OT constraints are categorical

OT Constraints Are Categorical
Author(s): John J. McCarthy
Source: *Phonology*, Vol. 20, No. 1 (2003), pp. 75-138
Published by: Cambridge University Press
Stable URL: http://www.jstor.org/stable/4420242
Accessed: 25/06/2009 18:33

Phonology **20** (2003) 75–138. © 2003 Cambridge University Press
DOI: 10.1017/S0952675703004470 Printed in the United Kingdom

*OT constraints are categorical**

John J. McCarthy
University of Massachusetts, Amherst

In Optimality Theory, constraints come in two types, which are distinguished by their mode of evaluation. Categorical constraints are either satisfied or not; a categorical constraint assigns no more than one violation-mark, unless there are several violating structures in the form under evaluation. Gradient constraints evaluate extent of deviation; they can assign multiple marks even when there is just a single instance of the non-conforming structure. This article proposes a restrictive definition of what an OT constraint is, from which it follows that all constraints must be categorical. The various gradient constraints that have been proposed are examined, and it is argued that none is necessary and many have undesirable consequences.

1 Introduction

In Optimality Theory (Prince & Smolensky 1993), a constraint can assign multiple violation-marks to a candidate. This happens in two situations. First, there can be several places where the constraint is violated in a single candidate, as when ONSET assigns two marks to the form [a.pa.i]. Second, some constraints measure the extent of a candidate's deviance from some norm. For instance, the constraint ALIGN(Ft, R; Wd, R) assigns three violation-marks to $[(\text{pá.ta})_{\text{Ft}}.\text{ka.ti.ma}]_{\text{Wd}}$, one mark for each syllable that separates the right foot edge from the right word edge.

Constraints of the first type are called CATEGORICAL. The majority of OT constraints that have been proposed are categorical. Categorical constraints never assign more than one violation-mark, unless the candidate

* For their comments, criticisms, and suggestions, I am grateful to the participants in the UMass phonology seminar and reading group: Michael Becker, Della Chambless, Paul de Lacy, Kathryn Flack, Maria Gouskova, Shigeto Kawahara, Jin-Hyung Kim, John Kingston, Steve Parker, Joe Pater, Lisa Selkirk, Taka Shinya, Monica Sieh, Melissa Svendsen, Anne-Michelle Tessier, Adam Werle, Ellen Woolford and Hosuk Yoon. I am likewise grateful to audiences at the 2nd North American Phonology Conference, especially Paul Boersma, Mary Bradshaw, Stuart Davis, San Duanmu, David Odden and Bert Vaux, and at SWOT 8, especially Diana Archangeli, Mike Hammond, Junko Itô, K. P. Mohanan, Jaye Padgett, Alan Prince, Paul Smolensky and Adam Ussishkin. I have also received valuable comments on the manuscript from Maria Gouskova, Jane Grimshaw, Linda Lombardi, Nicole Nelson, Orhan Orgun, Alan Prince, Colin Wilson and Ellen Woolford. I owe a special debt to the *Phonology* associate editor and four anonymous reviewers who provided a total of 30 single-spaced pages of commentary on my initial submission. I hope this article is the better for all this assistance.

under evaluation contains more than one instance of the marked structure or the unfaithful mapping that the constraint proscribes. Constraints of the second type are called GRADIENT. Gradient constraints predominantly come from the alignment family (McCarthy & Prince 1993a), though other types of gradient constraints have been proposed (see §3). Gradient constraints can assign multiple violation-marks even when there is just one instance of a marked structure or an unfaithful mapping.[1]

In this article, I argue that OT's universal constraint component CON permits only categorical constraints. The argument for categoricality has several components. After establishing what it means for a constraint to be categorical or gradient and what this says about CON (§2), the article goes on in §3 to present a taxonomy of gradient constraints and to examine all known proposals for gradient constraints from outside the alignment family. These constraints, it will be shown, have an obvious and arguably necessary reformulation in categorical terms.

The discussion then turns to alignment constraints. In §4, morphology–prosody alignment is scrutinised. In agreement with earlier research, I show that extant morphology–prosody alignment constraints are never evaluated gradiently and in fact must not be, or else typologically unattested patterns will be predicted. The correspondence-based ANCHOR constraints (McCarthy & Prince 1995, 1999) are categorical and replace alignment in this empirical domain.

Another line of analysis, taken up in §5, looks at how alignment constraints have been applied to infixation. Standard treatments involve gradient alignment, but I present two cases where these standard treatments prove inadequate. A new set of categorical constraints on affix placement – such as PREFIX, PREFIX/σ and PREFIX/Ft – will be proposed and it will be shown that they render the gradient constraints unnecessary. Ultimately, this is an argument from Occam's Razor: since categorical constraints prove to be sufficient, there is no reason to have gradient evaluation.

The argument in §6 turns to stress, where gradient alignment constraints have been heavily exploited. The *locus classicus* of gradient alignment, directionally iterative foot-parsing, has been convincingly reanalysed by Kager (2001) in terms of categorical constraints on clashes and lapses. Thus the first goal of §6 is to review Kager's results, which not only show that gradient alignment is dispensible, but that it can be pernicious to a sound stress typology. This argument, like the one from morphology–prosody alignment, is both indirect and direct: gradient alignment is unnecessary, and its presence in CON leads to unwelcome predictions.

In the same section, I also examine other stress phenomena that initially seem to require gradient alignment: non-iterative foot-parsing and

[1] 'Gradient' has sometimes been used in senses other than the one employed here. For example: Harrikari (1999) uses the phrase 'gradient OCP' to refer to a set of OCP constraints distinguished by locality, and not to a single gradient constraint. Constraints that assess forms continuously (i.e. numerical optimisation) have also been called 'gradient'.

assignment of main stress. These are shown to have a similar basis: categorical constraints on the location of the head foot that are descendants of Prince's (1983) End Rule. Here again the argument for categoricality and against gradience comes from Occam's Razor: gradient alignment is doing no necessary work, its functions having been usurped by categorical constraints, which are needed anyway.

The last substantive section (§7) switches to autosegmental phenomena. Docking of morphemic features and tones (§7.2) arguably falls under the same rubric as infixation. Flop or reassociation processes (§7.3) exemplify the effects of categorical COINCIDE constraints (Zoll 1996). These constraints, I will argue, avoid an unwanted typological prediction of gradient alignment under ranking permutation. Finally, §7.4 looks at autosegmental spreading processes. A novel constraint is proposed that combines the properties of two current non-alignment-based approaches to spreading.

To sum up, the thesis of this article is that the known applications of gradient constraints in OT can and in many cases should be reanalysed with categorical constraints. Many of the categorical constraints that step into this role have been proposed previously; those that are novel here are independently motivated. Overall, this argument is the natural sequel to a remark by Prince & Smolensky (1993: 88): 'the division of constraints into those which are binary and those which are not ... is not in fact as theoretically fundamental as it may at this point appear'. Gradient constraints are not an essential element of OT; they are an imposition on it, as is apparent once we set out to define what it is that OT constraints do.

2 Gradient and categorical constraints

A classic OT constraint can be regarded as a function from an input (in the case of markedness constraints) or an input/output pair (in the case of faithfulness constraints) to zero or more violation-marks. To say that a constraint is gradient or categorical, then, is to say something about this function. To say that *all* constraints are categorical is to say something about OT's universal theory of constraints CON, which not only lists the constraints but can also impose restrictions on them (for discussion, see McCarthy 2002b: 17ff).

Categorical markedness constraints have been formulated in diverse ways in the literature, but most if not all can be stated in terms of a prohibited phonological constituent and a (possibly null) contextual condition under which it is prohibited:[2]

(1) *Schema for categorical markedness constraint*

$*\lambda/C \equiv$ For any λ satisfying condition C, assign a violation-mark.

[2] See Eisner (1999) and Potts & Pullum (2002) for other developments along these general lines. See McCarthy (2002a, 2003a) for another application of the notion 'locus of violation'.

The letter λ is mnemonic for the LOCUS of violation. It is the phonological constituent that the markedness constraint militates against (compare the 'focus' of a constraint in Crowhurst & Hewitt 1997). As noted in §1, categorical constraints may assign multiple violation-marks when there are multiple loci of violation in the form under evaluation. It is, then, a general fact about categorical markedness constraints that when two candidates cand1 and cand2 contain equal numbers of loci of violation of some constraint C, C assigns an equal number of violation-marks to cand1 and cand2. The principal thesis of this article is that the theory of CON limits all markedness constraints to schema (1).

Schema (1) requires a couple of remarks before we go on. First, (1) requires all markedness constraints to be formulated negatively; they are prohibitions rather than admonitions. This is consistent with generally accepted practice, and has even been argued to be necessary (de Lacy 2002). Second, certain constraints have a symmetric character that makes the choice of a locus of violation arbitrary, as Maria Gouskova and Alan Prince have pointed out. This is true, for instance, of the *LAPSE constraints in (31) below, which are violated by sequences of unstressed syllables. This arbitrariness, though, is only a problem for the analyst seeking to translate various previously proposed markedness constraints into a consistent format like (1). A theory of CON is not obliged to make that translation easy or even fully determinate.

Gradient markedness constraints, of which alignment is the principal example, cannot in general be stated within the strictures of (1). Here is a definition of alignment, expanding on McCarthy & Prince (1993a: 80), that makes the assignment of violation-marks fully explicit (cf. Ellison 1994, Zoll 1996):

(2) *Expanded schema for alignment constraint*[3]

ALIGN(Cat1, Edge1; Cat2, Edge2; Cat3) ≡

∀Cat1 if ∃Cat2, assign one violation-mark ∀Cat3 that intervenes between Edge1 of Cat1 and the nearest Edge2 of some Cat2,

where

Cat1, Cat2 are prosodic or morphological categories, Cat3 is a prosodic category and Edge1, Edge2 ∈ {Right, Left}.

This formulation makes explicit what is usually implicit in analyses that use alignment: some specific unit of distance (Cat3) is used to determine the extent of violation.[4] Reference to the *nearest* Cat2 also makes explicit

[3] The five arguments of an alignment constraint will sometimes be abbreviated when there is no danger of ambiguity. Same-edge alignment constraints may omit the Edge1 argument, and Cat3 will often be left off unless it is the focus of discussion.

[4] The identity of Cat3 is often taken for granted in applications of alignment, and so this argument is missing from the original definition. This is a mistake, since it is at least conceivable that two alignment constraints might differ only in the choice of Cat3. Nonetheless, Mester & Padgett (1994: 81) mention the possibility of predicting the identity of Cat3 from the rest of the constraint. For related discussion, see §5.2.

that which has always been assumed implicitly: for example, in a recursive structure like ... $(\sigma\sigma)_{Ft}\ \sigma]_{Wd}\ \sigma]_{Wd}$, ALIGN(Ft, R; Wd, R; σ) assigns only one violation-mark because only one syllable intervenes between $)_{Ft}$ and the nearest $]_{Wd}$.

There is obviously a very big difference between (1) and (2) in the way that violation-marks are assigned. Categorical markedness constraints assign one mark for each locus – that is, for each instance of the offending constituent. Gradient alignment constraints can and often do assign more than one violation-mark for each instance of Cat1. The marks assigned to each Cat1 are then lumped together in evaluating the entire candidate.

The classic example of gradient behaviour is ALLFTR (McCarthy & Prince 1993a, following a suggestion by Robert Kirchner). This constraint asserts that every foot should be final in the prosodic word. In terms of (2), it says ALIGN(Ft, R; Wd, R; σ). When it evaluates a candidate with several unaligned feet, it treats each foot as a separate potential locus of violation, just like a categorical constraint, but then it assesses each foot gradiently. The violation-marks accumulated from these two sources are treated homogeneously, as (3) shows.

(3) *Evaluation by* ALLFTR

	Ft-1	Ft-2	Ft-3	ALLFTR
a. $[(\acute{\sigma}\sigma)_1(\acute{\sigma}\sigma)_2(\acute{\sigma}\sigma)_3\sigma]$	*****	***	*	*********
b. $[(\acute{\sigma}\sigma)_1(\acute{\sigma}\sigma)_2\sigma(\acute{\sigma}\sigma)_3]$	*****	***	∅	********
c. $[(\acute{\sigma}\sigma)_1\sigma(\acute{\sigma}\sigma)_2(\acute{\sigma}\sigma)_3]$	*****	**	∅	*******
d. $[\sigma(\acute{\sigma}\sigma)_1(\acute{\sigma}\sigma)_2(\acute{\sigma}\sigma)_3]$	****	**	∅	******

In (3a), for instance, the three feet are misaligned by five, three and one syllables, respectively. So this candidate receives nine violation-marks from ALLFTR.

Perhaps the most remarkable thing about gradient ALLFTR and ALLFTL is that they are able to distinguish between (3b) and (3c). This property is important when other constraints rule out (3a) and (3d), so that (3b) and (3c) compete directly. As we will see in §6.2, though, empirical findings about stress typology do not support this aspect of the gradient theory (Kager 2001); the (3b)/(3c) distinction does not seem to be an authentic, independent difference between languages.

Significantly, there is no reasonable way of reproducing this apparently unnecessary distinction using the categorical schema in (1). Consider hypothetical categorical constraints against non-peripheral feet: *FT / σ __ and *FT / __ σ. These constraints do not differentiate (3b) and (3c); they assign equal marks to both. Or consider hypothetical categorical constraints against a syllable that is preceded or followed at any distance by a foot: *σ / FT ... __ and *σ / __ ... FT. Again, these constraints cannot distinguish (3b) from (3c), because both have five syllables that are preceded by some foot and five that are followed by some foot.

Several anonymous reviewers have raised an objection that goes something like this. Suppose CON supplies categorical constraints against feet that are preceded or followed by at least three syllables (cf. Karttunen 1998, who adopts a similar artifice to deal with multiple loci of violation): *FT / $\sigma\sigma\sigma$ __ and *FT / __ $\sigma\sigma\sigma$. These constraints can distinguish (3b) from (3c), since (3b) has two feet, each of which is followed by at least three syllables, but (3c) has only one foot meeting that condition. It would appear, so the objection goes, that the power of gradient ALLFTR/L has been reproduced using only a categorical schema.

