Lexical insertion occurs in the phonological component*

To appear in Bernard Tranel (ed.), *Understanding Allomorphy: Perspectives from Optimality Theory*. London: Equinox.

Matthew Wolf
University of Massachusetts Amherst
March 24, 2007

1. Introduction

Mester (1994), Kager (1996), Mascaró (1996), and Tranel (1996a,b) propose that phonologically-conditioned suppletive allomorphy is an instance of the Emergence of the Unmarked (or TETU: McCarthy & Prince 1994). If a morpheme has two listed alternants X and Y, which respectively appear in the environments A_B and C_D, a TETU analysis requires two things to hold:

(1)
a. Some (markedness) constraint M_1 that exerts the preference AXB ⟩ AYB must dominate all constraints that exert the preference Y ⟩ X.

b. Some other (markedness) constraint M_2 that exerts the preference CYD ⟩ CXD must dominate all constraints that exert the preference X ⟩ Y.

For an example, consider the distribution of the 3rd person masculine singular pronominal enclitic in Moroccan Arabic, which is [u] after a consonant-final stem and [h] after a vowel-final stem. Because [u] and [h] bear no phonological resemblance to one another, it is clear that they need to be analyzed as separate, lexically-listed items, and not as different realizations of the same underlying form. Mascaró's analysis assumes that the 3rd-person-masculine-singular morpheme has two underlying forms: {/u/, /h/}. Each candidate produced by GEN bears a correspondence relation to exactly one of these underlying forms[1]. Hence, each candidate is pressured to be faithful to only one of the underlying forms. Mascaró (1996) analyzes this system of allomorphy in TETU terms by assuming the ranking ONSET » NOCODA. When the stem ends in a vowel, a candidate that picks /h/ thus beats candidates that pick /u/, because having an onsetless syllable is worse than having a word-final coda:

* Thanks to Michael Becker, Paul de Lacy, Kathryn Flack, Karen Jesney, Shigeto Kawahara, John Kingston, John McCarthy, Joe Pater, Alan Prince, Kathryn Pruitt, Bruce Tesar, Anne-Michelle Tessier, and the audience at HUMDRUM 2006 for helpful feedback related to this material. All errors are mine.

[1] The requirement that each candidate picks exactly one, and never zero or more than one, of the competing underlying forms, is effectively a stipulation; as we will see in §3, the proposal advanced in this paper is able to derive this economy of allomorphs without recourse to stipulation.

(2) *Moroccan Arabic: 'his error'*[2]

/xtˤa – {h, u}/ *Inputs:*	*Outputs:*	ONSET	NOCODA
/xtˤa-h/	→ [xtˤah]		1
/xtˤa-u/	~ a. [xtˤa.u]	W 1	L

By contrast, when the stem ends in a consonant, choosing /u/ will not result in an onsetless syllable. ONSET is therefore indifferent to the choice of allomorphs in the context of a C-final stem, and NOCODA will exert the decisive preference in favor of a /u/-selecting candidate over an /h/-selecting one:

(3) *Moroccan Arabic: 'his book'*

/ktab – {h, u}/ *Inputs:*	*Outputs:*	ONSET	NOCODA
/ktab-u/	→ [kta.bu]		
/ktab-h/	~ a. [ktabh]		L_1

In terms of the schema given in (1), ONSET is the constraint M_1, which prefers /h/ over /u/ in the context of a V-final stem. This constraint is ranked above the constraint NOCODA (M_2), which will prefer /u/ over /h/, regardless of whether the stem is C-final or V-final.

Because there are many dimensions of markedness, which often conflict with one another, it is often possible to identify two markedness constraints that could serve as M_1 and M_2, and to rank them as required. There are, however, systems of allomorphy where identifying such a pair of constraints would clearly be difficult if not impossible. The most widely-discussed example of this sort comes from the allomorphy of the Ergative suffix in Dyirbal (Dixon 1972, McCarthy & Prince 1990, 1993a: ch. 7, Bonet 2004, Paster 2005, Bye 2006). In this language, the Ergative is marked by [-ŋku] on disyllabic vowel-final noun stems, and [-ku] on longer vowel-final stems:[3]

(4) *Dyirbal Ergative*[4]

yaɽa-ŋku	'man-ERG'
yamani-ku	'rainbow-ERG'
palakara-ku	'they-ERG'

[2] I use comparative tableaux throughout; see Prince (2003) on this format and the motivations for it.

[3] The analysis in this paper will consider only the allomorphs that appear with V-final stems. Stems ending in a nasal or [j] mark the Ergative with a stop homorganic to the stem-final consonant, followed by [u]; when the stem ends in [l], [r], or [ɽ], the Ergative is marked by deletion of that consonant, plus suffixation of [-ɽu].

[4] A note on orthography: Dyirbal has only a single (voiceless unaspirated) stop series; for the sake of minimizing confusion, I use IPA [p t k …] for Dyirbal stops rather than the [b d g …] orthography employed by Dixon (1972).

In order to analyze the Dyirbal Ergative as TETU, we would have to find some constraint that preferred [-ŋku] over [-ku], either in general or just in the context of being attached to a disyllabic stem. Given the great similarity of [-ŋku]'s and [-ku]'s phonological shapes—as well as the marked status of nasality, of consonant clusters, and of N̥C sequences—it is far from clear that such a markedness constraint could be found.

Problems of this sort have produced two types of theoretical response in OT. One is to assume that allomorphs like [-ŋku], rather than being preferred over competing allomorphs by any markedness constraint, are the beneficiaries of an arbitrary preference, with that preference enforced in the phonology in various ways (McCarthy & Prince 1993a: Ch. 7, Kager 1996, Picanço 2002, Bonet 2004, Bonet, Lloret & Mascaró 2005, Mascaró 2005, Kikuchi 2006). The other is to take cases of allomorphy like (4) as evidence that allomorph selection takes place outside of the phonology proper, via a pre-EVAL filter on GEN (Lapointe 1999) or a module of inviolable constraints that filters the output of EVAL (Bye 2006).[5]

This paper will argue for the first view, with allomorph selection in the phonology and arbitrary preference called upon for systems like the Dyirbal Ergative. The distribution of listed allomorphs in systems like this can be described in phonological terms, and parsimony would lead us to assume as the null hypothesis that all natural-language phenomena that can be so described are part of the phonological module of the grammar. In order to maintain the null hypothesis, we will need some means for an arbitrary preference among allomorphs to be exerted within the phonological grammar. However, as I will argue, all existing proposals about how to implement such a preference suffer from significant problems.

I therefore propose a new means of exerting arbitrary preferences among allomorphs. The idea is this: the input to the phonology consists not of pre-selected morphemes, but rather of linearized bundles of morpho-syntactic features, and the phonology is responsible for choosing which morphemes are to be associated with each feature-bundle. Individual lexical entries are specified as spelling out one or more morphosyntactic features, and there are constraints which favor spelling out as many features as possible.[6] These constraints are freely re-rankable with phonological markedness and faithfulness constraints, and so the pressure to spell out as many features as possible can be overridden on phonological grounds.[7] In this system, arbitrary

[5] A third idea, advanced in Klein (2003), is that phonologically perverse allomorph distribution be explained by assuming that morphemes include markedness violations in their lexical entries, as in Golston (1996), and that those violations have to be preserved in the output. This proposal surely overgenerates: for instance if a morpheme /ata/ were lexically specified with a violation of NOCODA, it would be syllabified [at.a], while a homophonous morpheme without the lexical violation would syllabify as [a.ta]. As contrastive syllabification is generally held to be impossible (Clements 1986, Hayes 1989, Blevins 1995, McCarthy 2003b, cf. Elfner 2006), lexical encoding of constraint violation is probably best eschewed.

[6] The idea that the input to the phonology contains not-yet-spelled-out morphosyntactic features also seems to be implicit in Kager (1996) and Yip (1998).

[7] As this paper was nearing completion I became aware of Teeple (2006), which argues for this same general point—that constraints governing lexical selection freely interact with those on phonology—but within a rather different framework, in which the construction of syntactic and morphological structure

preferences among allomorphs such as [-ŋku] over [-ku] in Dyirbal can be attributed to the assumption that [-ŋku] spells out a proper superset of the morphosyntactic features that [-ku] does.

The remainder of this paper is organized as follows: Section 2 shows the need for arbitrary preference in the Dyirbal Ergative, by arguing in detail that no combination of empirically well-justified markedness constraints can be recruited to exert the preference in favor of [-ŋku] which would be required either in general or just in the contexts where [-ŋku] appears. Section 3 lays out my proposal, illustrating it with an analysis of the Dyirbal facts. Section 4 considers previous approaches to arbitrary preference among allomorphs, arguing that my proposal enjoys a number of empirical and conceptual advantages over each of them. Section 5 considers some extensions of the proposal concerning phonologically-induced affix insertion and phonologically-conditioned deponency. Section 6 concludes.

