smolensky 2000, Learnability in Optimality Theory. *Phonology* 20:436-446.

Bybee Hooper, Joan L. 1979. Child morphology and morphophonemic change. *Linguistics* 17:21-50.

Bybee, Joan L. 1988. Morphology as lexical organization. In *Theoretical morphology: approaches in modern linguistics*, eds. Michael Hammond and Michael Noonan, 119-141. Orlando: Academic Press.

Chen, Matthew. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4:109-149.

Chen, Matthew. 2000. *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Cheng, Robert L. 1968. Tone sandhi in Taiwanese. *Linguistics* 41:19-42.

Chomsky, Noam, and Halle, Morris. 1968. *The sound pattern of English*. New York: Harper & Row.

Crosswhite, Catherine, Alderete, John, Beasley, Tim, and Markman, Vita. 2003. Morphological effects on default stress placement in novel Russian words. In *Proceedings of the West Coast Conference on Formal Linguistics 22*, eds. Gina Garding and Mimu Tsujimura, 151-164. Somerville, MA: Cascadilla Press.

Frazier, Melissa. 2006. Accent in Proto-Indo-European athematic nouns: Antifaithfulness in inflectional paradigms, University of North Carolina: MA thesis.

Gregersen, Edgar A. 1972. Consonant polarity in Nilotic. In *Proceedings of the third annual conference on African linguistics*, ed. Erhard Voeltz, 105-109. Bloomington: Indiana University.

Hargus, Sharon. 1993. Modeling the phonology-morphology interface. In *Phonetics and phonology, Volume 4: Studies in Lexical Phonology*, eds. Sharon Hargus and Ellen Kaisse, 45–74. San Diego: Academic Press.

Hayes, Bruce, Tesar, Bruce, and Zuraw, Kie. 2003. OTSoft 2.1, software package. UCLA and Rutgers University.

Hayes, Bruce. 2004. Phonological acquisition in Optimality Theory. In *Fixing priorities: Constraints in phonological acquisition*, eds. René Kager, Joe Pater and Wim Zonneveld, 158-203. Cambridge: Cambridge University Press.

Hayes, Bruce P. 1999. Phonological restructuring in Yidin and its theoretical consequences. In *The derivational residue in phonological Optimality Theory*, eds. Ben Hermans and Marc van Oostendorp, 175-205. Amsterdam: John Benjamins.

Herzog, George. 1964. Drum signaling in a West African tribe. In *Language in culture and society*, ed. Dell Hymes, 312-329. New York: Harper & Row, Publishers.

Horwood, Graham. 2001. Anti-faithfulness and subtractive morphology. Ms. Rutgers University.

Hsieh, Feng-fan. 2005. Tonal chain-shifts as anti-neutralization-induced tone sandhi. In *Proceedings of the 28th Annual Penn Linguistics Colloquium, University of Pennsylvania Working Papers in Linguistics 11.1*, eds. Sudha Arunachalam, Tatjana Scheffler, Sandhya Sundaresan and Joshua Tauberer, 99–112. Philadelphia: University of Pennsylvania.

Itô, Junko, Kitagawa, Yoshihisa, and Mester, Armin. 1996. Prosodic faithfulness and correspondence: Evidence from a Japanese argot. *Journal of East Asian Linguistics* 5:217-294.

Itô, Junko, and Mester, Armin. 1999. The phonological lexicon. In *The handbook of Japanese linguistics*, ed. Natsuko Tsujimura, 62-100. Malden, MA: Blackwell.

Kager, René. 1997a. On affix allomorphy and syllable counting. In *Interfaces in Phonology. Studia Grammatica 41*, ed. Ursula Kleinhenz, 155-171. Berlin: Akademie Verlag.

Kager, René. 1997b. Generalized alignment and morphological parsing. *Rivista di Linguistica* 9:245-282.

Kiparsky, Paul. 1983. Lexical phonology and morphology. In *Linguistics in the morning calm*, ed. I. S Yang, 3-91. Seoul: Hansin.

Kiparsky, Paul. 1998/to appear. *Paradigm effects and opacity*: CSLI, Stanford University.

Lin, Jo-Wang. 1994. Lexical government and tone-group formation in Xiamen Chinese. *Phonology* 11:237-275.

Lombardi, Linda. 1999. Positional Faithfulness and voicing assimilation in Optimality Theory. *Natural Language and Linguistic Theory* 17:267-302.

Lombardi, Linda. 2001. Why place and voice are different: Constraint-specific alternations in Optimality Theory. In *Segmental phonology in Optimality Theory: Constraints and representations*, ed. Linda Lombardi, 13-45. Cambridge: Cambridge University Press.

Lubowicz, Ania. 2003. Local Conjunction and Comparative Markedness. *Theoretical Linguistics* 29:101-112.

Lubowicz, Ania. to appear. Infixation as morpheme absorption. In *Phonological argumentation: Essays on evidence and motivation*, ed. Steve Parker. London: Equinox.

Martin, Jack. 1988. Subtractive morphology as dissociation. In *Proceedings of the Seventh West Coast Conference on Formal Linguistics*, 229-240.

McCarthy, John J., and Prince, Alan. 1994. Two lectures on Prosodic Morphology. Ms., *Rutgers Optimality Archive, 59*. University of Massachusetts, Amherst and Rutgers University.

McCarthy, John J., and Prince, Alan. 1995. Faithfulness and reduplicative identity. In *University of Massachusetts Occasional Papers 18, Papers in Optimality Theory*, eds. Jill Beckman, Suzanne Urbanczyk and Laura Walsh, 249-384. Amherst, MA: Graduate Linguistic Student Association.