There are two answers to this objection. The first is that the full power of gradient ALLFTR/L has not really been recaptured. ALLFTR/L can make similar distinctions in even longer words, but the counting constraints *FT / $\sigma\sigma\sigma$ __ and *FT / __ $\sigma\sigma\sigma$ cannot. More counting constraints could be added, but since CON is finite, it will never be possible to reproduce the full effect of ALLFTR/L. What we seek are analyses of stress systems, not analyses of individual words that we happen to encounter. Clearly, counting constraints like *FT / $\sigma\sigma\sigma$ __ do not analyse the same systems that ALLFTR/L does.[5]

An even more telling response to the reviewers' objection is that constraints like *FT / $\sigma\sigma\sigma$ __ contravene a widely assumed (though often tacit) principle of linguistic metatheory: rules and constraints are local, a requirement often expressed by saying that rules or constraints do not count beyond two in their definitions (Chomsky 1965: 55, Hayes 1995, McCarthy & Prince 1986, Nelson & Toivonen 2001). For example, no language requires the presence of at least three round vowels to initiate rounding harmony, nor do we ever find that complementisers may be doubly but not trebly filled. The impossibility of constraints like *FT / $\sigma\sigma\sigma$ __ is therefore quite independent of alignment, OT and even phonology.

In fact, the impossibility of such constraints is already implicit in the markedness constraint schema (1). The locus of violation λ is a single phonological constituent. Paul Smolensky suggests that the contextual condition C also be limited to mentioning a single phonological constituent that is separate from λ. This is a strong claim; if it proves correct, then constraints are inherently local because they can never mention more than two distinct constituents and a relation between them, such as adjacency or shared membership in a superordinate constituent.

Now consider faithfulness constraints. In correspondence theory (McCarthy & Prince 1995, 1999), the standard faithfulness constraints are inherently categorical in their assessments. For example, MAX assigns a violation-mark for each input segment without an output correspondent. DEP does the same, but with input and output transposed. IDENT(F)

[5] It might be objected that OT itself involves counting violations. As has been emphasised repeatedly (Prince & Smolensky 1993), the key notion in OT is comparison, not counting. Furthermore, this objection blurs an important distinction within OT between the constraints and EVAL. Constraints like *FT / $\sigma\sigma\sigma$ __ build counting right into their definitions; this has nothing to do with how EVAL compares candidates.

assigns a mark for each input/output segmental pair differing in the value of feature F. The less familiar UNIFORMITY and INTEGRITY, which prohibit segmental coalescence and diphthongisation respectively, are also inherently categorical: they assign one violation-mark for each segment that has multiple correspondents. I-CONTIGUITY, which prohibits internal deletion, O-CONTIGUITY, which prohibits internal epenthesis, and the ANCHOR constraints that prohibit peripheral deletion and epenthesis are contextually restricted versions of MAX and DEP, so they are categorical just as MAX and DEP are. (For more on the ANCHOR constraints, see §4.) Overall, then, these faithfulness constraints are in accordance with the categorical markedness schema (1), except that their loci of violation are mappings of input and/or output constituents, rather than output constituents themselves.

This leaves LINEARITY as the only constraint whose status *vis-à-vis* gradience is as yet unclear. LINEARITY forbids metathesis, and the intuition we seek to capture is that the non-local metathetic mapping $/\alpha\beta\gamma/ \rightarrow [\gamma\alpha\beta]$ is less faithful than its local counterpart $/\alpha\beta\gamma/ \rightarrow [\alpha\gamma\beta]$. Thus, non-local metathesis, which is notably rare (non-existent according to Poser 1982), can occur only when local metathesis is unsatisfactory.

The need to distinguish local and non-local metathesis leads Hume (1998) to describe LINEARITY as a gradient constraint (cf. Carpenter 2002), but this conclusion does not necessarily follow. The input $/\alpha\beta\gamma/$ can be regarded as asserting three linear-precedence relations: $\alpha > \beta$, $\beta > \gamma$ and $\alpha > \gamma$. Categorical LINEARITY assigns a violation-mark for each input precedence relation that the output contradicts: two marks for $[\gamma\alpha\beta]$ and one mark for $[\alpha\gamma\beta]$. Though this sort of constraint goes beyond the highly limiting schema (1), it is nonetheless categorical in its assessments.

Local constraint conjunction (Smolensky 1995) creates new constraints, markedness or faithfulness.[6] The local conjunction of constraints C1 and C2 in domain D, written $[C1\&C2]_D$, is violated if and only if both C1 and C2 are violated by the same instance of D. Given how local conjunction is defined, $[C1\&C2]_D$ is necessarily categorical, even if C1, C2 or both are gradient. Therefore, local conjunction is not a potential source of new gradient constraints and may be safely set aside.

To sum up, the claim that all OT constraints are categorical has here been reduced to the claim that all markedness constraints conform with the constraint schema (1) and all faithfulness constraints are of the standard categorical types in correspondence theory. Gradient constraints, most prominently alignment, are not compatible with (1), nor can all the effects of gradience be obtained using (1). If it proves true, as I argue below, that gradient constraints are not required in OT, then it is possible to maintain a restrictive claim about CON: all markedness constraints are based on (1), and faithfulness constraints assign at most one mark for each unfaithful mapping.

[6] For related work on local conjunction, see the references cited in McCarthy (2002b: 43).

Before venturing into the realm of the empirical, it may be necessary to clarify a limit on the goals of this article. The proposal made here is that (1) sets down a standard that all markedness constraints must meet; the proposal does not say that everything meeting this standard is an actual constraint in CON. In other words, (1) presents necessary but not sufficient conditions for valid markedness constraints. Like other constraint schemata in the literature (e.g. McCarthy & Prince 1993a, Smolensky 1995, Eisner 1999, Baković & Wilson 2000, Wilson 2000, 2001, Potts & Pullum 2002, Smith 2002), this one does not obviate the need for other formal or substantive limits on what constraints are possible.

3 Bounded gradience

Gradient constraints in the OT literature are not limited to alignment. A taxonomy of attested types of gradience is useful to organise the discussion. In (4), the known types of gradient constraints are classified according to the dimension along which violations are assessed.

(4) *Types of gradience in the OT literature*

a. *Horizontal gradience*
Assign violation-marks in proportion to distance in the segmental string. Example: ALIGN(Ft, R; Wd, R; σ), ALIGN(Pfx, L; Wd, L; Seg) (used in infixation – see §5).

b. *Vertical gradience*
Assign violation-marks in proportion to levels in a hierarchy. Example (Prince & Smolensky 1993: ch. 4, (66)): NON-FINALITY ≈ no head of Wd is final in Wd. The Wd is headed by its main-stressed foot and recursively by the head syllable of that foot. One violation-mark is assigned for each of these that is final in Wd. E.g. Latin *[a(mó:)] gets two marks and [(ámo)] gets only one.

c. *Collective gradience*
Assign violation-marks in proportion to the cardinality of a set. Example (Padgett 1995a, 2002): CONSTRAINT(Class) ≈ assign one violation-mark for each member of the feature-class Class that does not satisfy CONSTRAINT. E.g. ASSIM[Place] assigns two marks to [an͡gba], one to [aŋ͡gba] and none to [aŋ͡mg͡ba].

d. *Scalar gradience*
Assign violation-marks in proportion to the length of a linguistic scale. Example (Prince & Smolensky 1993: 16): HNUC ≡ 'a higher sonority nucleus is more harmonic than one of lower sonority', i.e. assign a nucleus one violation-mark for each degree of sonority less than the sonority of *a*.

The example given are typical of these attested types of gradience.

Alignment constraints – and apparently *only* alignment constraints – involve horizontal gradience. In other words, constraints that are horizontally

gradient always conform to something like (2), and they assess candidates differently depending on how far it is between the two constituent edges.

Vertical gradience is not encountered nearly as often as horizontal gradience. The only cases I have found use the prosodic hierarchy (Selkirk 1980) to determine the extent of violation. This is true for the version of NON-FINALITY described by Prince & Smolensky, for WEAKEDGE in Spaelti (1994) and for the EXHAUSTIVITY constraint of Selkirk (1995, 1996). (WEAKEDGE assigns a violation-mark for each prosodic category whose right-periphery is non-empty – $[((\text{dog})_{\sigma})_{\text{Ft}}]_{\text{Wd}}$ receives three marks but $[((\text{do})_{\sigma})_{\text{Ft}}\ \text{g}]_{\text{Wd}}$ (with 'final consonant extrametricality') gets only one. EXHAUSTIVITY prohibits non-strict-layering in the prosodic hierarchy; if a phonological phrase directly dominates a syllable, as in $\{\text{to}_{\sigma}\ [(\text{Bill}_{\sigma})_{\text{Ft}}]_{\text{Wd}}\}_{\text{PPh}}$, then two marks are incurred because two levels, word and foot, have been skipped.)

Collective gradience is developed formally in the context of Padgett's work on feature classes; no other cases are known to me. The idea is that markedness or faithfulness constraints referring to a feature class are gradient over the members of that class.

Scalar gradience appears in Prince & Smolensky (1993) in the form of the constraints HNUC and PKPROM. The former evaluates syllable nuclei for their sonority, assigning marks in proportion to their sonority level. The latter favours stressed syllables with greater intrinsic prominence (weight or sonority); violations are reckoned in terms of the prominence level of the stressed syllable. Other examples of scalar gradience include the markedness constraints RAISING and REDUCE in Kirchner (1996), the syllable-contact sonority constraint SYLLCONT proposed by Bat-El (1996: 302) and the faithfulness constraint MAX[+nas] in Zhang (2000: 451).

There is a basic bifurcation between horizontal gradience and the other types. Horizontal gradience is responsible for unboundedly many constraint violations even when there is a single locus of violation (i.e. a single instance of Cat1 in (2)). There is no non-arbitrary limit on how many segments, syllables or other Cat3 units can separate the constituent edges mentioned in an alignment constraint. But constraints that are vertically, collectively or scalarly gradient are always limited in how many marks they can assign to a single locus of violation. Constraints assessing vertical gradience cannot assign more violation-marks than there are levels in the hierarchy. Constraints that are collectively gradient cannot assign more violation-marks than there are members of the set. Constraints that are scalar gradient cannot assign more violation-marks than there are steps in the scale. And since the linguistic hierarchies, sets or scales referred to in these constraints are always finite, the constraints are always bounded in their assessments.

This bifurcation between the unbounded and the bounded is important because there is an obvious alternative account of bounded gradience: posit separate categorical constraints for each level of the hierarchy, each member of the set or each step on the scale, instead of one gradient constraint for all hierarchy levels or set members. The basic techniques for

doing this are introduced in Prince & Smolensky (1993: ch. 8), where HNUC is deconstructed in this way. Instead of a gradient constraint that assigns zero marks to the nucleus *a*, one mark to the nuclei *i* and *u*, two marks to the nuclei *r* and *l* and so on, there is a universally fixed hierarchy of constraints, derived from the sonority hierarchy: ... ≫ *NUC/*r,l* ≫ *NUC/*i,u* ≫ *NUC/*a*. Alternatively, HNUC can be deconstructed into a set of constraints in a subset or 'stringency' relationship (Prince 1998, de Lacy 2002): *NUC/*r,l*, *NUC/*r,l,i,u*, *NUC/*r,l,i,u,a* ... Either way, the constraints involved are strictly categorical, yet they equally well express gradient HNUC's implication that, say, *r* and *l* are worse nuclei than the vocoids.

In fact, as Prince & Smolensky show, gradient HNUC *must* be deconstructed into categorical constraints in order to account for observed syllable-structure typology. Gradient HNUC suffices in their analysis of Berber, where the descriptive problem involves deciding which of two adjacent segments should be made into a syllable nucleus. HNUC correctly favours [tzm̩t] over *[tz̩mt] 'it (FEM) is stifling'. But there is no way to use HNUC to account for the absolute absence of low-sonority nuclei in the inventories of other languages. Inventory restrictions are obtained in OT by the ranking of markedness constraints with respect to faithfulness. The relative ranking of HNUC and FAITH is uninformative; there is no way to use these two constraints to specify that English allows vocoid, liquid and nasal nuclei, but Spanish allows only vocoids. With deconstructed HNUC, though, this distinction is easy: in English, FAITH dominates *NUC/*m,n*, but in Spanish FAITH is dominated by *NUC/*r,l*. Gradient HNUC is therefore inadequate on typological grounds. But once the categorical *NUC constraints are introduced to solve the typology problem, gradient HNUC is superfluous, even in Berber. We are therefore free (in fact, obliged by Occam's Razor) to remove gradient HNUC from CON.

Replacing gradient HNUC with categorical constraints is possible because the sonority scale is finite. (See also Ellison 1994: 1008 on 'constraints which use a finite alphabet of marks'.) The other boundedly gradient constraints can also be replaced by categorical constraints for the same reason. The vertically gradient version of NON-FINALITY in Prince & Smolensky (1993) can be replaced by separate constraints requiring non-finality of the main-stressed foot (NON-FINALITY(Ft)) and non-finality of the main-stressed syllable (NON-FINALITY($\acute{\sigma}$)), as Prince & Smolensky themselves suggest in a footnote on the same page. That this should be done is shown by a typological argument (Gouskova 2003): in Hopi, where the final syllable is unstressed but footed, only NON-FINALITY($\acute{\sigma}$) is active, whereas Latin shows activity by NON-FINALITY(Ft). EXHAUSTIVITY can also be deconstructed into categorical constraints, one for each level of the prosodic hierarchy. In fact, it standardly is deconstructed, since PARSE-σ ('every syllable belongs to a foot') is just the foot-syllable version of EXHAUSTIVITY.

As for collective gradience, any constraint that is gradient over a set of elements can be replaced by individual constraints on each member of the

set. For example, the gradient constraints IDENT[colour] and SPREAD [colour] in Padgett (2002) are superfluous if there are categorical IDENT constraints for each of the vowel-colour features [back] and [round]. If CON includes categorical IDENT[round], IDENT[back], SPREAD[round] and SPREAD[back], then gradient IDENT[colour] and SPREAD[colour] are superfluous because their presence will have no visible effect on the resulting factorial typology: a grammar with the ranking ⟦IDENT[colour] ≫ SPREAD[colour]⟧ cannot be distinguished empirically from a grammar with the ranking ⟦IDENT[round], IDENT[back] ≫ SPREAD[round], SPREAD[back]⟧. Pursuing a suggestion by Alan Prince, Padgett (1995a) briefly considers eliminating IDENT[round] and IDENT[back] from CON, but it has not been shown that this move is possible in this specific case or generally. (Backing harmony in Finnish presents obvious problems analogous to the difficulty that HNUC encounters in distinguishing English from Spanish.) More recently, Padgett (2002: 82) has explicitly allowed constraints to refer both to classes and to individual features. But, as was just shown, the gradient, class-referring constraints are unnecessary if there are similar categorical constraints on the individual features.