2. The case for arbitrary preference in the Dyirbal Ergative

In order to justify treating the Dyirbal Ergative as involving an arbitrary preference between listed allomorphs, we need to show that no markedness constraints whose existence is empirically justifiable could produce the observed distribution of surface forms. That is, we need to show that there is no markedness constraint whose effects are observable in other languages which would exert the required preference in favor of [-ŋku] over [-ku] in the environments where it appears.

In order to show this, let's first consider the possibility that [-ŋku] and [-ku] are derived from a single underlying form, which is not an altogether implausible idea, given that the two surface allomorphs are so phonologically similar. Suppose first that the shared UR were /ku/. This would have to be transformed via epenthesis into [-ŋku] just in case the Ergative marker were suffixed to a disyllabic stem. Such epenthesis would be hard to phonologically motivate, since Dyirbal has trochaic feet constructed from the left (Dixon 1972, McCarthy & Prince 1993a: main stress is on the initial syllable, with secondary stress on all non-final odd-numbered syllables). Hence the epenthesis would involve adding a coda to an unstressed syllable, which will actually be disfavored by constraints like the Weight-to-Stress Principle (much as observed by Paster 2005), as well as the general NoCoda constraint. Moreover, even if there were a constraint that favored adding codas to unstressed syllables, calling on it for Dyirbal would incorrectly predict that the same epenthesis would occur when /ku/ was attached to a four-syllable stem like /palakara/ 'they', resulting in unattested
*[(pála)$_{\text{HeadFt}}$(kàraŋ)$_{\text{Ft}}$-ku].

Now suppose instead that the shared UR of the Ergative suffix were /-ŋku/. The /ŋ/ would have to delete whenever /-ŋku/ were attached to a greater-than-disyllabic stem. The problem here would be to explain why the /ŋ/ would delete to avoid adding a coda to

(and not just the selection of the lexical items that spell out that structure) occurs in the same grammar as the phonology. Limits of space and time regrettably prevent me from exploring differences between the two frameworks in the current paper.

the unstressed second syllable of a non-head foot, as in [(pá.la)$_{\text{HeadFt}}$(kà.ra)$_{\text{Ft}}$-ku]$_{\text{Wd}}$ 'they-ERG', but not to avoid adding a coda to the unstressed second syllable of a head foot, as in [(yá.ɽaŋ)$_{\text{HeadFt}}$-ku]$_{\text{Wd}}$ 'man-ERG'. Conceivably, one might suggest that /ŋ/ escapes deletion only when it is a coda of a syllable in the head foot, due to positional faithfulness (Beckman 1998). This, however, cannot be the case in Dyirbal, since the language has numerous suffixes that contain coda segments which are not subject to deletion, even outside of the head foot. For instance, the transitive verbalization of [wa.ɽu] 'bend' is [(wá.ɽu)$_{\text{HeadFt}}$-mal] 'make bendy' (Dixon 1972: 86), and not *[(wá.ɽu)$_{\text{HeadFt}}$-ma], as we would expect if affix segments in coda position were only exempt from deletion inside the head foot[8].

There is, in sum, little hope of giving an analysis of the [-ŋku]~[-ku] allomorphy of the Dyirbal ergative which derives the two surface allomorphs from some common underlying form. We now turn to consider whether a TETU analysis like the one given for Moroccan Arabic would be possible. Such an analysis would presumably take the Ergative suffix to have the two underlying forms {/-ŋku/, /-ku/}. Because these two forms are quite similar to one another, relatively few markedness constraints will distinguish between them. It is reasonably clear, though, that those markedness constraints that do will reliably prefer [-ku] over [-ŋku]. *NASAL would be one such constraint; NOCODA would be another, and *NC̥ ('No sequences of a nasal followed by a voiceless segment': Pater 1999) would be a third. In the case of greater-than-disyllabic stems, these constraints would be assumed to be responsible for the observed victory of a /-ku/-selecting candidate over a /-ŋku/-selecting one:

(5) Dyirbal: 'rainbow-ERG'

yamani + {/-ŋku/, /-ku/} *Inputs:* *Outputs:*		M	*NASAL	NOCODA
yamani + ku	→ [ya.ma.ni.ku]		2	
yamani + ŋku	~ a. [ya.ma.niŋ.ku]		W_3	W_1

Candidate (5)a, by virtue of choosing the underlying form /-ŋku/, has one more nasal consonant and one more coda than does the winner, which chooses /-ku/. Since the two candidates are identical but for the presence of that coda nasal, it would be reasonable to assume that no other markedness constraints distinguish between them, and hence the /ku/-selecting candidate wins, as we in fact observe in Dyirbal.

Meanwhile, there is a passive onlooker in tableau (5): the constraint I have labeled M. This is the constraint that will have to prefer [-ŋku] over [-ku] when the stem is disyllabic, and which will have to dominate constraints like *NASAL and NOCODA, in

[8] The transitive verbalizer exhibits an allomorphy that potentially challenges any attempt to give a purely markedness-based analysis of [-ŋku]~[-ku] not involving arbitrary preference. This verbalizer is [-mal] after a two-syllable stem, but [-(m)pal] after a longer stem. Dixon gives no details of when the [m] of the second allomorph does and does not appear, but what we potentially have here is a case where the allomorph with an initial cluster is preferred with longer stems, which is just the opposite of what takes place with the Ergative suffix.

order to override those constraints' preference for [-ku] regardless of the length of the stem. The problem is that there likely is no such constraint. As we have noted, CON is unlikely to contain any universal markedness constraint that would prefer [-ŋku] over [-ku] in general—and certainly not any markedness constraint that would exert that preference just in case the preceding stem were disyllabic.

There is one conceivable markedness-based strategy for enforcing a preference for [-ŋku] over [-ku]. This would be to assume that [-ku] is indexed to an Alignment constraint (McCarthy & Prince 1993b) which forbids it to be affixed to a disyllabic stem. As mentioned, a disyllabic stem in Dyirbal will be co-extensive with the head foot of the prosodic word. The requisite Alignment constraint would then require that [-ku] not have the head foot immediately to its left. Such an Alignment constraint, stated in negative terms, goes outside the schema of Generalized Alignment constraints as they are proposed in McCarthy & Prince (1993b), wherein these constraints are always stated in positive terms. There is probably good reason for not expanding the possible form of Alignment constraints in this manner: negative alignment constraints (e.g. 'the {L, R} edge of morpheme X must not be aligned with the {L, R} edge of a stressed syllable') would predict, among other things, systems of infixation where infixes target the edgemost *un*stressed syllable/foot/etc. Such systems do not seem to be attested (Yu 2003, Fitzpatrick to appear).

In sum then, there is clearly no hope for a TETU analysis of the allomorphy of the Dyirbal Ergative because there is no plausibly-extant markedness constraint that will exert the required preference for [-ŋku] in the contexts where it appears. Therefore, if we are to have the choice of [-ŋku] vs. [-ku] to take place within the phonological component of the grammar, we will require some mechanism for exercising an arbitrary preference in favor of [-ŋku]. The next section lays out my proposal about how to enforce that preference within EVAL.

3. Encoding arbitrary preference in spell-out

The allomorphs [-ŋku] and [-ku] stand in a special-general relationship: [-ŋku] appears in one specific context (after a disyllabic stem), and [-ku] appears elsewhere. In this case, the context of the special allomorph is phonologically defined. There are, however, many other instances of competing special vs. general allomorphs, where the context of the special allomorph is morphosyntactically defined.

In realizational theories of morphology like A-Morphous Morphology (Anderson 1992) and Distributed Morphology (Halle & Marantz 1993, 1994), the latter type of special/general competition is handled using versions of what Halle (1997) formulates as the Subset Principle. This is the requirement that, of all the available lexical items or realization rules in a language, the one employed in a given morphological position is the one which realizes or 'spells out' the largest possible subset of the morphosyntactic features in that position, without any mismatches. A simple example comes from agreement morphology on Dutch strong adjectives (Sauerland 1995). Neuter singular strong adjectives have no overt agreement morphology; all other adjectives bear a suffix

–e. In terms of two binary morphosyntactic features [±neuter] and [±plural], the distributions of these two affixes can be described as shown below:

(6)

	[-neuter]	[+neuter]
[-plural]	-e	Ø
[+plural]	-e	-e

A Subset Principle-based account of these facts would posit (as Sauerland 1995 does) that the Dutch lexicon contains the following two agreement suffixes:

(7) Ø spells out [+neuter, -plural]
-e spells out no features

Dutch adjectives, we may assume, include an Agreement position bearing specifications for the features [±neuter] and [±plural]. Per the Subset Principle, the morphology will associate with this position the morpheme which spells out the largest subset of these features it can, with no mismatches. So, when an adjective bears the features [+neuter, -plural], the phonologically-null morpheme Ø will be chosen, since it is specified in the lexicon as spelling out both of these features, whereas the competing suffix *–e* spells out no agreement features. On the other hand, if the adjective is, say, [+neuter, +plural], insertion of Ø will not be an option, because Ø is specified to spell out [-plural], which would be a featural mismatch. Hence *–e* is used instead.