McCarthy, John J. 1998. Morpheme structure constraints and paradigm occultation. In *Proceedings of the Chicago Linguistic Society 5, Vol. II: The Panels*, eds. M. Catherine Gruber, Derrick Higgins, Kenneth Olson and Tamra Wysocki, 123–150. Chicago: Chicago Linguistics Society.

Merchant, Nazarré, and Tesar, Bruce. 2006. Learning underlying forms by searching restricted lexical subspaces. In *Proceedings of the 41st Chicago Linguistics Society*.

Mohanan, K. P. 1982. Lexical Phonology, MIT: Doctoral dissertation.

Moreton, Elliott. 2004. Non-computable functions in Optimality Theory. In *Optimality Theory in Phonology*, ed. John J McCarthy, 141-163. Malden, MA: Blackwell Publishing.

Mortensen, David. 2002. Semper Infidelis: Theoretical dimensions of tone sandhi chains in Jingpho and A–Hmao. Ms. University of California, Berkeley.

Okoth-Okombo, Duncan. 1982. *Dholuo morphophonemics in a generative framework*. Berlin: Dietrich Reimer Verlag.
Oostendorp, Marc van. 2005. Expressing inflection tonally: Meertens Instituut.
Pater, Joe. 2005. Learning a stratified grammar. In *Proceedings of the 29th Boston University Conference on Language Development*, eds. Alejna Brugos, Manuella R. Clark-Cotton and Seungwan Ha, 482-492. Somerville, MA: Cascadilla Press.
Poser, William. 1990. Word-internal phrase boundary in Japanese. In *The Phonology-Syntax Connection*, eds. Sharon Inkelas and Draga Zec, 279-287. Chicago: University of Chicago Press.
Prince, Alan, and Smolensky, Paul. 1993/2004. *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell.
Prince, Alan. 2003. Arguing optimality. In *Papers in Optimality Theory II*, eds. Andries Coetzee, Angela Carpenter and Paul de Lacy, 269-304. Amherst, MA: GLSA.
Prince, Alan, and Tesar, Bruce. 2004. Learning phonotactic distributions. In *Fixing priorities: Constraints in phonological acquisition*, eds. René Kager and Joe Pater, 245-291. Cambridge: Cambridge University Press.
Prince, Alan. 2007. The pursuit of theory. In *The Cambridge handbook of phonology*, ed. Paul de Lacy, 33-60: Cambridge University Press.
Revithiadou, Anthi. 1999. Headmost accent wins: Head dominance and ideal prosodic form in lexical accent systems, Leiden University: Doctoral dissertation.
Rose, Sharon. 1997. Theoretical issues in Comparative Ethio-Semitic phonology and morphology, McGill University: Doctoral dissertation.
Rubach, Jerzy. 1993. *The lexical phonology of Slovak*. Oxford: Oxford University Press.
Selkirk, Elisabeth. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Shaw, Patricia. 1991. Consonant harmony systems: The special status of coronal harmony. In *The special status of coronals*, eds. C. Paradis and J. F. Prunet, 125-157. New York: Academic Press.
Smolensky, Paul. 2006. On theoretical facts and empirical abstractions. In *Wondering at the Natural Fecundity of Things: Essays in Honor of Alan Prince*, eds. Eric Baković, Junko Itô and John J. McCarthy. Santa Cruz, CA: Linguistics Research Center.
Spencer, Andrew. 1998. Morphophological operations. In *The Handbook of Morphology*, eds. Andrew Spencer and Arnold M. Zwicky, 123-143. Oxford: Blackwell Publishers.
Stampe, David. 1979. *A dissertation on Natural Phonology*: Outstanding dissertations in linguistics. New York: Garland.
Stiebels, Barbara, and Wunderlich, Dieter eds. 2000. *Lexicon in focus*. Berlin: Akademie Verlag.
Stonham, John. 1994. *Combinatorial Morphology*. Amsterdam: John Benjamins.
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 22nd West Coast Conference on Formal Linguistics*, eds. G. Gardin and M. Tsujimura, 477-490. Somerville, MA: Cascadilla Press.
Tesar, Bruce. 2004. Using inconsistency detection to overcome structural ambiguity in language learning. *Linguistic Inquiry* 35:219-253.
Tesar, Bruce. 2006. Faithful contrastive features in learning. *Cognitive Science* 30:863-903.
Tessier, Anne-Michelle. 2007. Biases and stages in phonological acquisition, University of Massachusetts, Amherst: Doctoral dissertation.
Truckenbrodt, Hubert. 1995. Phonological phrases: Their relation to syntax, focus, and prominence, MIT: Doctoral dissertation.
Tsay, Jane, and Myers, James. 1996. Taiwanese tone sandhi as allomorph selection. In *Proceedings of the Twenty-Second Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on the Role of Learnability in Grammatical Theory*, eds. Jan Johnson and Matthew L. Juge, 394-405. Berkeley, CA: Berkeley Linguistics Society.

Tucker, Archibald N, and Creider, Chet A. (ed. by). 1994. *A grammar of Kenya Luo (Dholuo)*. Köln: Rüdiger Köppe Verlag.
Wolf, Matthew. 2005. For an autosegmental theory of mutation. In *University of Massachusetts Occasional Papers in Linguistics 32: Papers in Optimality Theory III*, eds. Leah Bateman, Michael O'Keefe, Ehren Reilly and Adam Werle, 315-404. Amherst: GLSA.
Wolfe, Patricia M. 1972. *Linguistic change and the great vowel shift in English*. Berkeley: University of California Press.
Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.
Zoll, Cheryl. 1996. *Parsing below the segment in a constraint based framework*: PhD Dissertation, University of California, Berkeley.
Zwicky, Arnold. 1988. Morphological rules, operations and operation types. In *Proceedings from the Fourth Eastern States Conference on Linguistics*, eds. Ann Miller and Joyce Powers, 318-334. Columbus: The Ohio State University.