To sum up, I have pointed out a distinction between boundedly and unboundedly gradient constraints. Boundedly gradient constraints have a straightforward and arguably necessary translation into categorical constraints. Instead of a single constraint that can be violated more than once, several categorical constraints are posited, one for each member of the scale, hierarchy or set over which violation is computed. This move is supported by typological arguments: the gradient constraint is able to express preferences in case of conflict, but it has no way of forbidding some members of the scale, hierarchy or set while permitting others. The boundedly gradient constraint therefore turns out to be insufficient and superfluous. Categorical constraints, consistent with the schema in (1), are necessary and they are enough.

Unbounded, horizontal gradience, on the other hand, cannot be translated into a finite set of categorical constraints. (That the constraint set in CON is finite is a basic assumption underlying the notion of factorial typology; Prince & Smolensky 1993.) For example, ALLFTR imposes a harmonic ordering of unlimited depth on words containing just a single foot: $[\ldots(\sigma\sigma)] \succ [\ldots(\sigma\sigma)\sigma] \succ [\ldots(\sigma\sigma)\sigma\sigma] \succ [\ldots(\sigma\sigma)\sigma\sigma\sigma] \succ \ldots$ No finite set of categorical constraints can duplicate this order. I will show that unbounded, horizontal gradience can also be eliminated, but the argument, which is developed in the subsequent sections of this article, is necessarily more complex than for bounded gradience.

4 Morphology–prosody alignment constraints

Constraints aligning the edges of morphological and prosodic constituents have been around since the beginning of OT. In fact, the very first alignment constraint to be proposed, ALIGN in Prince & Smolensky's (1993)

analysis of Lardil, says that the right edge of the stem must coincide with the right edge of a syllable. Similar constraints demand alignment of root or stem edges with the edges of prosodic words (McCarthy & Prince 1993a).

Interestingly, morphology–prosody alignment constraints are, in actual practice, never evaluated gradiently, an observation due to Merchant (1995). Indeed, it can be shown that many morphology–prosody alignment constraints *must not* be evaluated gradiently, or else incorrect results are predicted. There is, then, a measure of arbitrariness in the treatment of alignment constraints in the literature: some, like those affecting stress, are evaluated gradiently, but others are not. Requiring all constraints to be categorical, as proposed here, will eliminate this arbitrariness.

An example of categorical alignment comes from the analysis of Axininca Campa (Payne 1981, Spring 1990, McCarthy & Prince 1993a, b). This language shows visible activity by the different-edge alignment constraint ALIGN(Sfx, L; Wd, R), dubbed SFX-TO-WD. Through interaction with other constraints, SFX-TO-WD ensures that roots that are less than two moras long are augmented with epenthetic *ta* when they occur before consonant-initial suffixes:

(5) *Augmentation in Axininca Campa*

	/na-piro-ānc^{h}i/	na**ta**pirotānc^{h}i	'carry on shoulder +
cf.	/na-ānc^{h}i/	natānc^{h}i, *na**tat**ānc^{h}i	VERITY + INF'

When SFX-TO-WD is satisfied, a suffix like *-piro* is immediately preceded by a prosodic word: [na**ta**]$_{Wd}$-pirotānc^{h}i. The prosodic word must contain a foot to serve as its head (§6.3), and that foot must be binary to satisfy FTBIN (Prince 1980, Broselow 1982, Hayes 1995, McCarthy & Prince 1996). Ranked above the faithfulness constraint DEP, SFX-TO-WD and FTBIN compel augmentation of roots like *na*, which cannot support a binary foot unaided.

The interesting situation arises when the same root appears before a vowel-initial suffix. As [natānc^{h}i] shows, there is no augmentation, just epenthesis of ONSET-satisfying *t*. Yet if SFX-TO-WD were evaluated gradiently, augmentation would ensue, and *[na**tat**ānc^{h}i] would be the outcome. Tableau (6) shows why:

(6) *Wrong augmentation with gradient* SFX-TO-WD

/na-ānc^{h}i/	ONSET	SFX-TO-WD	DEP
☞ a. [nat‸ānc^{h}i]		****!	*
☞ b. [[(na**ta**)]**t**‸ānc^{h}i]		*	***
c. [[(na**ta**)]‸ānc^{h}i]	*!		**

To assess SFX-TO-WD gradiently, it is necessary to determine the distance in segments between the left edge of the suffix, indicated by '‸', and the

nearest right prosodic-word edge '] ' (see (2)). Perfect satisfaction of SFX-TO-WD is ruled out by top-ranking ONSET, leaving the choice to the best candidate among those that violate SFX-TO-WD. That candidate is *[[(na**ta**)]tānc^{h}i], since its alignment disparity is just the single segment *t*. The actual winner [na**t**ānc^{h}i] fares worse on gradient SFX-TO-WD.

If SFX-TO-WD is a categorical constraint, however, then the outcome is correct. As a categorical constraint, it requires that each suffix be immediately preceded by $]_{Wd}$. Both (6a) and (6b) have a single locus of categorical SFX-TO-WD violation: each has a suffix that is not immediately preceded by a $]_{Wd}$. So both candidates receive a single mark from SFX-TO-WD, a tie, which leaves the choice up to lower-ranking DEP. It resists augmentation, awarding the honours to [na**t**ānc^{h}i]. It is therefore crucial to the analysis of Axininca Campa that SFX-TO-WD be a categorical constraint.

Another example of this type comes from the phonology of Makassarese (Aronoff *et al.* 1987, McCarthy & Prince 1994). In this language, word-final codas are limited to [ʔ] and [ŋ]. Roots ending in any other consonant receive an epenthetic copy of the preceding vowel, plus a final [ʔ]:

(7) *Epenthesis in Makassarese*

/rantas/	rántas**a**ʔ	'dirty'
/tetter/	téttere**ʔ**	'quick'
/jamal/	jámala**ʔ**	'naughty'

That the final [Vʔ] sequence is indeed epenthetic is shown by the antepenultimate stress, because stress falls on the penult in words without epenthesis, and by the absence of the epenthetic segments in suffixed forms like /tetter-aŋ/ → [tettéraŋ] 'quicker'.

Vowel epenthesis is a straightforward indication that CODACOND dominates DEP. Epenthesis of [ʔ] shows the action of a less familiar markedness constraint, FINALC, which prohibits vowel-final prosodic words. It too is ranked above DEP, as tableau (8) shows:

(8) CODACOND, FINALC ≫ DEP *in Makassarese*

/rantas/	CODACOND	FINALC	DEP
☞ a. rantas**a**ʔ			**
b. rantas	*!		
c. rantas**a**		*!	*

But with FINALC ranked above DEP, underlying vowel-final words should also get epenthetic [ʔ]. This is incorrect, as shown by forms like /lompo/ → [lompo], *[lompo**ʔ**] 'big'. We have here a ranking paradox: [rantas**a**ʔ] requires FINALC ≫ DEP, but [lompo] requires DEP ≫ FINALC.

This paradox leads McCarthy & Prince (1994) to propose that alignment is what blocks [ʔ]-epenthesis in [lompo]. Ranked above FINALC, ALIGN (Stem, R; σ, R) blocks epenthesis in [lompo]; ranked below CODACOND,

it does not block epenthesis in [rantas**aʔ**]. Tableau (9) shows how this analysis works:

(9) CODACOND ≫ ALIGN(Stem, R; σ, R) ≫ FINALC ≫ DEP *in Makassarese*

	/rantas/	CODACOND	ALIGN(St,σ)	FINALC	DEP
a.	☞ i. rantas**aʔ**		*		**
	ii. rantas	*!			
	iii. rantas**a**		*	*!	*
b.	/lompo/				
	☞ i. lompo			*	
	ii. lompo**ʔ**		*!		*

Here, as in Axininca, it is crucial that the morphology–prosody alignment constraint be evaluated categorically. The candidates [rantas**aʔ**] (9a) and *[rantas**a**] (9c) must tie on alignment, so as to leave the choice up to FINALC. If alignment were assessed in the expected gradient fashion, then better-aligned *[rantas**a**] would wrongly win.

These examples and others (Merchant 1994, 1995, Noske 1999, Walker 2002) show that known cases of morphology–prosody alignment do not involve gradient evaluation; they must be treat categorically. To clinch the argument, I present here a hypothetical example where gradient morphology–prosody alignment predicts an unattested and presumably impossible phonological system.

Imagine a language where ONSET dominates ALIGN(Stem, L; Wd, L). Assume, too, that recursion of the category Wd is permitted because NON-REC(Wd) is low-ranked (Selkirk 1995, 1996). Now, consider the effect of joining a CVCVC-prefix to a vowel-initial root in this hypothetical language (which resembles Italian – Peperkamp 1997: 81).

(10) *Gradient alignment (hypothetical)*

CVCVC-[VCVCV]$_{Stem}$	ONSET	ALIGN(St,Wd)	NON-REC(Wd)
☞ a. [$_{Wd}$CVCV[$_{Wd}$C-VCVCV]]		*	*
☞ b. [$_{Wd}$CVCVC-V[$_{Wd}$CVCV]]		*	*
c. [$_{Wd}$CVCVC-VCVCV]		*****!	
d. [$_{Wd}$CVCVC-[$_{Wd}$VCVCV]	*!		*

Perfect alignment is impossible because of high-ranking ONSET. If ALIGN is enforced gradiently, then the winner is (10a) or (10b), which misalign the stem and the prosodic word by a single segment and thereby triumph over the grossly misaligned (10c). If, however, ALIGN is enforced categorically, then (10a, b) tie with (10c), leaving the choice up to NON-REC(Wd), which favours the simpler non-recursive structure of (10c).

In general, as (10) shows, gradient alignment predicts the existence of prosodic constituents that come close to but don't perfectly align with

morphological constituents. For example, the highlighted consonant in [$_{Wd}$CVCVC-V[$_{Wd}$**C**VCV]] (10b) would be expected to show the phonology of a word-initial consonant, even though it is root-internal. But no such evidence has been found, and unless it is, this typological prediction of gradient alignment is not supported by the facts.[7] The examples discussed in this section and in the literature cited show, on the contrary, that morphology–prosody misalignment is all or nothing; there is no advantage to being only slightly misaligned. Gradient alignment of prosodic and morphological constituents predicts a dubious language typology, but if the same constraints are evaluated categorically, no typological problems ensue.

This argument also shows that gradient alignment of morphological and prosodic constituents cannot simply be augmented with categorical alignment. The unwanted prediction in (10) is not avoided merely by positing categorical alignment constraints; rather, it requires the elimination of gradient alignment from CON. Gradient alignment at the morphology–prosody interface is not only superfluous, but wrong.

If, as I have argued, constraints requiring coincidence of morphological and prosodic edges are categorical, then it makes very little sense to keep calling them alignment constraints. In fact, the ANCHOR constraints of McCarthy & Prince (1995, 1999) are a correspondence-based categorical replacement for gradient alignment of morphological and prosodic constituents. They are correspondence-based because they relate grammatical structure, which is normally regarded as a property of inputs, to prosodic structure, which is reliably present only in outputs.

To account for the full range of morphology–prosody alignment effects, two ANCHOR constraints are required. The first, ANCHOR proper, is defined in (11). It accounts for same-edge alignment cases, such as Makassarese. C_I stands for a morphological constituent in the input, and C_O stands for a prosodic constituent in the output. The correspondence relation is $\mathfrak{R}$.

(11) ANCHOR(C_I, C_O, E)

If x = Edge(C_I, E) and y = Edge(C_O, E) then x$\mathfrak{R}$y.

'Any element at the designated edge of C_I has a correspondent at the same edge of C_O.'

Definition: Edge(X, {L, R}) ≡ the segment standing at the L/R edge of X.

In other words, the segment that begins (or ends) the input morphological constituent C_I must stand in correspondence with the segment that begins (or ends) the output prosodic constituent C_O. ANCHOR constraints may be substituted for any same-edge constraint on the alignment of morphological and prosodic constituents. For example, ALIGN(Stem, R; σ, R) in (9) can and should be replaced by the categorical constraint ANCHOR(Stem, σ, R).

[7] Peperkamp (1997: 81) assumes that a structure analogous to (10a, b) is correct for Italian, but she doesn't defend this assumption or consider candidates analogous to (10c).

The subcategorisational or different-edge alignment constraints, like SFX-TO-PRWD in Axininca, are replaced by another type of ANCHOR constraint:

(12) D-ANCHOR(C_I, C_O, E)

If $x = \text{Edge}(C_I, E)$ and $y = \text{Edge}(C_O, \bar{E})$, then $x\Re x'$ and x' is immediately adjacent to y.

'Any element at the designated edge of C_I has a correspondent that is adjacent to an element at the opposite edge of C_O.'

So, for example, the alignment constraint SFX-TO-PRWD is replaced by ANCHOR(Sfx, Wd, L).

Constraints based on these schemata are inherently categorical, in the same way that faithfulness constraints in general are categorical. ANCHOR is satisfied or not, depending on whether the required correspondence relation exists. Likewise, D-ANCHOR is satisfied or not, depending on whether the required correspondence and adjacency relations exist. That these constraints should be preferred to alignment follows from the overall argument of this section: requirements that the edges of morphological and prosodic constituents coincide or be adjacent are, in known cases, always enforced categorically, and gradient evaluation leads to implausible predictions about language typology.