Let's now illustrate how this model of competition between Ø and *–e* could be handled in an OT framework. We will need, first of all, a constraint that will enforce the pressure to spell out as many features as possible. The family of such constraints can be defined as follows:

(8)
SPELL-OUT(F): For every instance of the morphosyntactic feature [F] residing on a feature-bundle β in the input, assign a violation-mark if a morpheme specified as spelling out [F] has not been inserted on β.

I will assume, along with Halle & Marantz (1993, 1994) that the input to the morphological component of the grammar consists of a set of bundles of morphosyntactic features, which occupy terminal nodes of the tree structure emitted by the syntax. Morphemes (or 'vocabulary items', in Distributed Morphology parlance) compete for insertion onto these terminals. Constraints of the SPELL-OUT(F) will favor insertion of morphemes which spell out more of the features on a given terminal over those that spell out fewer features, since every un-spelled-out feature X results in a violation of SPELL-OUT(X).[9]

[9] For previous OT implementations of Distributed Morphology, which make very similar proposals, see Noyer (1993), Bonet (1994), and Trommer (2000).

The second thing we need for our analysis of Dutch is a constraint which will over-ride the preferences of SPELL-OUT(F) by penalizing the insertion of morphemes whose featural specifications don't match those of the terminal they're inserted onto. This constraint can be defined as follows:

(9)

MATCH(F): For every instance of the morphosyntactic feature F which a morpheme is specified as spelling out, assign a violation-mark if that morpheme has been inserted onto a feature-bundle which does not contain F.

We are now equipped to provide an OT analysis of Dutch. First, consider what happens when the input contains a [+neuter, -plural] feature-bundle:

(10)

[+neuter, -plural]	MATCH ([+neuter])	SPELL-OUT ([+neuter])	SPELL-OUT ([-plural])
→ [+neuter, -plural] | *{+neuter, -plural}*$_1$			
~ a. [+neuter, -plural] | *{}*$_2$ e$_2$		W$_1$	W$_1$

The notation I use for candidates will require some explanation. In square brackets at the top of each cell are the feature-bundles emitted by the syntax, which form the input to the morphology. In italics are the abstract morphemes that are inserted onto each of those feature-bundles, expressed as the set of features which the morpheme spells out. At the bottom of each cell, coindexed with the morphemes, are the phonological structures that make up the underlying form associated with each one of those morphemes. No such structure appears in the winning candidate in tableau (10), because the morpheme inserted in this candidate is Ø, which has an empty underlying form.

In the case of (10), the winning candidate is the one which employs the morpheme Ø. This morpheme spells out both of the input features [+neuter] and [-plural], and it does not spell out any features that are absent in the input. As such it incurs no violations of any SPELL-OUT(F) or MATCH(F) constraints. By contrast, candidate (10)a, which has inserted *–e*, violates both SPELL-OUT([+neuter]) and SPELL-OUT([-plural]), because *–e* is not specified as being able to spell out either of these features.

Something different will happen when the input contains a different pair of features, say [-neuter, -plural]:

(11)

[-neuter, -plural]	MATCH ([+neuter])	SPELL-OUT ([+neuter])	SPELL-OUT ([-plural])
→ [-neuter, -plural] | $\{\}_2$ e$_2$		1	1
~ a. [-neuter, -plural] | *{+neuter, -plural}*$_1$	W$_1$	L	L

As before, the candidate which inserts *–e* does so at the cost of violating SPELL-OUT([+neuter]) and SPELL-OUT([-plural]), since *–e* is not specified to spell out any features. However, the candidate which inserts Ø now incurs a violation of MATCH([+neuter]), because Ø is specified as spelling out [+neuter], a feature-value which is not present in the input. If MATCH([+neuter]) dominates both SPELL-OUT([+neuter]) and SPELL-OUT([-plural]), then *–e* will be used rather than Ø when the input contains the feature [-neuter]. Endowing the same high rank on MATCH([-plural]) will, analogously, give the desired result for [+plural] inputs.

In the Dutch example, then, the special-general relationship between Ø and *–e* can be accounted for by assuming that the special allomorph Ø spells out a proper superset of the features that the elsewhere allomorph *–e* does. If lexical selection is governed by the Subset Principle, then the assumptions shown in (7) regarding the lexical specifications of the two morphemes assures that Ø will always be chosen over *–e* in the contexts where Ø is in fact found, with *–e* emerging elsewhere, when selection of Ø would result in a prohibited morphological mismatch.

The competition between [-ŋku] and [-ku] in the Dyirbal Ergative has a parallel character: [-ŋku] is the special case, and [-ku] emerges in its place when use of [-ŋku] would be prohibited. There is the difference, though, that whereas the Dutch special-case allomorph Ø is blocked by a morphosyntactic factor (avoidance of feature mismatches), the Dyirbal special-case allomorph [-ŋku] is blocked by a phonological factor: it cannot be suffixed to a longer-than-disyllabic stem.

The two examples from Dutch and Dyirbal seem to indicate that one pressure within the grammar—the need to prefer privileged, special-case allomorphs—can be over-ridden by both by a pressure to avoid feature mismatches, and also can by the prosodic restrictions on the distribution of affixes. If, in the interest of theoretical economy, we treat the preference for the privileged allomorph in each case as stemming from the same mechanism (a pressure to spell out as many morphosyntactic features as possible), then this suggests that lexical selection occurs in the same OT grammar as the phonology, wherein the constraints responsible for morphological pressures like maximizing feature spell-out interact with—and can be violated under the compulsion of—phonological constraints.

By far the standard assumption in all schools of generative phonology is that the lexical items which express morpho-syntactic features are selected before the phonology begins. (Even in theories with serial morphology-phonology interleaving like Lexical Phonology, the typical assumption is that the insertion of morphemes at a given level precedes the phonology relevant to that level.) The phonology has no role to play in selecting lexical items (that is, selecting which underlying form to employ as the exponent of a given set of morphosyntactic features) except in the event that a single lexical entry has more than one underlying form, as assumed in TETU analyses of listed allomorphy like the one of Moroccan Arabic discussed earlier.

I propose instead that no lexical items have been selected at the point where the phonology gets underway. The input to the phonology consists, I suggest, only of a set of linearly-ordered morpho-syntactic feature bundles. For instance, in Dyirbal, the input for 'man-ERGATIVE' is not /yaɽa-{/ŋku/, /ku/}/ but instead what we can represent as //MAN-ERGATIVE//. In this latter representation, 'MAN' stands for the bundle of semantic and morpho-syntactic features corresponding to the meaning 'man', and 'ERGATIVE' represents the bundle of morpho-syntactic features corresponding to Ergative case.

When candidates are constructed in a TETU analysis of listed allomorphy, GEN has two dimensions of choice in candidate construction. Not only do candidates vary in their output phonological shape, but they also vary in the underlying form to which that output bears a correspondence relation. In TETU analyses, GEN's range of choice with respect to the input is relatively limited: when a morpheme like the Moroccan Arabic 3rd person masculine singular clitic has more than one underlying form, GEN has to pick exactly one of them to which the output bears a correspondence relation and is thus pressured to be faithful to. In the model proposed here, the situation is similar, but GEN's freedom of choice with respect to underlying forms is more pervasive: for each of the feature-bundles in an input representation like //MAN-ERGATIVE//, the grammar can pick zero or more lexical items to associate with it, and it is the underlying forms of those lexical items to which the output representation of a given candidate bears a correspondence relation. In principle, *any* morpheme in the lexicon can be chosen for insertion onto *any* of the feature-bundles, and *any number* of morphemes can be inserted onto a single feature bundle. We thus will require that there be some mechanism to evaluate how good a given morpheme is as a speller-out of a given bundle of morphosyntactic features, and a means to enforce an economy of allomorphs, so that piling up of multiple lexical items on the same feature-bundle will not be the normal case.