5 Alignment and infixation

5.1 Statement of the problem

The theory of infixation originally proposed by Prince & Smolensky (1991, 1993: 33ff) holds that infixes are imperfect prefixes or suffixes – imperfect because the constraints aligning them peripherally, ALIGN(Pfx, L; Wd, L; Seg) and ALIGN(Sfx, R; Wd, R; Seg), are crucially dominated and may be violated. (These constraints may in fact align affixes with the stem rather than the word; I will ignore this detail in what follows.) Often, familiar markedness constraints like ONSET or NOCODA are responsible for non-peripheral placement of an affix.[8]

[8] Applications and extensions of this idea appear in McCarthy & Prince (1993a, b), Noyer (1993), Akinlabi (1996), Urbanczyk (1996), Buckley (1997), Fulmer (1997), Spaelti (1997), Boersma (1998), Carlson (1998), de Lacy (1999), Stemberger & Bernhardt (1999) and McCarthy (2000a). For a different approach, which replaces alignment with faithfulness, see Horwood (to appear).

The existence of constraints like ALIGN(-*um*-, Wd, L) is sometimes offered as proof that OT has language-particular constraints. This point is somewhat jesuitical. ALIGN(Pfx, Wd, L) and ALIGN(Sfx, Wd, R) offer a universal framework for stating constraints on affix placement. That individual affixes must somehow be identified as prefixes or suffixes on a language-particular basis comes as no surprise. A real 'language-particular constraint', if any exist, would presumably have the character of the language-particular rules in other theories: a one-time ad hoc statement with no typological commitments whatsoever.

For example, in Prince & Smolensky's analysis of Tagalog, infixation of the actor-focus morpheme *-um-* is attributed to a constraint hierarchy where NoCoda crucially dominates gradient Align(*-um-*, L; Wd, L). This ranking leads to less-than-perfect alignment with consonant-initial words like [s**um**ulat] 'to write' or [pr**um**eno] 'to brake'. The tableau in (13) shows how infixation is achieved:

(13) *Infixation with gradient alignment*

/um-preno/	NoCoda	Align(*-um-*,Wd,L)
☞ a. pr**um**eno		**
b. **um**preno	*!	
c. p**um**reno	*!	*
d. pren**um**o		****!

The gradience of alignment is called on to decide in favour of (13a) [pr**um**eno] over (13d) *[pren**um**o]. The prefix *-um-* is infixed no more than is necessary to optimise performance on NoCoda. Since [pr**um**eno] and *[pren**um**o] tie in their NoCoda performance, the better-aligned one wins.

When fuller and more exact data from Tagalog are considered, however, further issues are disclosed. This evidence comes from Orgun & Sprouse (1999: 203ff), though I depart from them in including initial [ʔ] in the analysis.[9] The data are given in (14).

(14) *Infixation in Tagalog*

a. C-*initial words*

sulat	s**um**ulat	'to write'
ʔabot	ʔ**um**abot	'to reach for'

b. CC-*initial words*

preno	pr**um**eno ~ p**um**reno	'to brake'
gradwet	gr**um**adwet ~ g**um**radwet	'to graduate'

c. m/w-*initial words*

mahal	*m**um**ahal	'to become expensive'
walow	*w**um**alow	'to wallow'

d. s + m/w-*initial words*

smajl	*s**um**majl[10] ~ *sm**um**ajl	'to smile'
swiŋ	s**um**wiŋ ~ *sw**um**iŋ	'to swing'

[9] That is, Orgun & Sprouse (1999), like Prince & Smolensky, transcribe 'to reach for' in (14a) as [abot] and [**um**abot]. This is consistent with Tagalog orthographic practice, but not with the actual pronunciation (Schachter & Otanes 1972: 26). Since OT constraints evaluate output forms, the initial [ʔ] in these words cannot properly be disregarded. See also Boersma (1998: 198) and Halle (2001: 156) on this point.

[10] The sequence [**um**mV] is excluded by a general prohibition against geminate *m* (Orgun & Sprouse 1999: 206, n. 11).

There are no surface vowel-initial words in Tagalog. When a word begins with a single consonant, *-um-* is infixed after that consonant, unless the word begins with a labial sonorant, in which case the verb has no *-um-* form.[11] With cluster-initial roots, *-um-* is, for at least some speakers, variably infixed after the first or the second consonant. When the initial cluster contains a labial sonorant, then forms with [m**um**V] and [w**um**V] sequences are again blocked, just as they are in the *m-* and *w*-initial roots.

Tagalog permits codas and complex onsets,[12] so the ranking in (15) can be safely assumed.

(15) *Some initial rankings for Tagalog*

DEP(V), MAX(C) ≫ NOCODA, *COMPLEXONS

Though they cannot compel unfaithfulness to the input because of (15), NOCODA and *COMPLEXONS play a role in analysing the [pr**um**eno] ~ [p**um**reno] variation. Orgun & Sprouse (1999) propose that these two constraints are formally tied, with one ranking or the other chosen randomly at EVAL time.[13] Thus, [pr**um**eno] ~ [p**um**reno] differ by trading better performance on one of these constraints for better performance on the other, as shown in (16).

(16) NOCODA *and* *COMPLEXONS *as tied constraints*

a. NOCODA ≫ *COMPLEXONS

/um-preno/	NOCODA	*COMPLEXONS
☞ i. pr**um**eno		*
ii. p**um**reno	*!	

b. *COMPLEXONS ≫ NOCODA

/um-preno/	*COMPLEXONS	NOCODA
i. pr**um**eno	*!	
☞ ii. p**um**reno		*

[11] Since only a fraction of all verbs are lexically marked to take *-um-* as their actor-focus marker, and since there are other actor-focus markers like *ma-*, *mag-* and *maŋ-*, it is no loss for a verb to be blocked from having an *-um-* form for phonological reasons. See Schachter & Otanes (1972: 284ff).

[12] Complex onsets may be permitted only initially; there is some reason to think that the same clusters are heterosyllabic word-medially. Schachter & Otanes (1972: 29) cite the word [libró] 'book' as evidence that 'the preference for short vowels in closed syllables is reflected in the pronunciation of certain loan words ... in which a vowel that is stressed in the language of origin is short in the Tagalog borrowing'. In short, this word is syllabified [lib.ró].

[13] There is a large body of work applying the idea of partially ordered or tied constraints to problems of phonological variation. For references, see McCarthy (2002b: 233).

If these constraints are formally tied in the grammar of Tagalog, and if a specific ranking is chosen at each application of EVAL, then the observed variation can be obtained.

The real focus of Orgun & Sprouse's analysis, however, is the role of labial sonorants in blocking *-um-* affixation. They propose a constraint, here called OCP[labial], that forbids sonorant labials in successive onsets. Most of the starred forms in (14c, d) violate this constraint: *[m**um**ahal], *[w**um**alow], *[sm**um**ajl], *[sw**um**iŋ]. They argue that merely ranking OCP[labial] among the other constraints is insufficient to block *-um-* affixation entirely with such words. Instead, OCP[labial] is promoted, on a language-particular basis, to a new grammatical component called CONTROL. The control component applies to the output of EVAL, blocking some candidates that EVAL has judged as optimal. Thus, constraints in the CONTROL component are inviolable and can cause derivations to crash. Their analysis is that EVAL proper emits *[m**um**ahal] as the most harmonic form, but then the derivation crashes when OCP[labial] sees *[m**um**ahal] in the CONTROL component.

Orgun & Sprouse's argument for enriching OT in this way comes from the impossibility of deeper infixation to satisfy OCP[labial]. The problem is that *[m**um**ahal]'s violation of OCP[labial] can be avoided by moving the infix further away from the initial [m], as in *[mah**um**al] or *[maha**lum**]. In a conventional OT analysis, without the CONTROL component, *[mah**um**al] should be fine because it violates only low-ranking ALIGN (*-um-*, Wd, L). The CONTROL component sidesteps this issue: the problematic candidate *[mah**um**al] gets no benefit from satisfying OCP[labial] because it has already lost in the EVAL component by virtue of its poor alignment.

Orgun & Sprouse hint, however, that the special, post-EVAL application of OCP[labial] could be avoided 'if ALIGN were supplemented with a constraint limiting *-um-* to the first syllable' (Orgun & Sprouse 1999: 207), a move they reject on the grounds that 'it clearly is not in the spirit of the alignment approach to infixation'. This critique seems apt if gradient alignment is *supplemented with* a categorical constraint, but not if it is *replaced by* a categorical constraint, as I will argue shortly. But first, I will present some necessary theoretical background to the reanalysis of Tagalog.

5.2 Categorical constraints on affix position

The gradient alignment constraints that have been applied to infixation in Tagalog and other languages can be replaced by categorical constraints. This section introduces these constraints and the following sections apply them.

One way to look at affix position in categorical terms is to require that the affix lie within a certain specified distance from the word periphery. If prefixation is exact, then the affix and beginning of the word coincide exactly (17a). Less exact prefixation – that is, infixation – might satisfy the requirement that there are no syllables to the left of the affix (17b).

Conceivably, an infix might be allowed to migrate inward by as much as a syllable but less than a foot (17c).

(17) *Categorial constraints on affix position*

a. PREFIX(-*af*-)

Wd

*-*af*- / Seg _ i.e. -*af*- is not preceded by a segment within the prosodic word

b. PREFIX/σ(-*af*-)

Wd

*-*af*- / σ _ i.e. -*af*- is not preceded by a syllable within the prosodic word

c. PREFIX/FT(-*af*-)

Wd

*-*af*- / Ft _ i.e. -*af*- is not preceded by a foot within the prosodic word

These constraint formulations assume the categorical schema (1) and some additional notational conventions. The category label Wd and the lines indicating constituent membership should be understood as saying that Wd dominates both *seg* and -*af*- in (17a), and furthermore that there is no other Wd that dominates either *seg* or -*af*- but not both. (This is roughly equivalent to the 'nearest Edge2 of some Cat2' clause in the alignment definition (2).) In addition, joint membership in the Wd constituent is enough; for example, it is not intended that σ and -*af*- are necessarily adjacent for (17b) to be violated, only that some σ precede -*af*- within Wd.

For any affix -*af*-, there will be the full suite of constraints in (17), if -*af*- is a prefix, or the SUFFIX counterparts of (17), if -*af*- is a suffix. Similar constraints exist for morphemes that are affixed to prosodic constituents like the head foot, rather than the word (see §5.4).

The PREFIX constraints form a stringency hierarchy in the sense of Prince (1998): violation of (17c) entails violation of (17b) entails violation of (17a). (This presupposes, as an anonymous reviewer points out, that the headedness requirement on prosodic constituents is wired into GEN, so that any constituent at level n is guaranteed to contain at least one constituent at level $n-1$.) Prince shows that constraints in a stringency relation never conflict, so they are never directly rankable. They can be ranked indirectly, through transitivity, as will be shown in (19).

These constraints build the unit of violation into the definition of the constraint, as has sometimes been assumed for gradient alignment (see (2) above and §5.3 below). But they operate categorically: the locus of violation is the prefix, and so none can assign more marks than there are prefixes in the form under evaluation. The distance between prefix and word edge

is relevant only to determining whether or not the constraint is violated, not how much it is violated.

The constraints in (17) will reappear in §7.2, when we examine the phonology of floating feature or tone morphemes.

5.3 Infixation in Tagalog

The categorical constraints on affix position, specifically (17b), permit an analysis of Tagalog that does not require Orgun & Sprouse's novel CONTROL component. The idea is that the infix can be misaligned by one or more segments, because PREFIX(-*um*-) is crucially dominated, but it cannot be misaligned by one or more syllables, as in *[mah**um**al], because PREFIX/σ(-*um*-) is undominated.[14]

The ranking of PREFIX/σ(-*um*-) is shown by candidates like *[m**um**ahal] and *[mah**um**al], but before analysing them, we need some background about how absolute ill-formedness is standardly addressed in OT. Within classic OT, which has no CONTROL component, a candidate can only lose because some other candidate wins. The ill-formedness of *[m**um**ahal] and *[mah**um**al], then, is an indication that some other candidate is favoured by the grammar. To address situations like this, Prince & Smolensky (1993: 48ff) hypothesise that the NULL OUTPUT is a member of every candidate set. In the context of their representational assumptions and the phenomena they were analysing, the null output was called the null parse, and it consisted of a segmental string without prosodic structure. In terms of Correspondence Theory (McCarthy & Prince 1995, 1999), the null output can be thought of as a candidate whose correspondence relation to the input is undefined.[15] I will use the symbol '⊙' to stand for this candidate. It is the candidate that beats *[m**um**ahal] and *[mah**um**al].

No matter what the input, the candidate ⊙ is among those emitted by GEN. Moreover, ⊙ is a surprisingly attractive candidate because it is as unmarked as can be. It vacuously satisfies every markedness constraint in CON. Markedness constraints either militate against the presence of

[14] In proposing a categorical approach to Tagalog infixation, I have been anticipated by Boersma (1998: 196–200). Boersma proposes a family of *SHIFT constraints defined as follows:

(i) *SHIFT($f : t ; g : u ; d$) ... A pair of contours (edges) at times t and u, defined on two perceptual tiers f and g and simultaneous in their specification, are not further apart in the output (if they occur there) than by any positive distance d.

[15] To be specific, suppose that the correspondence relation maps every segment of the input to one or more segments of the output, or otherwise to the empty string e. Mappings of input segments to e violate MAX; all other mappings obey it. The null candidate has no phonological content and no mappings from the input, because the correspondence relation is undefined. Therefore, it vacuously satisfies MAX, unlike a candidate with one or more true segmental deletions. For further development and applications of the null output as a candidate, see the references cited in McCarthy (2002b: 230).

structure – like NoCoda – or they require structure, when present, to have certain properties – like Onset or many alignment constraints. Since ⊙ has no structure whatsoever, it is never in danger of violating either kind of markedness constraint. Furthermore, because its input–output correspondence relation is undefined, ⊙ vacuously satisfies all faithfulness constraints. (Faithfulness constraints are defined on correspondence relations; if the correspondence relation of some candidate is undefined, then no faithfulness constraint can possibly be violated. See note 15.) By assumption, ⊙ violates just one constraint, which Prince & Smolensky call MParse.

To be specific, the constraint MParse(-*um*-) is violated by the candidate ⊙ whenever the input contains the morpheme -*um*-. Since verbs with -*um*- do sometimes have codas or complex onsets, we can infer that MParse(-*um*-) dominates NoCoda and *ComplexOns (see (18a)). Furthermore, since -*um*- is misaligned by one or more segments, we can conclude that MParse(-*um*-) also dominates Prefix(-*um*-) (see (18b)).[16]

(18) a. MParse(-*um*-) ≫ NoCoda = *ComplexOns

/um-preno/	MParse(-*um*-)	NoCoda = *ComplexOns
☞ i. p**um**.re.no		*
☞ ii. pr**u.me**.no		*
iii. ⊙	*!	

b. MParse(-*um*-) ≫ Prefix(-*um*-)

/um-sulat/	MParse(-*um*-)	Prefix(-*um*-)
☞ i. s**u.m**u.lat		*
ii. ⊙	*!	