The first criterion can be implemented by SPELL-OUT(F) constraints, as I now proceed to illustrate.[10] In what follows, I'll adopt—merely for the sake of illustration, rather than for that of endorsement—the theory of case features proposed in Halle & Vaux (1998). In this theory, Ergative case comprises a bundle of four morphosyntactic

[10] MATCH(F) constraints will not figure in our discussion of Dyirbal, since we will only be considering Ergative inputs, and consequently opportunities to create a feature mismatch will not arise.

features [-oblique, +structural, +superior, -free]. To make [-ŋku] the privileged allomorph, we need then only assume that it spells out some larger subset of these features than [-ku] does, for example:

(12)
a. /-ŋku/ spells out [-oblique, +structural, +superior, -free]
b. /-ku/ spells out [-oblique, +structural, +superior]

Assuming this particular (arbitrarily-chosen) subset-superset relation between the two allomorphs, candidates that insert /-ŋku/ onto the ERGATIVE feature-bundle are preferred over those that insert /-ku/ onto that same feature-bundle by the constraint SPELL-OUT([-free]):

(13) *-ŋku/ gets inserted on Case-slot of disyllabic Ergative nouns*

[MAN]—[-obl, +str, +sup, -fr] *Inputs:*	*Outputs:*	SPELL-OUT([-free])	*NASAL
→ [MAN]—[-obl, +str, +sup, -fr] | | '*man*'$_1$ *{-obl, +str, +sup, -fr}*$_2$ $y_1a_1ʗ_1a_1$ - $ŋ_2k_2u_2$	[yaʗaŋku]		1
~ a. [MAN]—[-obl, +str, +sup, -fr] | | '*man*'$_1$ *{-obl, +str, +sup}*$_3$ $y_1a_1ʗ_1a_1$ - k_3u_3	[yaʗaku]	W_1	L

Now that we are doing phonology and lexical insertion in a single tableau, our candidates will require elaboration into input and output levels. The notation in the 'Inputs' column is identical to that used earlier in the analysis of Dutch. The phonological structures in this column are the underlying forms of the morphemes inserted in each candidate. The output form (which stands in correspondence with that phonological underlying forms and which is evaluated by markedness constraints) appears in the 'Outputs' column.

As we've noted, there are many makedness constraints which will prefer [-ku] over [-ŋku]; in tableau (13) *NASAL is chosen as a representative example. In order for [-ŋku] to ever be able to win, the preferences of these constraints must be overridden by the preference of some higher-ranked constraint. SPELL-OUT([-free]) is just such a constraint: it assigns one violation-mark to the /-ku/-selecting candidate (13)a, because /-ku/ does not spell out [-free], but no violation-marks to the winning, /-ŋku/-selecting candidate, because /-ŋku/ does spell out [-free].

This much will get us the desired results for disyllabic stems. For longer stems, we need /-ku/ rather than /-ŋku/ to be chosen, and this will require that some constraint that prefers /-ku/ over /-ŋku/ just in case the stem is more than two syllables long to be

ranked above SPELL-OUT([-free]). I will follow McCarthy & Prince (1993a) in invoking the following constraint to disprefer /-ŋku/ with longer stems:

(14) AFFIX-TO-FOOT
'The left edge of [ŋku] coincides with the right edge of the head foot'[11]

In Alignment terms, AFFIX-TO-FOOT is equivalent to ALIGN([-ŋku], L, HeadFt, R). As mentioned in the previous section, Dyirbal has trochaic stress with main stress on the initial syllable, and therefore a disyllabic stem will be co-extensive with the head foot. When the stem is disyllabic, /-ŋku/-selecting candidates will satisfy AFFIX-TO-FOOT, because /-ŋku/ appears immediately after the head foot, as AFFIX-TO-FOOT demands. Candidates that select /-ku/ will also satisfy AFFIX-TO-FOOT, albeit vacuously, because /-ku/ is not indexed to that constraint. As such, AFFIX-TO-FOOT exerts no preference between the competing allomorphs when the stem has two syllables, and the preferences of SPELL-OUT([-free]) become decisive.

By contrast, when the stem has more than two syllables, AFFIX-TO-FOOT will disprefer the use of /-ŋku/, and /-ku/ will then emerge as the optimal choice:

(15) */ku/ gets inserted on Case slot of longer Ergative nouns*

[ROOT]—[-obl, +str, +sup, -fr] *Inputs:*	*Ouputs:*	AFFIX-TO-FOOT	SPELL-OUT([-free])
→ [ROOT]—[-obl, +str, +sup, -fr] | | '*rainbow*'$_1$ *{-obl, +str, +sup}$_3$* $y_1a_1m_1a_1n_1i_1$ - k_3u_3	[yamaniku]		1
~ a. [ROOT]—[-obl, +str, +sup, -fr] | | '*rainbow*'$_1$ *{obl, +str, +sup, -fr}$_2$* $y_1a_1m_1a_1n_1i_1$ - $ŋ_2k_2u_2$	[yamaniŋku]	W_1	L

As before, using /-ku/ results in violation of SPELL-OUT([-free]), but this is tolerated for the sake of avoiding instances of /-ŋku/ that are not aligned with the head foot.

So far in this section, we have only considered candidates that insert just one of /-ŋku/ vs. /-ku/. What about a candidate like *[yaɽa-ŋku-ku] where the two allomorphs of the Ergative suffix pile up in the output? In an analysis with disjunctive underlying forms like the one given for Moroccan Arabic, candidates like this are eliminated by

[11] While this constraint may seem to be entirely *ad hoc*, it is not unusual for affixes to be indexed to prosodic size requirements on the bases that host them. For example, English comparative *–er* seems to require a single-foot base: *smarter* but **intelligenter*. See the discussion of Prosodic Circumscription in McCarthy & Prince (1990) for more examples.

stipulating that each candidate has to pick exactly one underlying form to use. In my proposal, as we will now see, there is no need for such a stipulation.

The reason for *[yaɽa-ŋku-ku]'s ungrammaticality resides in phonological markedness. Whenever a phonologically non-null morpheme is inserted, it results in an increased amount of structure in the output, which (in general) means more markedness violations. If lexical insertion occurs in the same module of the grammar as the phonology, the markedness incurred by extra structure assures that the grammar will insert lexical items only to the extent that is necessary to satisfy SPELL-OUT(F) constraints. Consider, for instance, what happens with //MAN-ERGATIVE// in Dyirbal:

(16) *Economy of allomorphs derives from phonological markedness*

[MAN]—[-obl, +str, +sup, -fr] *Inputs:*	*Outputs:*	SPELL-OUT(F)	*[+dorsal]
→ [MAN]—[-obl, +str, +sup, -fr] | | '*man*'$_1$ *{-obl, +str, +sup, -fr}*$_2$ $y_1a_1ɽ_1a_1$ - $ŋ_2k_2u_2$	[yaɽaŋku]		2
~ a. [MAN]—[-obl, +str, +sup, -fr] | \ '*man*'$_1$ *{-obl, +str, +sup}*$_3$ *{-obl, +str, +sup, -fr}*$_2$ $y_1a_1ɽ_1a_1$ - $ŋ_2k_2u_2$ - k_3u_3	[yaɽaŋkuku]		W_3
~ b. [MAN]—[-obl, +str, +sup, -fr] | '*man*'$_1$ $y_1a_1ɽ_1a_1$	[yaɽa]	W_4	L

The winning candidate, [yaɽa-ŋku], spells out all four of the morphosyntactic features that make up ERGATIVE, and hence receives no violation-marks from any SPELL-OUT(F) constraints. However, by having the affixal segments /-ŋku/ present in the output, it violates *[+dorsal], as well as *[+nasal], *[+round], and whatever other markedness constraints disprefer the presence of any of the segments [-ŋku].

The losing candidate *[yaɽa], which inserts no Ergative affix, avoids these markedness violations, but it violates the various SPELL-OUT(F) constraints applicable to the features that make up ERGATIVE. This means that those SPELL-OUT(F) constraints must dominate such markedness constraints as *[+dorsal]. However, these markedness still play a crucial role in making *[yaɽa-ŋku-ku] harmonically bounded. Because inserting /-ŋku/ alone suffices to satisfy all of the SPELL-OUT(F) constraints, there is nothing gained in performance on those constraints by inserting /-ku/ as well. In the

absence of a morphological rationale for inserting both Ergative allomorphs, the presence of an extra [+dorsal] consonant (and an extra round vowel, etc.) in*[yaɽa-ŋku-ku] results in its being harmonically bounded by the attested winner [yaɽa-ŋku].

The present theory, then, is able to derive economy of lexical insertion in the same way that OT derives other kinds of structural economy like limits on epenthesis: for the most part, more structure means more markedness violations (and more DEP violations, in the case of epenthesis), so structure will be added only to the minimum extent required for the satisfaction of higher-ranked constraints.[12, 13] This result gives the lexical-insertion-in-the-phonology approach a clear advantage over a theory which allows morphemes to have multiple underlying forms, which require an independent stipulation to rule out candidates in which multiple allomorphs pile up in the output.