These ranking arguments exemplify what Legendre *et al.* (1998: 257, n. 9) call a 'harmony threshold' that is set by MParse. Because ⊙ obeys every constraint except MParse, no winning candidate derived from an input with -*um*- can violate any constraint ranked higher than MParse. Therefore, all constraints that words with -*um*- are observed to violate must be ranked below MParse.

The harmony threshold works to our advantage when it comes to dealing with the effects of OCP[labial]. Because /um-mahal/ maps most harmonically to ⊙, all non-null candidates derived from this input must violate constraints ranked higher than MParse(-*um*-). This includes not only OCP[labial], to rule out *[m**um**ahal], but also Prefix/σ(-*um*-), to rule out *[mah**um**al].

[16] In tableau (18a), the ' = ' symbol and the absence of a vertical line indicate that two constraints are formally tied.

(19) OCP[labial], PREFIX/σ(-*um*-) ≫ MPARSE(-*um*-)

/um-mahal/	OCP[lab]	PREFIX/σ(-*um*-)	MPARSE(-*um*-)	PREFIX(-*um*-)
☞ a. ⊙			*	
b. m**u.m**a.hal	*!			*
c. ma.h**u.m**al		*!		*

This tableau shows a key result. We know from (18) that MPARSE(-*um*-) dominates PREFIX(-*um*-), since otherwise -*um*- would never be infixed. To this, (19) adds the information that PREFIX/σ(-*um*-) dominates MPARSE. Therefore, MPARSE separates the two PREFIX constraints in the hierarchy. This shows that they must indeed be separate constraints, as proposed in §5.2.[17] (To complete the argument at the level of analytic detail, it is also necessary to consider dissimilated candidates like *[m**un**ahal], which show that IDENT[Place] dominates MPARSE(-*um*-).)

This categorical approach is usually regarded as incompatible with gradient alignment theory, whence Orgun & Sprouse's argument for a post-EVAL check by OCP[labial]. If there is a single gradient alignment constraint ALIGN(-*um*-, L; Wd, L), then it must either dominate MPARSE (-*um*-) or be dominated by it. Either way, the wrong result is obtained.

Gradient alignment theory could be modified to achieve similar results by building the counting unit into the definition of the constraint. This is, in fact, the implication of the Cat3 argument in (2). If two otherwise identical gradient alignment constraints can differ only in the quantum of violation, as (2) implies, then gradient ALIGN(-*um*-, L; Wd, L; σ) can be ranked above MPARSE and gradient ALIGN(-*um*-, L; Wd, L; Seg) can be ranked below it. This move might be seen as the easiest answer to vexed questions about how to count violations of gradient constraints: for every gradient constraint, there are several versions distinguished solely by the counted unit.

If Tagalog is to be analysed within the strictures of standard input/GEN/EVAL/output OT, then either categorical PREFIX/σ(-*um*-) or gradient ALIGN(-*um*-, L; Wd, L; σ) is needed. The categorical constraints are also sufficient for Tagalog, as shown in (19) above and (22)–(24) below. The enriched gradient alignment theory may work in Tagalog, but the gradience part of it plays no actual role. Categorical constraints are needed anyway; their existence in OT is not in doubt. Since categorical constraints are also sufficient, as I have argued here and elsewhere in this article, then standard Occamite reasoning demands that gradient constraints be eliminated.

It remains only to clear up a few remaining points about Tagalog and to show the efficacy of the entire analysis before moving on to another example. If deep infixation *à la* *[mah**um**al] is not an option, then why not

[17] As Klein (2002: 9–10) points out, PREFIX(-*um*-) is never visibly active in Tagalog. But this scarcely supports his conclusion that it can be dropped from the analysis. By a central premise of OT, constraints may be low-ranked, but they are never literally absent from the grammar of any language.

skip infixation entirely with such words, opting for *[ʔ**um**mahal] or *[ʔ**um**walow]? There is a local explanation for the ill-formedness of *[ʔ**um**mahal] – *mm* clusters aren't allowed (see note 10) – but there is no such explanation for *[ʔ**um**walow]. In fact, we know that *-um-* words specifically *can* contain *mw* clusters because of examples like [s**um**wiŋ]. So *[ʔ**um**walow] must be ruled out for another reason: its epenthetic initial consonant.

(20) ONSET, DEP(C) ≫ MPARSE(*-um-*)

/um-walow/	ONS	DEP(C)	MPARSE(*-um-*)
☞ a. ⊙			*
b. **um**.wa.low	*!		
c. ʔ**um**.wa.low		*!	

Because no surface from of Tagalog violates ONSET, we can safely conclude that it is undominated. This tableau shows that DEP(C) is also high-ranked, crucially dominating MPARSE(*-um-*). This forecloses the last way that /um-walow/ could map to a non-null output. (There will be a bit more to say at the end of this section about the treatment of ONSET violators.)

To sum up, we have seen evidence for the following ranking in Tagalog:

(21) *Tagalog ranking summary*

ONSET, OCP[labial], PREFIX/σ(*-um-*), DEP(C), MAX(C), DEP(V)
≫ MPARSE(*-um-*)
(no *-um-* form violates preceding constraints)
≫ NOCODA, *COMPLEXONS, PREFIX(*-um-*)
(*-um-* forms can have codas and complex onsets, and *-um-* can be infixed)

On the basis of words beginning with labial sonorants, it has been established that ONSET, OCP[labial], PREFIX/σ(*-um-*) and DEP(C) all dominate MPARSE(*-um-*). MAX(C) and DEP(V) were previously shown to dominate NOCODA and *COMPLEXONS (see (15)). The location of MPARSE(*-um-*) in the hierarchy sets the harmony threshold: actually occurring *-um-* words can only violate lower-ranking constraints. Those constraints are NO CODA, *COMPLEXONS and PREFIX(*-um-*). The following tableaux certify the validity of these ranking arguments:

(22)

/um-sulat/	ONS	OCP	PRE-FIX/σ	DEP(C)	MPARSE	NOCODA=*COMP ONS	PRE-FIX
☞ a. s**u.m**u.lat						*	*
b. su.l**u.m**at			*!			*	*
c. **um**.su.lat	*!					**	
d. ʔ**um**.su.lat				*!		**	*
e. ⊙					*!		

In (22), candidates without infixation are ruled out by the undominated constraints ONSET or DEP(C). Excessive infixation is excluded by PREFIX/σ(-*um*-). Since there is a form (22a) that violates none of these constraints, there is an alternative to the null output.

(23)

/um-preno/	ONS	OCP	PRE-FIX/σ	DEP(C)	MPARSE	NOCODA=*COMP ONS	PRE-FIX
☞ a. **pum**.re.no						*	*
☞ b. pr**u.me**.no						*	*
c. **um**.pre.no	*!					* *	
d. ʔ**um**.pre.no				*!		* *	*
e. ⊙					*!		

Tableau (23) presents two winners, depending on which order of NOCODA and COMPLEXONS is chosen at EVAL time. As in the previous tableau, candidates without infixation violate undominated ONSET or DEP(C). The null output loses because there are other candidates that violate none of the constraints that dominate MPARSE(-*um*-).

(24)

/um-walow/	ONS	OCP	PRE-FIX/σ	DEP(C)	MPARSE	NOCODA=*COMP ONS	PRE-FIX
☞ a. ⊙					*		
b. w**u.m**a.low		*!				*	*
c. wa.l**u.m**ow			*!			*	*
d. **um**.wa.low	*!					**	
e. ʔ**um**.wa.low				*!		**	*

In (24), the null output wins because the alternatives are all worse: a [w**um**V] sequence (24b) that violates OCP[labial]; deep infixation (24c), contrary to the dictates of PREFIX/σ(-*um*-); or the usual problems with ONSET and DEP(C) (24d, e).[18]

A final remark. This analysis requires that the underlying form of [ʔabot] is /ʔabot/ and not /abot/, treating this word exactly on a par with

[18] Alan Prince raises an important typological question: is deep infixation ever possible in any language under any ranking (cf. McCarthy & Prince 1993b)? Samek-Lodovici (1993) finds a possible example involving a geminating (i.e. mora) infix, though cases similar to [wal**um**ow] are not known to me. With further refinement, the theory developed here may offer an explanation for this typological gap. If shallow infixation is ruled out, then deep infixation competes with suffixation (cf. Noyer 1993, Fulmer 1997). Deep infixation and suffixation tie on the constraints PREFIX and PREFIX/σ; if some other constraint, such as morpheme CONTIGUITY (Kenstowicz 1994, McCarthy & Prince 1999), disfavours infixation, then suffixation must win. This cannot be the full story, however, because other constraints may also militate against suffixation (e.g. the OCP rules out both [w**um**alow] and [walow**um**]).

/sulat/. Independently, there is good reason to assume that the underlying form is indeed /ʔabot/: there are no [ʔ]/∅ alternations, and the root-initial [ʔ] shows up even after consonant-final prefixes. Of course, under richness of the base (Prince & Smolensky 1993), the grammar of Tagalog is also responsible for correctly disposing of hypothetical V-initial roots. They must be treated unfaithfully, because Tagalog words never begin with a vowel, but no active alternations show how they are treated – the traditional assumption that hypothetical /apak/ becomes [ʔapak] is without empirical support. If we nonetheless assume in the absence of evidence that /apak/ → [ʔapak] is the right disposition of V-initial words, then the ranking [[ONSET ≫ DEP(C)]] must be added to the grammar in (21). Hypothetical /apak/ then would surface as [ʔapak]. This verb would have no *-um-* form, because DEP(C) dominates MPARSE(*-um-*) (see (20)). Do such verbs exist? They could: some surface [ʔ]-initial verbs don't take *-um-*, as predicted, but that could also be because verbs in general are lexically marked to take *-um-*. In any case, richness of the base does not challenge the analysis presented here.

5.4 Infixation in Nakanai

Nakanai is an Austronesian language spoken in New Britain, carefully described and analysed by Johnston (1980). In this language, as in Tagalog, infixation shows the effect of categorical alignment: a morphological process is blocked when misalignment is too severe. Specifically, segment-sized alignment discrepancies are permitted, but misalignment by a syllable or more is not.

Nakanai disallows codas and allows onsetless syllables freely. Each vowel is said by Johnston to constitute a syllable on its own, so a word like [a.u] 'to steer' is disyllabic. Stress falls strictly on the penult. Words are minimally disyllabic in size.

Nakanai forms nominalisations by inserting *-il-*[19] before the main-stressed vowel in words containing exactly two syllables. In longer words, however, nominalisations are formed by suffixing *-la* instead:

(25) *Nakanai nominalisation*

a. *Disyllables*		b. *Longer words*	
iláu	'steering'	sageg**éla**	'happiness'
tilága	'fear'	vikuéla	'fight'
gilógo	'sympathetic'	vigilemulimul**íla**	'story'

The size of the entire word, and not just the size of the root, is decisive. For instance, the last example is based on the disyllabic root [gile] 'to sift'.

[19] There are additional alternations in the form of the infix. It is *i* before *l-* or *r*-initial roots. Its vowel is *u* before *u* or *o* in the next syllable. And its consonant is *r* in agreement with an *r* in the next syllable (cf. Cohn 1992).

This rather puzzling distribution of the *-il-* and *-la* alternants can be made sense of when it is recalled that Nakanai has penult stress. The *-il-* alternant is attracted to the left edge of the word – it is a formal prefix – like Tagalog *-um-*. It is also attracted to the main-stressed syllable. In disyllables, the main-stressed syllable is also the initial syllable, so both desiderata for *-il-* placement can be more or less satisfied. In longer words, however, there is no way to attach *-il-* to the penult main-stressed syllable and also keep it close to the beginning of the word.

To get the analysis rolling, we first need to make some assumptions about the source of the *-il-*/*-la* alternation. The form of these two alternants is not phonologically predictable, though their distribution is. There is a large literature on this kind of allomorphy. The basic idea is that allomorphs are listed together in the lexicon, so an underlying representation will contain a set of alternants, such as /{il, la}-sagege/.[20] When GEN constructs candidates, it uses both input alternants. This means that [s**il**agege], [sagege**la**], [**la**sagege], [sa**la**gege] and [sagege**il**] are among the candidates that incur no faithfulness violations. (Of course, unfaithful candidates like [s**ul**agege] or [sagege**a**] are also in the mix.) The choice of the winning candidate – and hence the selection of *-il-* or *-la* – is as usual the responsibility of EVAL.

The allomorph *-il-* is a formal prefix with its distribution under the control of undominated PREFIX/σ(*-il-*). The allomorph *-la* is a formal suffix, and since it is never infixed, its distribution is governed by undominated SUFFIX(*-la*). To understand the *-il-*/*-la* alternation, we first need to get a handle on a couple of descriptive problems: *-la* functions as kind of default, occurring only when *-il-* is blocked, and *-il-* is attracted to the stressed syllable.

The first thing to address is *-la*'s default status. Because *-la* does not occur with disyllables (*[tag**ala**]), there must be some cost associated with it. The cost is not input–output faithfulness, however, since neither *-il-* nor *-la* is more faithful. One possibility is that the affixal alternants are lexically prioritised, as Bonet *et al.* (2003) have argued for Catalan. Another possibility is that *-la*'s cost is measured by output–output faithfulness to stress (Kenstowicz 1996, 1997, Benua 1997, Alber 1998, Kager 2000, McCarthy 2000b, Pater 2000) or paradigm uniformity (Raffelsiefen 1995, 1999, Kenstowicz 1996: 385, McCarthy 1998). Kager proposes the following constraint:

(26) OO-PK-MAX (after Kager 2000: 127)

Let α be a segment in the base and β be its OO correspondent in the derived form. If α is a stress peak, then β is a stress peak.

[20] The idea of lexical entries as sets of allomorphic alternants originated with Hudson (1974) and is adopted by Hooper (1976). There is a considerable literature applying OT to problems in allomorphy or lexical selection, much of it cited in McCarthy (2002b: 183–184).