The theory being advocated here has a second conceptual advantage: we no longer need to assume that morphemes can have multiple underlying forms. If lexical insertion precedes the phonology, then analyzing allomorphic alternation that takes place in the phonology requires us to assume that the lexical items chosen by the morphosyntax can have disjunctive phonological content, as we saw in the analysis of Moroccan Arabic. If, on the other hand, the *phonology* is responsible for choosing which lexical items to insert, we are free to regard sets of competing 'allomorphs' like /-ŋku/ and /-ku/ as fully separate lexical entries, each with just a single underlying form.

In this section, I have shown that a lexical-insertion-in-the-phonology approach is capable of giving a descriptively adequate account of the arbitrary preference required for the Dyirbal Ergative, and that it additionally achieves conceptual advantages over a multiple-underlying-forms theory. In the next section, I compare my theory with existing proposals about arbitrary preference among allomorphs, and show that mine has advantages over each of them.

4. Arbitrary preference among allomorphs: Previous proposals

The first OT-based proposal for implementing an arbitrary preference among allomorphs is found in the analysis of the Dyirbal Ergative in McCarthy & Prince (1993a: ch. 7). They argue for a serial analysis in which the privileged allomorph /-ŋku/ is 'tried' first. That is, in the first pass of constraint evaluation, the input contains only the allomorph /-ŋku/ rather than the pair of underlying forms {/-ŋku/, /-ku/}. They further

[12] For relevant discussion of economy of structure in OT, see Grimshaw (2003) and Gouskova (2003).

[13] Given my claim that phonological markedness will countervail against the insertion of (most) phonologically non-null morphemes, the analysis of Dutch presented earlier will require some revision. Specifically, we may need to discard the assumption that *–e* spells out literally *no* features, since the markedness violations that come from adding this vowel would not be counterbalanced by any gain in performance on SPELL-OUT(F) constraints. This matter might be solved by assuming that the features [±neuter] and [±plural] are terminals of a morphosyntactic feature geometry (as in Harley & Ritter 2002), and that *–e* spells out a single non-terminal feature of the tree which would be shared by all of the cells in (6)—perhaps simply the root node RE ('referring expression'). Alternatively, we could posit a constraint which demanded that every feature-bundle have at least one morpheme inserted onto it; such a constraint would favor the insertion of *–e* over the insertion of no morpheme at all, even if *–e* spells out no features.

assume that the allomorph /-ŋku/ is indexed to the constraint AFFIX-TO-FOOT defined in (14). If the constraint AFFIX-TO-FOOT, along with ALIGN([ŋku],R, PWd, R), which will militate against infixing /-ŋku/, are ranked above the anti-null-parse constraint MPARSE (Prince & Smolensky 2004 [1993], Wolf & McCarthy to appear), the null parse will win on the first pass of constraint evaluation if the stem involved is greater than disyllabic:

(17) */yamani-ŋku/ leads to null parse*

/yamani-ŋku/	AFFIX-TO-FOOT	ALIGN([ŋku],R, PWd,R)	MPARSE
→ *null parse*			$_1$
~ a. $[(\text{yá.ma})_{\text{HeadFt}}(\text{ nì.ŋku})_{\text{Ft}}]_{\text{Wd}}$	W_1		L
~ b. $[(\text{yá.ma})_{\text{HeadFt}}(\text{ŋkù.ni})_{\text{Ft}}]_{\text{Wd}}$		W_1	L

The fully-faithful candidate (17)a violates AFFIX-TO-FOOT because the suffix [-ŋku] is not adjacent to the head foot. AFFIX-TO-FOOT is satisfied in (17)b, which infixes /-ŋku/ so as to place it immediately to the right of the head foot, but this involves displacing [-ŋku] from the right edge of the PWd, and hence violating ALIGN([ŋku],R, PWd, R). Both of these constraints are ranked above MPARSE, and so the null parse (which, by hypothesis, violates no constraint but MPARSE) is the winner.

Because the pass of constraint evaluation with /-ŋku/ in the input fails to produce any output, the grammar then tries again with the elsewhere allomorph /-ku/ in the input instead. Because /-ku/ is not indexed to AFFIX-TO-FOOT, that constraint can no longer be violated, and the fully-faithful candidate now defeats the null parse:

(18)

/yamani-ku/	AFFIX-TO-FOOT	ALIGN([ŋku],R, PWd)	MPARSE
→ $[(\text{yá.ma})_{\text{Ft}}(\text{ nì.ku})_{\text{Ft}}]_{\text{Wd}}$			
~ a. *null parse*			W_1

While this approach does work, it suffers from the conceptual drawback of having to impose an external mechanism of 'order of trying' to get the desired result. Since competition is the essence of OT, we would prefer on grounds of parsimony for competition like that between /-ŋku/ and /-ku/ in Dyirbal to be hashed out *within* an OT grammar.[14]

The MPARSE approach also suffers from a second, more serious drawback: it is unable to account for systems of allomorphy in which some realization (whether faithful

[14] The MPARSE approach to arbitrary preference does have the possible advantage of being able to accommodate competition between synthetic and periphrastic expressions with the same meaning, where the two competing forms might not plausibly belong to the same candidate set, at least not at the level of the phonology. The question of whether competition between synthesis and periphrasis exists is a controversial one, and addressing it would be well beyond the scope of this paper; Embick & Marantz (2006) is a recent entry denying that there is such competition, which cites a number of works taking the opposite view.

or unfaithful) of both allomorphs would be more harmonic than the null parse. Consider, as an example, the indefinite article in English, which is *a* ([eɪ]~[ə]) before a following consonant-initial word and *an* ([æn]) before a following vowel-initial word. If there were an arbitrary preference between these two allomorphs, then one of them would have to yield the null parse as winner when it was 'tried' in the environment where it does not appear. If *a* were the special case, the null parse would have to be the optimal output for an input like *a apple*, and likewise if *an* were the special case, the null parse would have to be the optimal output for an input like *an duck*. The problem here is that these inputs would respectively result in the phonotactically-permitted outputs [ə.ʔæp̩l] and [æn.dʌk]. The null parse would never win, and the elsewhere allomorph would never get to be tried.

Given this, the MPARSE approach would require that systems of listed allomorphy which involve arbitrary preference have a fundamentally different architecture from at least some systems that do not. In the former case, there would be only one allomorph in the input at a time, with allomorphs being tried in the requisite order, whereas in the latter case, both allomorphs would have to be present in the input simultaneously, as in the analysis given for Moroccan Arabic in (2)-(3).

This situation stands in contrast to that of the lexical-insertion-in-the-phonology view advocated here. In the present proposal, systems that involve arbitrary preference and those that do not can be analyzed using exactly the same architecture. For a case like that of the Moroccan Arabic 3rd person masculine singular clitic, where there is no arbitrary preference, we simply need to assume that the two lexical items /h/ and /u/ each spell out exactly the same set of morphosyntactic features. This means that all constraints of the SPELL-OUT(F) and MATCH(F) families will be indifferent as to whether to use /h/ or /u/, and the choice will be left entirely up to the phonological constraints.

The second existing proposal about arbitrary preference is advanced by Bonet, Lloret, and Mascaró (2005) and Mascaró (2005), and also is used in Bonet (2004) and Kikuchi (2006)[15]. It involves the following constraint:

(19) PRIORITY. Respect lexical priority (ordering) of allomorphs.
Given an input containing allomorphs m_1, m_2, …, m_n, and a candidate m_i', where m_i' is in correspondence with m_i, PRIORITY assigns as many violation marks as the depth of ordering between m_i and the highest dominating morph(s).
(Definition from Mascaró 2005: 15)

PRIORITY-based analyses are architecturally identical in form to multiple-UR TETU analyses like the one presented earlier for Moroccan Arabic: there is only one pass of constraint evaluation, and all of the competing allomorphs in the input at once. The following tableaux, adapted from Bonet (2004), illustrates how the PRIORITY approach handles the Dyirbal facts:

[15] The *ad hoc* constraint GENITIVE=/-n/ which Kager (1996) uses for Djabugay is effectively equivalent to the PRIORITY proposal.

(20) Dyirbal 'man.ERG' with PRIORITY

/ yaɽa -{ŋku, ku}/	AFFIX-TO-FOOT	PRIORITY
→ [(yá.ɽa)$_{\text{HeadFt}}$ ŋku]$_{\text{Wd}}$		
~ a. [(yá.ɽa)$_{\text{HeadFt}}$ ku$_{\text{Ft}}$]$_{\text{Wd}}$		W_1

With a disyllabic stem, as in (20), [-ŋku] and [-ku] can both be suffixed to the head foot. Candidate (20)a thus satisfies AFFIX-TO-FOOT, because [-ŋku] is so suffixed, and (20)b satisfies the same constraint, vacuously because /-ku/ is not indexed to AFFIX-TO-FOOT. Since AFFIX-TO-FOOT is indifferent as to the choice of allomorphs, the choice is made by the lower-ranked constraint PRIORITY. The winning candidate chooses the first-listed underlying form /-ŋku/ and thus gets no marks from PRIORITY. By contrast, candidate (20)a chooses the second-listed underlying form /-ku/. It therefore gets one mark from PRIORITY, and thus loses.

Now consider what happens with a greater-than-disyllabic stem:

(21) Dyirbal 'rainbow.ERG' with PRIORITY

/yamani-{ŋku, ku}/	AFFIX-TO-FOOT	PRIORITY
→ [(yá.ma)$_{\text{HdFt}}$(nì.ku)$_{\text{Ft}}$]$_{\text{Wd}}$		1
~ a. [(yá.ma)$_{\text{HdFt}}$(nì.ŋku)$_{\text{Ft}}$]$_{\text{Wd}}$	W_1	L

Because, when the stem is more than two syllables long, the head foot is no longer at the right edge of the Prosodic Word, neither allomorph of the Ergative suffix can be suffixed to the head foot. As a result, the candidate that chooses the first-listed underlying form /-ŋku/ incurs a violation from AFFIX-TO-FOOT. The candidate that chooses /-ku/ gets no such violation (again vacuously because /-ku/ is not indexed to AFFIX-TO-FOOT), and so the /-ku/-selecting candidate now emerges as the winner, because AFFIX-TO-FOOT is higher-ranked than PRIORITY.

The PRIORITY approach is thus able to get the desired results because the constraint PRIORITY can exert whatever pairwise preferences are needed: we simply list the desired preference-order among allomorphs into the lexicon, and PRIORITY does the rest. This approach is not without its problems, though. The main one is that, unless we stipulate a universal upper bound on the number of listed allomorphs that a single morpheme can have, PRIORITY will have to evaluate candidates gradiently, since it assigns (*n*-1) violation-marks to candidates that pick the *n*th highest-priority allomorph. For example, Bonet, Lloret, & Mascaró (2005) propose that the masculine gender suffix in Catalan has three underlying forms, which are ordered in the preference hierarchy {∅>u>ə}[16]. To give the desired effects, PRIORITY needs to assign one violation-mark to candidates that pick /-u/ and two violation-marks to candidates that pick /-ə/. Given that, outside of suppletive allomorphy, gradient evaluation is never necessary and is

[16] There are other analyses which posit the existence of more than two allomorphs of a single morpheme: Mascaró (2005) argues that the infinitive marker in Baix Empordà Catalan has six and that the Classical Arabic definite article has fourteen, while Wolf (to appear) proposes that morphemes that trigger the 'mixed mutation' in Breton have four.

frequently empirically undesirable (McCarthy 2003a, 2004), we have good reason to look for an alternative to PRIORITY which requires only categorical evaluation. SPELL-OUT(F) constraints are just such an alternative: they assess categorically, assigning a single violation mark for each instance of the feature F in the input that is not spelled out in the output.

The third and final existing proposal about arbitrary preference also uses categorical constraints. Picanço (2002) proposes that, when a morpheme has multiple allomorphs, each allomorph is indexed to a PARSE-MORPH constraint (Akinlabi 1996). The constraint PARSE-MORPH(X) is violated by a given candidate if that candidate fails to pick allomorph *X*, so the ranking of the various PARSE-MORPH constraints will determine the order of (arbitrary) preference among the allomorphs. For instance, in Dyirbal, for input /yaɽa-{/ŋku/, /ku/}/, the candidate [yaɽa-ŋku] violates PARSE-MORPH(-gu), while *[yaɽa-ku] violates PARSE-MORPH(-ŋku). If PARSE-MORPH(-ŋku) is higher-ranked, then [yaɽa-ŋku] will be preferred over *[yaɽa-ku]. There is a problem, though: there is no clear reason why both of these candidates should not be bested by *[yaɽa-ŋku-ku], which satisfies *both* PARSE-MORPH constraints by virtue of preserving both allomorphs in the output. Because it contains no mechanism to assure (or even violably prefer) the mutual exclusivity of the various allomorphs of a single morpheme, a PARSE-MORPH approach to arbitrary preference does not look like the right way to go.[17] By contrast, as shown in (16), the lexical-insertion-in-the-phonology approach is able to derive the required economy of allomorphs from the effect of phonological markedness constraints.

5. Extensions

I have advocated in this paper the proposition that the insertion and selection of lexical items takes place entirely in the same OT grammar as the phonology. While the need for arbitrary preference in certain systems of allomorphy provided the motivation for this proposal, I will now proceed to argue that it results in several further advantages.

One potentially advantageous consequence is that morphemes could be inserted for solely phonological reasons. That is, if "insert a lexical item" is one of the operations available to GEN, this could be done not only to satisfy constraints like SPELL-OUT(F) but also as an alternative to epenthesis. There are situations in a few languages for which essentially this analysis has been suggested. Hale (1973) argues that the following word-final augmentation rule exists in certain Western Desert dialects such as Pitjantjatjara:

(22) ∅→ pa / C___#

[17] Steriade (1999, 2000) does propose that properties of different allomorphs can be combined in the output. For a critique of PARSE-MORPH constraints as an approach to consonant mutation and other 'process' morphology (the purpose for which Akinlabi 1996 proposes them), see Wolf (to appear).

The process is clearly conditioned by the presence of what would otherwise be a word-final consonant—in Pitjantjatjara, when the stem is followed by a V-final suffix, the augmentative /pa/ doesn't show up:

(23)

uninflected	*ergative*	*dative*	
maṇkurpa	maṇkur-tu	maṇkur-ku	'three'
punpunpa	punpun-tu	punpun-ku	'fly'

The /pa/ also appears after certain verbal suffixes: /-n, -ṇin, -ŋin, -nin/ ~ /-npa, -ṇinpa, -ŋinpa, -ninpa/.

This augmentation process is theoretically challenging because the marked status of [+labial] place means that epenthesis of [+labial] consonants should be impossible. The tableau below illustrates the analysis that I propose for Pitjantjatjara, and the markedness problem that would arise for the assumption that the augmentative [-pa] were epenthetic:

(24)

[THREE] *Inputs:*	*Outputs:*	DEP	*C]$_{PWd}$	*[labial]
→ [THREE] | *'three'*$_1$ {}$_2$ $m_1a_1n_1k_1u_1r_1$ p_2a_2	[man.kur.pa]			$_2$
~ a. [THREE] | *'three'*$_1$ $m_1a_1n_1k_1u_1r_1$	[man.kur]		W_1	L_1
~ b. [THREE] | *'three'*$_1$ $m_1a_1n_1k_1u_1r_1$	[man.kur.**ta**]	W_2		L_1
~ c. [THREE] | *'three'*$_1$ $m_1a_1n_1k_1u_1r_1$	[man.kur.**pa**]	W_2		$_2$

*(Epenthetic segments are in **bold**)*

The markedness constraint responsible for /pa/-insertion is what we can call *C]$_{PWd}$, which bans Prosodic Words from ending in a consonant (see Flack in prep. for

extensive typological justification of this constraint). For the input //[THREE]//, i.e. the root meaning 'three' alone, with no inflection, the winning candidate is [mankurpa]. This candidate has inserted the root morpheme /mankur/, as well as the semantically-empty affix /pa/. The presence of /pa/ means that the winning candidate satisfies *C]$_{PWd}$, but it also means that the candidate incurs an extra violation of *[labial], by virtue of containing the segment /p/.

One competitor of the observed winner is (24)a, which inserts only the root morpheme /mankur/. This candidate does better than the winner on *[labial], due to the absence of [-pa], but it loses by virtue of violating the higher-ranked constraint *C]$_{PWd}$. Of greater interest are the competitors (24)b-c, with epenthesis. Both of these candidates violate the anti-epenthesis constraint DEP, by virtue of epenthesizing the sequences [ta] or [pa]. By contrast, the winner does not violate DEP, because all of its surface segments—including the [pa]—stand in correspondence with the segments in the underlying form of some morpheme. Because of this, candidate (24)c is harmonically bounded by the winner—the two do identically on all markedness constraints, but only (24)c violates DEP.

Crucially for my argument, (24)c is *also* harmonically bounded by (24)b, which epenthesizes [ta] rather than [pa]. The two perform identically on DEP and *C]$_{PWd}$, but (24)b is more harmonic than (24)c because the coronal [t] is less marked than the labial [p]. This means that if lexical insertion were not available as a 'repair' in the phonological component of the grammar—that is, if the depicted winner in tableau (24) were not a possible candidate—there would be no way for *C]$_{PWd}$–violation to be avoided by insertion of [pa], since [pa]-epenthesis should always be harmonically bounded by [ta]-epenthesis (except in specific contexts that might favor the presence of a labial, e.g. adjacent to another labial).