When a word takes the suffix *-la* or almost any other suffix in Nakanai, the stress shifts to the new penult: [sagége]/[sagegé**la**]. Stress shift is a violation of OO-PK-MAX: the stress peak $é_4$ in [$s_1a_2g_3é_4g_5e_6$] does not stand in correspondence with a segment that is a stress peak in [$s_1a_2g_3e_4g_5é_6$**la**]. The infix *-il-* does not affect stress placement, since it falls to the left of the stressed nucleus: [tága]/[**til**ága]. On grounds of OO-PK-MAX alone, the *-il-* alternant is favoured.

While *-la* is suffixed, *-il-* is attracted to the main stress. It is not unusual to find infixes that are tropic to stress: reduplication in Samoan targets the main-stressed syllable (Marsack 1962, Broselow & McCarthy 1983), as does possessive suffixation in Ulwa (Hale & Lacayo Blanco 1989, McCarthy & Prince 1990). The central analytic idea is that affixes may be prefixed or suffixed to the head foot rather than the prosodic word (Broselow & McCarthy 1983, Inkelas 1989, McCarthy & Prince 1990, 1993b). The responsible constraints follow the same general pattern as (17): AFX-TO-HD(*-il-*) is violated by an *il*-containing candidate where *-il-* is separated by one or more segments from the head foot. This constraint is undominated in Nakanai, since *-il-* is never found in any other context. (Forms like [**il**taga] are ruled out by another undominated constraint, NOCODA.)

With these preliminaries taken care of, we are now in a position to explain the conditions on the *-il-*/*-la* alternation. The key idea is that *-il-* cannot stray from the first syllable of the word because the categorical constraint PREFIX/σ(*-il-*) is undominated. When AFX-TO-HD(*-il-*) and PREFIX/σ(*-il-*) cannot both be satisfied, as is the case with trisyllabic and longer words, then the *-la* allomorph appears instead, even though it is dispreferred by OO-PK-MAX. This allomorph sidesteps both of the problematic constraints, since they pertain only to *-il-* and not to *-la*.

One element of the analysis, then, is crucial domination of *la*-disfavouring OO-PK-MAX by PREFIX/σ(*-il-*):

(27) PREFIX/σ(*-il-*) $\gg$ OO-PK-MAX

/{il, la}-sagege/	PREFIX/σ(*-il-*)	OO-PK-MAX (cf. *sagége*)
☞ a. sa.ge.gé.**la**		*
b. sa.g**i**.**l**é.ge	*!	

Choosing the suffixed *-la* alternant in (27) leads to stress shift on the OO dimension, but the alternative of placing the *-il-* alternant more than a syllable away from the left word edge is ruled out by an undominated constraint.[21]

[21] Other undominated constraints exclude some plausible competitors for the winner in (27). In *[sagége**la**], OO-PK-MAX is satisfied by treating *-la* as a stress-neutral suffix. But *-la-*, like nearly all Nakanai suffixes, is stress-determining, not stress-neutral. This means that the metrical constraints responsible for penult stress must dominate OO-PK-MAX. Another reasonable-looking competitor is [sa**la**gége], with

Another element of the analysis is crucial domination of OO-PK-MAX by AFX-TO-HD(-*il*-):

(28) AFX-TO-HD(-*il*-) ≫ OO-PK-MAX

/{il, la}-sagege/	AFX-TO-HD(-*il*-)	OO-PK-MAX (cf. *sagége*)
☞ a. sa.ge.gé.**la**		*
b. **si.la**.gé.ge	*!	

Taken together, the ranking arguments in (27) and (28) establish the conditions for choice between the -*il*- and -*la* allomorphs. For -*il*- to occur, it cannot be displaced by as much as a syllable from the beginning of the word or at all from the main stress. If these conditions are not satisfied, then the -*la* allomorph occurs instead, even though its presence forces a stress shift in violation of OO-PK-MAX.

For present purposes, the most important thing about the prefix -*il*- is that does not fall exactly at the left word edge in forms like [**til**ága] (see (29)). Deviation by a whole syllable is not possible, as the tableau (27) shows, but deviation by just a segment is tolerated. This demonstrates that OO-PK-MAX dominates PREFIX(-*il*-):

(29) OO-PK-MAX ≫ PREFIX(-*il*-)

/{il, la}-taga/	OO-PK-MAX (cf. *tága*)	PREFIX(-*il*-)
☞ a. **ti.l**á.ga		*
b. ta.gá.**la**	*!	

In [**til**ága], the allomorph -*il*- is misaligned by a segment from the left edge of the word. But this misalignment is not as big as a whole syllable, so the undominated PREFIX/σ constraint is not violated. That is why disyllabic words – and only disyllabic words – take the -*il*- allomorph.

This covers the main points of the analysis, summarised by the ranking in (30).

(30) *Nakanai ranking summary*
PREFIX/σ(-*il*-), AFX-TO-HD(-*il*-) ≫ OO-PK-MAX ≫ PREFIX(-*il*-)

The constraint OO-PK-MAX favours -*il*- over -*la*, so it sets a kind of threshold: forms with -*il*- can violate only constraints ranked below OO-PK-MAX. The constraints ranked above OO-PK-MAX are those that block -*il*- in favour of -*la*. These constraints say that the allomorph -*il*- is

infixed -*la*. But -*la* is never infixed, so SUFFIX(-*la*) is undominated, also crucially ranked above OO-PK-MAX.

simultaneously attracted to the left word edge and the main stress. It cannot be displaced from the word edge by a syllable or more, nor from the head foot, so when a word is longer than a single foot, the *-il-* allomorph fails completely and *-la* takes its place. But *-la* has a cost: because it is a stress-determining suffix in a language with penultimate stress, it produces a stress alternation. The allomorph *-il-* avoids this alternation, and that option is taken when *-il-* can get close enough to its preferred locus so that it violates only the low-ranking PREFIX constraint.

5.5 Summary

Nakanai and Tagalog show that affix-position constraints must categorically distinguish the extent to which an affix is malpositioned. Classic gradient alignment constraints cannot do this. The separate ranking required in Nakanai and Tagalog is not an option with classic alignment. (See Klein 2002 for a further argument in support of quantisation based on evidence from infixation in Chamorro.)

As I noted in §5.3, it is certainly possible to construct a theory with gradient alignment and violation quanta. Indeed, such a theory is contemplated in the formalisation of gradient alignment in (2). Tagalog and Nakanai could be analysed in this theory, substituting gradient ALIGN (*-um-*/*-il-*, L; Wd, L; σ) in tableaux (24) and (27).

The problem with this gradience-*cum*-quantisation theory is not descriptive coverage – it is a richer theory, after all – but parsimony. The categorical constraints in (17) are sufficient for Tagalog and Nakanai. These constraints take over the actually observed functions of gradience. For example, Prince & Smolensky (1993) attribute the ill-formedness of words like *[pren**um**o] to gradient ALIGN. But categorical PREFIX/σ (*-um-*) is sufficient to rule out *[pren**um**o], as shown in (23). So, although gradience and quantisation are not logically incompatible, they compete for the same explanatory turf. Constraints like PREFIX/σ(*-um-*) have the violation quanta without the trappings of gradience. As we have seen in Tagalog and Nakanai, and as I argue elsewhere in this article, the need for gradience is very much in doubt. Since OT indisputably has categorical constraints, and plenty of them. Occamite reasoning demands that we rid the theory of gradient constraints if categorical ones are sufficient.

6 Alignment and stress

6.1 Introduction

The hypothesis that all OT constraints are categorical faces perhaps its greatest challenge from stress theory. Since the very beginning of research on OT, starting with Prince & Smolensky's (1993) constraint EDGEMOST, gradient alignment constraints have been used to analyse stress

phenomena. As shown in (3), gradient constraints like ALLFTR are the basis of directional foot-parsing. Yet ALLFTR and its congeners cannot be reconstructed as categorical constraints conforming to (1). The problem is that ALLFTR treats each foot as a locus of violation and evaluates each such locus for distance from a word edge (see §2 for the argument).

Recent work by Kager (2001) offers a very different perspective on directional foot-parsing. Using constraints on stress clashes and lapses, Kager is able to obtain a typology of directionality effects that better fits the facts than the ALLFTR approach. This work is summarised in §6.2.

Gradient alignment has also been important in controlling the uniqueness and location of the main-stress foot. In §6.3, I show how and why simple, categorical constraints based on the End Rule of Prince (1983) prove to be sufficient.

6.2 Directional foot-parsing

As shown in (3), gradient alignment constraints of the ALIGN(Ft, Wd) variety, such as ALLFTR, simulate the effects of directionality iterative foot-parsing in rule-based metrical phonology (Prince 1976, Halle & Vergnaud 1978, Hayes 1980 and many others). Summarising Kager (2001), this section shows that this application of gradient alignment is not only unnecessary but actually problematic, since it yields an overly rich typology. This article's goal of showing that all constraints are categorical is supported by eliminating this otherwise compelling example of gradience.

Through permuted ranking, ALIGN(Ft, Wd) allows free choice of parsing direction independent of foot type. But the choice is not really free: there are no convincing examples of right-to-left iambic stress systems (Kager 1993, McCarthy & Prince 1993b, Hayes 1995: 262ff). Furthermore, Kager claims that bidirectional systems, with a single foot at one end and iteration from the other end, never iterate from the main stress toward the secondary. That is, while there are languages like Polish, which parses heptasyllables as $[(_{\text{l}}\sigma\sigma)(_{\text{l}}\sigma\sigma)\sigma(\text{'}\sigma\sigma)]$, there are no solid examples of languages with the parse $[(_{\text{l}}\sigma\sigma)\sigma(_{\text{l}}\sigma\sigma)(\text{'}\sigma\sigma)]$ (though see note 23). Again, free permutation of gradient ALIGN(Ft, Wd) predicts that these non-existent patterns should occur.

Instead of the gradient ALIGN(Ft, Wd) constraints, Kager proposes an enriched theory of the constraint *LAPSE (Selkirk 1984, Nespor & Vogel 1989, Hung 1994, Green & Kenstowicz 1995, Elenbaas & Kager 1999, Gordon 2002). A lapse is a sequence of unstressed syllables, independent of foot structure. Though lapses are marked configurations generally, Kager hypothesises that they are less marked in two positions, word-finally and adjacent to the main-stressed syllable, and more

marked in another position, word-initially. The responsible constraints are these:

(31) *Lapse constraints in Kager* (2001) (*definitions reformulated to conform to* (1))

a. *LAPSE
$*\breve{\sigma}$ / __ $\breve{\sigma}$
i.e. assign one violation-mark for each pair of adjacent unstressed syllables.

b. LAPSE-AT-END
$*\breve{\sigma}$ / __ $\breve{\sigma}\alpha$, where α is non-null
i.e. assign one violation-mark for each pair of adjacent unstressed syllables that is not word-final.

c. LAPSE-AT-PEAK
$*\breve{\sigma}$ /α __ $\breve{\sigma}\beta$, where α does not end and β does not begin with $\acute{\sigma}$
i.e. assign one violation-mark for each pair of adjacent unstressed syllables that is not adjacent to the main-stressed syllable.

d. *INITIALLAPSE
$*\breve{\sigma}$ /$_{\text{Wd}}$[__ $\breve{\sigma}$
i.e. assign one violation-mark for each pair of adjacent unstressed syllables that is word-initial.

The constraint *LAPSE militates against stress lapses generally, but the constraints LAPSE-AT-END and LAPSE-AT-PEAK can license lapses in those specific environments by disallowing them everywhere else. Conversely, *INITIALLAPSE disfavours lapses word-initially. Directionality effects are obtained from the interaction of these categorical constraints, with a better fit to observation than the gradient ALIGN(Ft, Wd) constraints.[22]

Table I shows how the effects of directionality follow in Kager's system. This table is an unranked tableau. It provides information about candidate performance on the various *LAPSE constraints without considering the constraints' ranking. In addition, the left and right ALIGN (Wd, Ft) constraints are included. I will discuss these constraints later in this section.

Table I uses seven-syllable words, since directionality is usually visible only in odd-parity words, and seven syllables are need to show the full range of observed patterns. Exhaustive parsing is assumed up to degeneracy – that is, FTBIN is undominated. The candidates are grouped according to the position of main stress (indicated by the numeral 1) and whether feet are iambic or trochaic. Each group, such as (a)–(d), contains a set of candidates that compete with one another, holding main-stress location and foot type constant. In effect, the candidates in each group compete in directionality only. Competition across groups also occurs, of

[22] The suggestion that directionality effects are reducible to constraints on the position of unfooted syllables was made to me in 1993 by Junko Itô and Armin Mester. At the time, I summarily (and, in retrospect, foolishly) rejected this idea.

		*Lapse	Lapse-at-End	Lapse-at-Peak	*Init Lapse	Align (Wd,L;Ft,L)	Align (Wd,R;Ft,R)
Trochaic: Main Stress Left							
a.	(10)(20)(20)0	*		*			*
b.	(10)(20)0(20)	*	*	*			
c.	(10)0(20)(20)	*	*				
d.	0(10)(20)(20)					*	
Trochaic: Main Stress Right							
e.	0(20)(20)(10)					*	
f.	(20)0(20)(10)	*	*	*			
g.	(20)(20)0(10)	*	*				
h.	(20)(20)(10)0	*					*
Iambic: Main Stress Left							
i.	(01)(02)(02)0						*
j.	(01)(02)0(02)	*	*	*			
k.	(01)0(02)(02)	*	*				
l.	0(01)(02)(02)	*	*		*	*	
Iambic: Main Stress Right							
m.	0(02)(02)(01)	*	*	*	*	*	
n.	(02)0(02)(01)	*	*	*			
o.	(02)(02)0(01)	*	*				
p.	(02)(02)(01)0						*

Table I

Typology of rhythmic stress systems (after Kager 2001).

course, but it involves further constraints (FtForm, Align(Hd, Wd)) that would not fit in the table.

The unranked tableau in Table I is useful because it permits quick inferences about which candidates are harmonically bounded. A candidate is harmonically bounded if it incurs a proper superset of a competitor's violation-marks. The harmonic bounding we are interested in is within a group, because the candidates within each group tie on all constraints that are not included in Table I. The shaded rows contain candidates that are harmonically bounded by other candidates in the same group. For example, (b) is harmonically bounded by (c), because (b) has a proper superset of (c)'s violation-marks. Rows that are not shaded are predicted to be possible in Kager's theory under some ranking(s) of the given constraints.