The prediction of markedness theory that marked segment types like labials can never be epenthetic is largely supported by typological surveys (e.g. de Lacy 2002). The analysis that I offer thus somewhat complicates the status of epenthetic quality as evidence about markedness, since any segment, no matter how marked, could in principle belong to the UR of a morpheme inserted for phonological reasons. This does not seem tremendously worrisome, though, as there are various diagnostics that will often be available to distinguish epenthetic segments from affix segments. For example, de Lacy (2002) notes that apparent epenthetic round vowels in Seri, Hungarian and Icelandic are restricted to particular morphological contexts, and suggests that these segments are therefore likely to be morphemes rather than true epenthetic segments. Hale (1973) identifies similar conditions on the distribution of augmentative /-pa/ in Pitjantjatjara —it does not appear with vocatives or after the 2nd person singular clitic /-n/—and argues therefrom that /-pa/ is a morpheme. Looking beyond surface evidence, there are also likely to be experimental means for disentangling the epenthetic vs. affixal status of segments.

A second, related, potential field of application of the theory proposed here is phonologically-conditioned deponency, i.e. insertion for phonological reasons of

morphemes that spell out features other than those present in the input. One widely-discussed case that has been argued to have this character occurs in French. Certain adjectives and determiners in that language show a suppletive alternation between a V-final form in the masculine and a C-final form in the feminine: [nuvo] *nouveau* 'new.MASC', [nuvɛl] *nouvelle* 'new.FEM'. However, when the adjective is followed by a V-initial noun, the feminine form is used, even if the noun is masculine: [nuvɛl ã] *nouvel an* 'new year'. (V-initial words that have the so-called *h aspiré* do not trigger unexpected use of the feminine.) Tranel (1996a,b) and Perlmutter (1998) analyze this alternation by assuming that a constraint against hiatus dominates a constraint demanding gender concord between the noun and the adjective—which, in terms of the analysis given earlier for Dutch, would presumably be a constraint of the MATCH(F) family. As the proponents of this analysis of French have noted, for that kind of constraint interaction to take place, the selection of morphemes based on morphosyntactic criteria like gender agreement can't take place before the phonology—the two have to happen in the same OT grammar, just as this paper has argued to be necessary for dealing with arbitrary preference among allomorphs.[18] Other possible cases of phonologically-conditioned deponency are reported in English (Dixon 1982) and Ondarroa Basque (Côté 1999, 2000). Also relevant to the matters discussed in this section is Cook (1971), who argues that a certain morpheme in Sarcee is inserted only when its presence is necessary to avoid accidental homophony in a paradigm.

In sum, then, it appears that a whole range of morphological preferences regarding lexical selection can be over-ridden on phonological grounds. The phonology can displace a privileged allomorph in favor of its elsewhere competitor (as in the Dyirbal ergative), it can force the insertion of a morpheme that does not match the morphosyntactic feature content of the word that bears it (as in French adjectives), and it can force the insertion of a morpheme whose presence is simply not needed for any morphosyntactic reason (as in Pitjantjatjara [pa]-augmentation). The convergence of all these cases strongly suggests that lexical insertion and the phonology occupy a single OT grammar, as this paper has argued.

6. Conclusion

Phonologists typically assume that morphemes drawn from the lexicon are chosen by the morpho-syntax before the phonology gets underway. In this paper, I have argued that there are empirical and conceptual benefits to assuming instead that morphemes are inserted within the phonology. With respect to the treatment of phonologically-conditioned suppletive allomorphy, I have argued that an approach based on that assumption offers the best available analysis of cases of allomorphy involving arbitrary preference, as in the Dyirbal ergative, avoiding problems faced by MPARSE, PRIORITY, and PARSE-MORPH theories. Additionally, unlike multiple-UR approaches to allomorphy, this approach avoids the need to impose an additional stipulation that candidates can use only one allomorph. I have also explored potentially beneficial extensions of the theory to insertion of morphemes for strictly phonological reasons, and

[18] See Janda (1998) and Lapointe & Sells (1997) for criticisms of this approach to the French data, and Steriade (2000) for a somewhat different view.

to phonologically-motivated deponency. It thus looks very much like placing lexical insertion fully inside the phonological component of the grammar will allow OT to make headway on several aspects of the phonology-morphology interface.

References

Akinlabi, Akinbiyi (1996). Featural affixation. *Journal of Linguistics* **32**, pp. 239-289. [ROA-185]

Anderson, Stephen R. (1992). *A-Morphous Morphology*. Cambridge: Cambridge University Press.

Beckman, Jill N. (1998). *Positional Faithfulness*. Ph.D. dissertation, University of Massachusetts, Amherst. [ROA-234]

Blevins, Juliette (1995). The syllable in phonological theory. In John A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell, pp. 206-244.

Bonet, Eulàlia (1994). The Person-Case constraint: A morphological approach. In Heidi Harley and Colin Phillips (eds.), *MIT Working Papers in Linguistics* **22**: *The Morphology-Syntax Connection*. Cambridge, MA: MITWPL, pp. 33-52.

Bonet, Eulàlia (2004). Morph insertion and allomorphy in Optimality Theory. *International Journal of English Studies* **4**.2, pp. 73-104. [Special issue on 'Advances in Optimality Theory', ed. by Paul Boersma & Juan Antonio Cutillas.] [ROA-734]

Bonet, Eulàlia, Maria-Rosa Lloret, and Joan Mascaró (2005). Lexical specification and ordering of allomorphs: two case studies. Ms., Universitat Autònoma de Barcelona and Universitat de Barcelona.

Bye, Patrik (2006). Allomorphy – selection, not optimization. Ms., Universitetet i Tromsø.

Clements, G.N. (1986). Syllabification and epenthesis in the Barra dialect of Gaelic. In Koen Bogers, Harry van der Hulst, and Maarten Mous (eds.), *The Phonological Representation of Suprasegmentals*. Dordrecht: Foris, pp. 317-336.

Cook, Eung-Do (1971). Phonological constraint and syntactic rule. *Linguistic Inquiry* **2**, pp. 465-478.

Côté, Marie-Hélène (1999). Edge effects in the prosodic hierarchy: Evidence from stops and affricates in Basque. In Pius Tamanji, Masako Hirotani, and Nancy Hall (eds.), *Proceedings of NELS 29*. Amherst: GLSA, pp. 51-65/

Côté, Marie-Hélène (2000). *Consonant Cluster Phonotactics: A Perceptual Approach*. Ph.D. dissertation, MIT, Cambridge, MA.

de Lacy, Paul (2002). *The Formal Expression of Markedness*. Ph.D. dissertation, University of Massachusetts, Amherst. [ROA-542]

Dixon, R.M.W. (1972). *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Dixon, R.M.W. (1982). *Where Have All the Adjectives Gone?: And Other Essays in Semantics and Syntax*. Berlin: Mouton.

Elfner, Emily (2006). Contrastive syllabification in Blackfoot. In Donald Baumer, David Montero, and Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, pp. 141-149.

Embick, David, and Alec Marantz (2006). Architecture and blocking. Ms., University of Pennsylvania and Massachusetts Institute of Technology. [LingBuzz/000262]
Fitzpatrick, Justin (to appear). A concatenative theory of possible affix types. In Andrés Pablo Salanova (ed.), *Papers from EVELIN 1*. Cambridge, MA: MITWPL.
Flack, Kathryn (in prep.) *The Typology of Word-Initial Restrictions*. Ph.D. dissertation, University of Massachusetts Amherst.
Golston, Chris (1996). Direct Optimality Theory: Representation as pure markedness. *Language* **72**, pp. 713-748.
Gouskova, Maria (2003). *Deriving Economy: Syncope in Optimality Theory*. Ph.D. dissertation, University of Massachusetts, Amherst. [ROA-610]
Grimshaw, Jane (2003). Economy of structure in OT. In Angela C. Carpenter, Paul de Lacy, and Andries W. Coetzee (eds.), *University of Massachusetts Occasional Papers in Linguistics* **26**: *Papers in Optimality Theory II*. Amherst: GLSA, pp. 81-120. [ROA-434].
Hale, Kenneth (1973). Deep-surface canonical disparities in relation to analysis and change: An Australian example. In Thomas Sebeok (ed.), *Current Trends in Linguistics*, vol. 11. The Hague: Mouton, pp. 401-458.
Halle, Morris (1997). Distributed Morphology: Impoverishment and Fission. In Benajamin Bruening, Yoonjung Kang, and Martha McGinnis (eds.), *MIT Working Papers in Linguistics* **30**: *PF: Papers at the Interface*. Cambridge, MA: MITWPL, pp. 425-449.
Halle, Morris, and Alec Marantz (1993). Distributed Morphology and the pieces of inflection. In Kenneth Hale and Samuel Jay Keyser (eds.), *The View From Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. Cambridge, MA: MIT Press, pp. 111-176.
Halle, Morris and Alec Marantz (1994). Some key features of Distributed Morphology. In Andrew Carnie, Heidi Harley, and Tony Bures (eds.), *MIT Working Papers in Linguistics* **21**: *Papers on Phonology and Morphology*. Cambridge, MA: MITWPL, pp. 275-288.
Halle, Morris and Bert Vaux (1998). Theoretical aspects of Indo-European nominal morphology: The nominal declensions of Latin and Armenian. In Jay Jasanoff, H. Craig Melchert, and Lisi Olivier (eds.), *Mír Curad: Studies in Honor of Calvert Watkins*. Innsbruck: Institut für Sprachwissenschaft der Universität Innsbruck, pp. 223-240.
Harley, Heidi, and Elizabeth Ritter (2002). Person and number in pronouns: A feature-geometric analysis. *Language* **78**, 482-526.
Hayes, Bruce (1989). Compensatory lengthening in moraic phonology. *Linguistic Inquiry* **20**, pp. 253-306.
Janda, Richard D. (1998). Comments on the paper by Perlmutter. In Steven G. Lapointe, Diane K. Brentari, and Patrick M. Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications, pp. 339-359.
Kager, René (1996). On affix allomorphy and syllable counting. In Ursla Klienhenz (ed.), *Interfaces in Phonology*. Berlin: Akademie Verlag, pp. 155-71. [ROA-88]
Kikuchi, Seiichiro (2006). On Galician definite article allomorphy. *On'in Kenkyu [Phonological Studies]* **9**, pp. 41-48.

Klein, Thomas B. (2003). Syllable structure and lexical markedness in creole morphophonology: determiner allomorphy in Haitian and elsewhere. In Ingo Plag (ed.), *The Phonology and Morphology of Creole Languages*. Tübingen: Niemeyer, pp. 209-228. [ROA-517]

Lapointe, Steven (1999). Stem selection and OT. In Geet Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1999*. Dordrecht: Kluwer, pp. 263-297.

Lapointe, Steven G., and Peter Sells (1997). Separating syntax and phonology in Optimality Theory: The case of suppletive segment/∅ allomorphy. Ms., University of California-Davis and Leland Stanford Junior University.

Mascaró, Joan (1996). External allomorphy as emergence of the unmarked. In Jacques Durand and Bernard Laks (eds.), *Current Trends in Phonology: Models and Methods*. Salford, Manchester: European Studies Research Institute, University of Salford, pp. 473-83.

Mascaró, Joan (2005). External allomorphy and lexical representation. Ms., Universitat Autònoma de Barcelona.

McCarthy, John J. (2003a). OT constraints are categorical. *Phonology* **20**, pp. 75-138.

McCarthy, John J. (2003b). Sympathy, cumulativity, and the Duke-of-York gambit. In Caroline Féry and Ruben van de Vijver (eds.), *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press, pp. 23-76

McCarthy, John J. (2004). Headed spans and autosegmental spreading. Ms., University of Massachusetts Amherst. [ROA-685]

McCarthy, John J., and Alan S. Prince (1990). Foot and word in Prosodic Morphology: The Arabic broken plurals. *Natural Language and Linguistic Theory* **8**, pp. 209-282.

McCarthy, John J., and Alan Prince (1993a). *Prosodic Morphology I: Constraint Interaction and Satisfaction*. Ms., University of Massachusetts Amherst and Rutgers University, New Brunswick, NJ. [ROA-482]

McCarthy, John J., and Alan Prince (1993b). Generalized Alignment. In Geert Booij and Jaap van Marle (eds.), *Yearbook of Morphology 1993*. Dordrecht: Kluwer, pp. 79-153. [ROA-7]

McCarthy, John J., and Alan Prince (1994). The Emergence of the Unmarked: Optimality in Prosodic Morphology. In Mercè Gonzàlez (ed.), *Proceedings of NELS 24*. Amherst: GLSA, pp. 333-379. [ROA-13]

McCarthy, John J. and Alan Prince (1995). Faithfulness and reduplicative identity. In Jill N. Beckman, Laura Walsh Dickey, and Suzanne Urbanczyk (eds.), *University of Massachusetts Occasional Papers in Linguistics* **18**: *Papers in Optimality Theory*. Amherst: GLSA, pp. 249-384. [ROA-103]

McCarthy, John J., and Alan Prince (1999). Faithfulness and identity in Prosodic Morphology. In René Kager, Harry van der Hulst, and Wim Zonneveld (eds.), *The Prosody-Morphology Interface*. Cambridge: Cambridge University Press, pp. 218-309.

Mester, Armin (1994). The quantitative trochee in Latin. *Natural Language and Linguistic Theory* **12**, pp. 1-61.

Noyer, Rolf (1993). Optimal Words: Towards a declarative theory of word-formation. Paper presented at ROW-1, New Brunswick, NJ.

Paster, Mary (2005). Subcategorization vs. output optimization in syllable-counting allomorphy. In John Alderete, Chung-hye Han, and Alexei Kochetov (eds), *Proceedings of the Twenty-Fourth West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla Proceedings Project, pp. 326-333.

Pater, Joe (1999). Austronesian nasal substitution and other *NC̥ effects. In René Kager, Harry van der Hulst, and Wim Zonneveld (eds.), *The Prosody-Morphology Interface*, ed., Cambridge: CUP, pp. 310-43.

Perlmutter, David M. (1998). Interfaces: Explanation of allomorphy and the architecture of grammars. In In Steven G. Lapointe, Diane K. Brentari, and Patrick M. Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications, pp. 307-338.

Picanço, Gessiane (2002) Tonal polarity as phonologically conditioned allomorphy in Mundurukú. In Julie Larson & Mary Paster (eds.), *Proceedings of the 28th Annual Meeting of the Berkeley Linguistics Society*. pp. 237-48. Berkeley, CA: Berkeley Linguistics Society, pp. 237-248.

Prince, Alan (1992). Quantitative consequences of rhythmic organization. In Karen Deaton, Manuela Noske, and Michael Ziolkowski (eds.), *Proceedings of the Parasession on the Syllable in Phonetics and Phonology*. Chicago: CLS, pp. 355-398.

Prince, Alan (2003). Arguing optimality. In Angela C. Carpenter, Paul de Lacy, and Andries W. Coetzee (eds.), *University of Massachusetts Occasional Papers in Linguistics* **26**: *Papers in Optimality Theory II*. Amherst: GLSA, pp. 269-304.

Sauerland, Uli (1995). The late insertion of Germanic inflection. Ms., MIT, Cambridge, MA.

Steriade, Donca (1999). Lexical conservatism in French adjectival liason. In J.-Marc Authier, Barbara E. Bullock, and Lisa Reed (eds.), *Formal Perspectives on Romance Linguistics : Selected Papers from the 28th Linguistic Symposium on Romance Languages (LSRL XXVIII) : University Park, 16-19 April 1998*. Amsterdam: John Benjamins, pp. 243-270.

Steriade, Donca (2000). Lexical conservatism and the notion "base of affixation". Ms., University of California, Los Angeles.

Teeple, David (2006). Lexical selection and strong parallel OT. Ms., University of California, Santa Cruz.

Tranel, Bernard (1996a). Exceptionality on Optimality Theory and final consonants in French. In Karen Zagona (ed.), G*rammatical Theory and Romance Languages*. Amsterdam: John Benjamins, pp. 275-91. [ROA-61]

Tranel, Bernard (1996b). French liason and elision revisited: A unified account within Optimality Theory. In Claudia Parodi, Carlos Quicoli, Mario Saltarelli, & Maria Luisa Zubizarreta (eds.), *Aspects of Romance Linguistics*. Washington: Georgetown University Press, pp. 433-455. [ROA-15]

Trommer, Jochen (2001). *Distributed Optimality*. Ph.D. dissertation, Universität Potsdam.

Wolf, Matthew (to appear). For an autosegmental theory of mutation. In Leah Bateman, Michael O'Keefe, Ehren Reilly, and Adam Werle (eds.), *University of Massachusetts Occasional Papers in Linguistics* **32**: *Papers in Optimality Theory III*. Amherst: GLSA. [ROA-754]

Wolf, Matthew, and John J. McCarthy (2007). Less than zero: Correspondence and the null output. In Curt Rice (ed.), *Modelling Ungrammaticality in Optimality Theory*. London: Equinox.
Yip, Moira (1998). Identity avoidance in phonology and morphology. In Steven G. Lapointe, Diane K. Brentari, and Patrick M. Farrell (eds.), *Morphology and its Relation to Phonology and Syntax*. Stanford: CSLI Publications, 216-246.
Yu, Alan C.L. (2003) *The Morphology and Phonology of Infixation*. Ph.D. dissertation, University of California, Berkeley.

Department of Linguistics
University of Massachusetts
South College, Room 226
Amherst, MA 01003 USA

mwolf [at] linguist [dot] umass [dot] edu