Kager observes that all of the non-harmonically-bounded trochaic systems are attested: (a) is Pintupi, (c) is Garawa, (d) is Wargamay, (e) is Warao, (g) is Piro and (h) is Cairene Arabic (substituting moras for syllables). The harmonically bounded pattern in (b) has not been reported; there are a few reports of (f) in the literature, but all are subject to other interpretations.[23] If indeed (b) and (f) are impossible, then this must count as evidence from language typology against gradient ALIGN(Ft, Wd) and in favour of the categorical constraints operating in Table I. The problem with gradient alignment is that it easily produces unattested (b) and its symmetric counterpart (f) – for instance, the ranking to get (b) with gradient constraints is [[ALIGN(Wd, L; Ft, L) ≫ ALIGN(Wd, R; Ft, R) ≫ ALIGN(Ft, L; Wd, L)]]. Gradient ALIGN(Wd, Ft) can do this because it allows direction of foot-parsing to be specified independently of foot form (trochaic or iambic).

Attestation of the non-harmonically-bounded iambic systems is less complete. Pattern (i) in Table I is Araucanian and (p) is Creek (again, moraic rather than syllabic). Neither (k) nor (o) has been observed. Still, this is real progress over the theory with gradient ALIGN(Ft, Wd). It predicts not only (k) and (o), but also the remaining patterns (j), (l), (m) and (n), all of which are unattested. Again the problem with ALIGN(Ft, Wd) in iambic systems is that it permits free choice of parsing direction, when in fact the choice is not free: iambic systems are consistently left-to-right. Kager's proposal explains why: the right-to-left iambic systems (l) and (m) are harmonically bounded because they have a marked initial lapse. Trochaic systems are not similarly asymmetric because an initial lapse is not a consequence of right-to-left trochaic parsing.

Kager also examines stress systems that allow degenerate feet and therefore have no lapses. In Murinbata, for example, seven-syllable words have the trochaic stress pattern (10)(20)(20)(2), while in Weri they have the iambic pattern (2)(02)(02)(01). This too looks at first glance like a directionality effect. With gradient alignment, one would say that ALIGN(Ft, Wd, L) is active in Weri, as the following tableau shows:

(32) *Gradient* ALIGN(Ft, Wd, L) ≫ ALIGN(Ft, Wd, R) *in Weri*

	ALIGN(Ft,Wd,L)	ALIGN(Ft,Wd,R)
☞ a. (2)(02)(02)(01)	*********	**********
b. (02)(2)(02)(01)	**********!	**********
c. (02)(02)(2)(01)	**********!	**********
d. (02)(02)(02)(1)	**********!	*********

This tableau reveals a curious property of gradient ALIGN(Ft, Wd) that was first noted by Crowhurst & Hewitt (1995). In systems that disallow

[23] The strongest apparent counterexample is Indonesian (Cohn 1989, Cohn & McCarthy 1994). Kager observes that the crucial forms are Dutch loans and may simply be reproducing stress from the original language.

degenerate feet, the effect of right-to-left footing is obtained from high-ranking ALIGN(Ft, Wd, R), as shown in (3). But in systems that permit degenerate feet, right-to-left footing requires high-ranking ALIGN (Ft, Wd, L). The directional sense of gradient alignment is oddly reversed, depending on which of FTBIN or PARSE-σ is ranked higher.

Apart from this formal peculiarity of gradient alignment, there is a typological problem as well. The left-to-right stress pattern in (32d) is attainable by simply permuting the gradient constraints, yet it does not seem to exist. This typological skew leads to Kager's further proposal that constraints on clash (i.e. adjacent stressed syllables), rather than gradient alignment, are responsible for apparent directionality effects in stress systems that permit degenerate feet. Of all the candidates in (32), only (a) avoids clash completely. It therefore satisfies *CLASH better than its competitors. This constraint and its allies are sufficient, Kager argues, to account for the full range of observed directionality effects in these systems. For further details, see Kager (2001).

A theory of directional foot-parsing based on the distribution of lapses and clashes is superior on typological grounds to the gradient ALIGN (Ft, Wd) constraints. Kager's results about the typology of stress systems converge with the overall argument of this article: the thesis that all constraints are categorical finds support from the discovery that a prime application of gradient alignment is deeply flawed. This convergence of results from very different directions is perhaps an indication that we are on the right track here.

It might appear that the gradient theory has not been entirely dismissed, though, because Table I still has left and right ALIGN(Wd, Ft) constraints. Though these constraints are formalised in McCarthy & Prince (1993a) using the standard gradient alignment schema, in actual analytic practice they are and always have been treated as categorical constraints. (The situation, then, is the same as with the morphology–prosody alignment constraints discussed in §4.) The constraint ALIGN(Wd, L; Ft, L) asks whether the word begins with a foot or not. If the candidate contains no more than one prosodic word, then no more than one violation-mark can be assigned. Since gradience plays no role in evaluating these constraints, alignment is at best unnecessary, and there is no barrier to replacing ALIGN(Wd, Ft) with overtly categorical constraints. One approach is to posit positional versions of the categorical foot-parsing constraint PARSE-σ (which is itself a member of the categorical deconstruction of EXHAUSTIVITY – see §3):

(33) *Positional* PARSE-σ

a. PARSE-σ_I
$*\sigma^\circ / [_{Wd}$ __ , where σ° denotes a syllable that is not contained in any foot.

b. PARSE-σ_F
$*\sigma^\circ /$ __ $]_{Wd}$

That is, no prosodic word can begin or end with a syllable that Wd immediately dominates, so peripheral syllables must be footed. Substituting these constraints into Table I and into the tableaux in Kager (2001) has no effect on the outcome, precisely because the ALIGN(Wd, Ft) constraints have never actually been treated gradiently. Alignment, gradient or otherwise, has been effectively eliminated from the theory of directional foot-parsing.[24]

6.3 Constraints on the head foot

6.3.1 *Introduction.* Every prosodic word contains one and only one head foot, which is the locus of main stress. The existence and uniqueness of the head foot are usually taken to be axiomatic – universal properties of GEN rather than violable constraints. To avoid unnecessarily complicating the discussion, I will stick with that assumption here.

Gradient alignment constraints affect the head foot in two important ways (McCarthy & Prince 1993a). First, gradient ALIGN(Ft, Wd) (i.e. ALLFTL/R) has been used to limit words to a single foot, the head, to account for non-iterative foot assignment. If ALLFTL/R is ranked above PARSE-σ, then the winning candidate will contain no more feet than the bare minimum, one, because any other foot is inevitably misaligned. Second, gradient ALIGN(Hd(Wd), Wd) is invoked to express the generalisation that main stress is located on the leftmost or rightmost foot.

There is, however, a categorical alternative to gradient ALIGN(Hd(Wd), Wd) that antedates OT. The End Rule of Prince (1983) is a categorical, inviolable-but-parametrised constraint on phonological representations. Here is one of several formulations of the End Rule that Prince considers:

(34) *End Rule* (Prince 1983: 19)
In a constituent C, the leftmost/rightmost entry at level α corresponds to an entry at level β, where β is the next level up from α in the prosodic hierarchy and β is the prosodic category that syntactic category C is related to.

For example, if C is the syntactic word, then β is the prosodic word, and α is the foot. The word-level End Rule is obeyed if the leftmost/rightmost foot is prominent at the level of the prosodic word – i.e. if it is the head of the prosodic word.

Because it has the form of a categorical markedness constraint, the End Rule can be carried over virtually unaltered into an OT context. In terms

[24] In a theory with various *LAPSE constraints, it might seem that PARSE-σ and its positional variants in (33) are superfluous. This does not seem to be true. There is a basic difference in form and function between *LAPSE and PARSE-σ; the former is part of the prominential system and is completely indifferent to foot structure; the latter is part of the prosodic-hierarchy system and is completely indifferent to prominence.

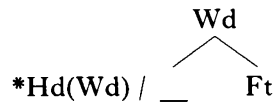
of the general schema (1), the End Rule can be translated into the constraints given in (35):

(35) *End Rule constraints*

a. ER-L

i.e. the head foot is not preceded by another foot within the prosodic word

b. ER-R



i.e. the head foot is not followed by another foot within the prosodic word

(The notational conventions assumed here are the same as in (17).) ER-L is satisfied if the head foot is the first foot in the prosodic word, regardless of whether it is literally initial in the word. ER-R is the same, with mirror symmetry. There is nothing new about either of these constraints; except that they are violable and non-parametrised, they are identical to Prince's End Rule.

In §6.3.2 and §6.3.3, I apply ER-L and ER-R to the one-foot-per-word phenomenon and main-stress placement.[25]

6.3.2 *One foot per word.* There are two situations of interest. One involves languages that are said to lack secondary stress, from which it is inferred that words contain no feet except for the head. The other involves the one-foot minimal-word template that is frequently encountered in reduplication. Each will be discussed in turn.[26]

As was just noted, gradient ALIGN(Ft, Wd), if it dominates PARSE-σ, will prevent iterative foot-parsing, since every additional foot contributes more alignment violations. Claims about non-iterative footing should be approached sceptically, however. Often, they rely on the original analyst's silence about secondary stress. For example, Latin has been described as having non-iterative stress, but this is only because the Latin pronunciation tradition does not include any information about secondary stress. In

[25] It is in principle possible to mimic the effects of ALIGN(Hd(Wd), Wd) using a categorical constraint, as has been pointed out to me by Colin Wilson and several anonymous reviewers: $*\sigma$ / $)_{Hd} \ldots \_\_ \ldots]_{Wd}$. Why is that possible for this alignment constraint but not others (see §2)? Because heads are guaranteed to be unique in Wd, so the first $\forall$ in the alignment definition (2) can be ignored. This shows, as I noted in §2, that categoricality is not a sufficient condition for licit constraints (e.g. locality conditions are likely to be relevant; cf. Eisner 1999).

[26] A direct assault on one-foot-per-word is also possible by invoking a constraint of the *STRUC family (Prince & Smolensky 1993), *STRUC(Ft), which assigns one violation-mark for every foot in the candidate under evaluation. But the need for and desirability of such constraints have been impeached on typological and other grounds by Gouskova (2003).

fact, though, we know from work by Mester (1994) that Latin did have iterative footing. The second syllable of words like [pudi:citiam] 'chastity (ACC SG)' or [vere:bamini] 'you (PL) were afraid', although underlyingly long, is observed to scan as short: [pudicitiam], [verebamini]. This shortening process makes sense if these words are parsed into a succession of bimoraic feet: [(pùdi)(cíti)am]. The case of Cairene Arabic is also instructive. In my observation, it does not have systematic secondary stress (though cf. Kenstowicz 1980, Welden 1980, Harms 1981), but there can be little doubt that there is an iterative foot-parse, since otherwise the position of main stress could not be explained (McCarthy 1979, Hayes 1995: 64–71).

This apparent disconnection between metrical structure and observed secondary prominence has led to various formal proposals (Halle & Vergnaud 1987, Blevins 1992, Crowhurst 1996, de Lacy 1998), though it seems equally reasonable to see it as an aspect of the phonetics–phonology mapping. As Hayes (1995: 119) writes, 'we might suppose that the phonetic and phonological rules of the language just happen not to provide any means of manifesting foot structure. This solution is viable, given what we have seen ... concerning the language-specific phonetic realization of stress.' The point is that even solid evidence for the absence of secondary stress, if such is possible, does not permit the inference that words have only one foot, because the range of ways in which metrical structure can be realised phonetically is so broad.

With these empirical caveats aside, the End Rule constraints in (35) offer a way of limiting words to a single foot. If the head foot is obliged to be both the leftmost and rightmost foot in the word, then it cannot be preceded or followed by other feet. Therefore, if these constraints are ranked above PARSE-σ and its positional variants, then the head foot will also be the only foot. Tableau (36) shows this result:

(36) *Non-iterative footing from End Rule constraints*

	ER-L	ER-R	PARSE-σ
☞ a. [(10)00] *or* [00(10)]			**
b. [(10)(20)]		*!	
c. [(20)(10)]	*!		

These constraints do not fully determine the outcome, of course. Constraints like those in (33) will decide whether (36a) or (36b) wins.

This sort of analysis might seem surprising from the perspective of a parametric stress theory like the one in Prince (1983). The original End Rule is either left or right, on a language-particular basis. But in OT, where all constraints are present in every grammar, there is no barrier to both ER-L and ER-R being active in a single language, and (36) shows why this may be a good idea.

Prosodic words are exactly one foot long in the minimal-word reduplicative template (McCarthy & Prince 1994, 1999). Via emergence of the

unmarked, gradient ALIGN(Ft, Wd) and PARSE-σ work together to determine the shape of the reduplicant. Ranked below MAX-IO and above MAX-BR, they ensure that the reduplicant is monopodal. The following hypothetical example is based on the Australian language Diyari (Austin 1981, Poser 1989):

(37) *Gradient* ALIGN(Ft, L; Wd, L; σ), PARSE-$\sigma \gg$ MAX-BR

/RED+ŋandawalka/	ALIGN(Ft,Wd)	PARSE-σ	MAX-BR
☞ a. [(ŋánda)]-[(ŋánda)(wàlka)]	**		*****
b. [(ŋánda)wa]-[(ŋánda)(wàlka)]	**	*!	***
c. [(ŋánda)(wàlka)]-[(ŋánda)(wàlka)]	****!		

Austin presents evidence from stress and allophony that the reduplicant is a separate prosodic word in Diyari, as indicated by the [ ] brackets. The role of gradient ALIGN(Ft, L; Wd, L; σ) in this analysis is to rule out (37c), with total reduplication. (It also ensures left-to-right foot-parsing in unreduplicated roots.) Total reduplication of a quadrisyllabic root produces a dipodal reduplicant, which is worse aligned than the monopodal reduplicant (37a). Ranked between MAX-IO and MAX-BR, ALIGN(Ft, Wd) controls the size of the reduplicant without affecting the size of the base.

Suitably ranked, the categorical End Rule constraints can produce the same result. Since Diyari is a language with main stress on the first foot and iterative footing, ER-L and PARSE-σ (or *LAPSE) must dominate ER-R:

(38) ER-L, PARSE-$\sigma \gg$ ER-R

	ER-L	PARSE-σ	ER-R
☞ a. [(ŋánda)(wàlka)]			*
b. [(ŋànda)(wálka)]	*!		
c. [(ŋánda)walka]		**!	

If ER-R is itself ranked above MAX-BR, then the monopodal reduplicant is obtained:

(39) PARSE-$\sigma \gg$ ER-R $\gg$ MAX-BR

/RED+ŋandawalka/	PARSE-σ	ER-R	MAX-BR
☞ a. [(ŋánda)]-[(ŋánda)(wàlka)]		*	*****
b. [(ŋánda)wa]-[(ŋánda)(wàlka)]	*!	*	***
c. [(ŋánda)(wàlka)]-[(ŋánda)(wàlka)]		**!	
d. [(ŋánda)walka]-[(ŋánda)(wàlka)]	**!	*	

In the failed candidate (39c), the reduplicant contains a foot that is separated by another foot from the left word edge. Since this violation-mark is avoidable by less zealous copying, and since MAX-BR is low-ranked, the first candidate wins. Gradient foot alignment is not crucial in accounting for the minimal-word template.

Before leaving the topic of non-iterative footing, it is worth pointing out a related empirical prediction that follows from eliminating gradient alignment. Imagine a language with the non-iterative stress pattern [$\sigma\sigma\sigma(\text{'}\sigma\sigma)\sigma$] or its moraic equivalent. The standard approach with gradient alignment posits the ranking [[NON-FIN(Ft) ≫ ALIGN(Ft, R; Wd, R; σ) ≫ PARSE-σ]]. Under this ranking, the final syllable is unfooted, and, since every word must contain at least one foot, the sole foot is aligned as far to the right as possible.

The iterative version of this stress pattern does not require gradient alignment – see (a), (h) in Table I. But the non-iterative version is potentially a problem for the theory sketched here. The intended output [$\sigma\sigma\sigma(\text{'}\sigma\sigma)\sigma$] ties with its competitors *[$\sigma\sigma(\text{'}\sigma\sigma)\sigma\sigma$] and *[$\sigma(\text{'}\sigma\sigma)\sigma\sigma\sigma$], and fares worse than *[$(\text{'}\sigma\sigma)\sigma\sigma\sigma\sigma$], on the positional PARSE-σ constraints in (33). The *LAPSE constraints in (31) are of no help either. Macedonian is said to exemplify this stress pattern (Franks 1989, Hammond 1989, Halle & Kenstowicz 1991), though again the inference that words contain only a single foot is insecure. If the empirical caveats raised at the beginning of this subsection can be resolved, so that the case for [$\sigma\sigma\sigma(\text{'}\sigma\sigma)\sigma$] metrical structure is on more solid footing, then the place to look for an analysis is in extensions to the *LAPSE constraints in (31). For example, the associate editor points out that a constraint against long lapses ($\breve{\sigma}\breve{\sigma}\breve{\sigma}$) word-finally would suffice (though see §2 on locality and counting in phonological constraints).

6.3.3 *The location of main stress.* Gradient alignment has also been used to assign main stress to the rightmost or leftmost foot. In rhythmic stress systems like those exemplified in Table I, main stress usually falls on the leftmost or rightmost foot, which need not be in absolute word-initial or word-final position. In prominence-driven stress systems, main stress falls on the leftmost or rightmost heavy syllable, with no limit on how far it can be displaced from the word edge. Standardly, minimal violation of gradient ALIGN(Hd, Wd) (the erstwhile EDGEMOST) is the source of these effects (McCarthy & Prince 1993b, Prince & Smolensky 1993).

Following Prince (1983), I have proposed that the End Rule constraints (35) are responsible for the location of main stress, and that these constraints give a categorical advantage to the first or last foot. There are two potential challenges to this view. First, prominence-driven stress systems have sometimes been analysed with syllable-counting ALIGN(Hd, Wd) constraints, and the End Rule constraints cannot directly reproduce this effect. Second, foot-extrametricality phenomena (Hayes 1995) appear to show a typical gradient pattern: main stress cannot appear on the final foot, so it goes on the one next to it.

In general, prominence-driven stress systems can be analysed using the categorical End Rule constraints, if the foot structure is properly understood. Prince (1985) and Bakovič (1998) propose that prominence-driven stress systems, which had sometimes been attributed to unbounded feet in the past (Halle & Vergnaud 1978, Hayes 1980), actually involve binary feet. Unlike rhythmic stress systems, though, feet are rather sparse in prominence-driven stress: feet parse all the heavy syllables, and otherwise they parse a pair of light syllables at the default edge. As in rhythmic stress systems, the first or last foot is singled out for main stress. Schematically, the stress patterns are like these:

(40) *Prominence-driven stress*

a. *Default to opposite : rightmost heavy, else leftmost*
LL(H)LL**(H)**LL
(LL)LL

b. *Default to same : leftmost heavy, else leftmost*
LL**(H)**LL(H)LL
(LL)LL

The letters H and L stand for heavy and light syllables, respectively, the parentheses bracket trochaic feet and the head foot is in boldface. In the default-to-opposite system, the last foot in the word takes the main stress; in the default-to-same system, the first foot takes the main stress.

In prominence-driven stress, feet are sparse, appearing only on heavy syllables except for the initial foot in words with no heavy syllables. This economical mode of foot-parsing follows from the stress/weight relating constraint in (41).[27]

(41) SWP

*Ĺ Stressed light syllables are prohibited/if stressed, then heavy.

SWP ensures that no light syllable projects a foot, unless there are no heavy syllables and the word would otherwise go unheaded. In prominence-driven systems like (40), SWP dominates PARSE-σ (and *LAPSE):

(42) SWP ≫ PARSE-σ

	SWP	PARSE-σ
☞ a. LL(H)LL**(H)**LL		******
b. (LL)(H)(LL)**(H)**(LL)	***!	

In words with no heavy syllables (e.g. **(LL)**LL), there is no way to provide the obligatory head foot without violating SWP. Violation is still minimal, though, so such words have only a single foot. It will be initial or final,

[27] SWP has been called Prokosch's Law (Prokosch 1939, Vennemann 1972), Obligatory Branching (Hayes 1980, Hammond 1986) and STRESS-TO-WEIGHT (Fitzgerald 1997). Cf. Prince (1990).

depending on which of the positional PARSE-σ constraints in (33) is ranked higher.

The rankings just presented will derive the right foot-parsing. Deploying the main stress on the first or last foot is then just a matter of whether ER-L or ER-R is ranked higher, as the following partial factorial typology shows:

(43) *Partial factorial typology for prominence-driven stress*

Leftmost heavy, else leftmost:
SWP $\gg$ PARSE-σ_{I} $\gg$ PARSE-σ_{F} *and* ER-L $\gg$ ER-R

Rightmost heavy, else leftmost:
SWP $\gg$ PARSE-σ_{I} $\gg$ PARSE-σ_{F} *and* ER-R $\gg$ ER-L

Leftmost heavy, else rightmost:
SWP $\gg$ PARSE-σ_{F} $\gg$ PARSE-σ_{I} *and* ER-L $\gg$ ER-R

Rightmost heavy, else rightmost:
SWP $\gg$ PARSE-σ_{F} $\gg$ PARSE-σ_{I} *and* ER-R $\gg$ ER-L

This account bears a more than passing resemblance to the analysis of prominence-driven stress in Prince (1983); the only real difference is the somewhat greater reliance on foot structure and foot-parsing constraints.

FOOT EXTRAMETRICALITY, a concept introduced by Hayes (1995: 77–78), presents another situation where gradient alignment of the head foot might seem essential. In the best-attested type, a left-to-right iambic stress system has main stress at the right, but not on a word-final syllable. Odd-parity sequences come out with penult stress (e.g. [(02)(01)0]), but even-parity sequences end up with antepenultimate stress because the final foot is extrametrical for the purpose of main-stress assignment: [(02)(01)⟨(02)⟩], as in Cayuga (tewa)(katá)⟨(wǫnyeʔ)⟩ 'I'm moving about'. Besides Cayuga (Hayes 222), other languages described as having this stress pattern are two varieties of Bedouin Arabic (227, 232), Eastern Ojibwa (216) and two varieties of Delaware (211). There is also a left-to-right trochaic version of the same pattern. It is described for Palestinian Arabic (125), Early Latin (180) and Egyptian Radio Arabic (130).[28]

The foot-extrametricality parsing can be easily and exactly reproduced in a theory with gradient alignment constraints. The constraint NON-FIN(Hd(Wd)) (Prince & Smolensky 1993: ch. 4, (53)), which bans the head

[28] Two right-to-left trochaic versions of the foot-extrametricality pattern are also described by Hayes (1995): 'the Grierson/Fairbanks stress rule' for Hindi (162) and Paamese (178). These cases seem less convincing. There are serious problems establishing what the Hindi stress facts really are (Ohala 1977, Hayes 1995). The foot-extrametricality analysis of Paamese is one approach to a rather tricky and not yet fully understood problem in lexical conditioning of stress (cf. Goldsmith 1990: 215–216). Another example: Buckley (1994) proposes that the invisibility of initial CVV syllables to stress in Kashaya is an effect of initial foot extrametricality. That analysis presents many complications, however, that go well beyond the issue of gradience.

foot from word-final position, must be ranked above gradient ALIGN-(Hd(Wd), R; Wd, R; σ):

(44) *Foot extrametricality pattern with gradient alignment*

	NON-FIN(Hd)	ALIGN(Hd,R)	ALIGN(Hd,L)
☞ a. [(02)(01)(02)]		**	**
b. [(02)(02)(01)]	*!		****
c. [(01)(02)(02)]		****!	

Gradient assessment is necessary to favour (44a) over (44c); if ALIGN(Hd, R) were evaluated categorically, then (44a) and (44c) would tie, and low-ranking ALIGN(Hd, L) would decide the matter, wrongly favouring stress on the first foot (44c) instead of the penultimate foot (44b). This becomes apparent if the gradient ALIGN constraints are replaced by categorical ER-L and ER-R, as in (45).

(45) *Tableau* (44) *with categorical constraints*

	NON-FIN(Hd)	ER-R	ER-L
a. [(02)(01)(02)]		*	*!
b. [(02)(02)(01)]	*!		*
☞ c. [(01)(02)(02)]		*	

With categorical constraints, if the final foot is skipped, then stress appears on the initial foot instead. Indeed, it would seem that the categorical End Rule constraints not only cannot produce the foot-extrametricality pattern, which is attested, but also predict an unattested pattern: penult stress in odd-parity words (e.g. [(02)(01)0]) contrasting with peninitial stress in even-parity words (e.g. [(01)(02)(02)]).

Both of these unwanted predictions turn out to be artefacts of the attempt to reproduce the foot-extrametricality analysis. What's needed is a way of obtaining the descriptive effects of foot extrametricality, though not the construct itself. The key idea is that the putative extrametrical foot is not there at all, so the output forms are actually odd-parity [(02)(01)0] and even-parity [(02)(01)00]. Importantly, the even-parity word ends in a lapse rather than a secondary-stressed foot. The responsible constraint is not NON-FIN(Hd(Wd)), but rather a similar constraint that replaces it, NON-FIN(Ft) (e.g. Kager 1999: 151):

(46) NON-FIN(Ft)

*Ft/ __]$_{Wd}$

'Word-final feet are prohibited.'

If NON-FIN(Ft) is ranked above *LAPSE, then the last two syllables will remain unfooted in even-parity words:

(47) *Iambic 'foot extrametricality' pattern categorically*

	NON-FIN(Ft)	*LAPSE	LAPSE-AT-END
☞ a. [(02)(01)00]		*	
b. [(02)(02)(01)]	*!		
c. [(02)(01)(02)]	*!		
d. [(02)0(01)0)]		*	*!

NON-FIN(Ft) rules out the full-parsing candidates (47b, c). The surviving candidates (47a, d) have a stress lapse. With LAPSE-AT-END anywhere in the hierarchy, this lapse must fall on the last two syllables. Since the last two syllables are adjacent to both the peak and the end, there is no better place for that lapse to go.

The winning candidate, (47a), has all of the observed descriptive effects of foot extrametricality without the extrametrical foot. The extrametrical foot is an analytic artefact, freely dispensed with when a better analysis comes along. Indeed, in none of the foot-extrametricality languages do we find reports of a word-final secondary stress as overt evidence of the extrametrical foot.

This line of analysis provides a categorical treatment of the foot-extrametricality pattern. It also avoids the problem noted below (45); with NON-FIN(Ft) replacing NON-FIN(Hd(Wd)), there is no way of getting (45c) to win in a language with penult stress in odd-parity words.

Another type of foot extrametricality is also robustly exemplified in Hayes (1995). This is extrametricality in clash. When moraic trochees parse HLL sequences, they produce a stress clash. According to Hayes, Bani-Hassan Arabic (366), Maithili (153), Manam (182) and Turkish (262) declare the final foot to be extrametrical when it is in clash with the penultimate foot: [**(H)**⟨(LL)⟩] (*vs.* clashless [L**(LL)**]).

Buckley (1998) presents an analysis of Manam that avoids foot extrametricality or its OT equivalent. In words of the form [**(H)**LL], there is a final lapse because *CLASH disfavours the alternative, *[(H)**('LL)**]. In addition, WSP rules out the opposite way of resolving the clash, *[H**(LL)**].

Even within the Hayes (1995) framework, foot extrametricality in clash seems like a dubious move. In Hayes' general schema for extrametricality rules (p. 58), the only context that can be mentioned is the right edge of some domain, such as the word, and he notes that even this much contextual information is redundant, because of factors like the Peripherality Condition (Harris 1983). A rule assigning extrametricality only in clash is therefore a big leap in expressive power, and this ought to encourage scepticism. In fact, there is good reason to be sceptical, since, as we have just seen, a plausible alternative exists.

Before closing out this section, two other potential cases of foot extrametricality should be mentioned. First, Hayes (1995: 262–263) very briefly describes a right-to-left iambic stress pattern with foot extrametricality in clash. The languages, all related (two very closely), are Javanese, Malay and Sarangani Manobo. Stress falls on the penult unless it contains schwa, in which case stress falls on the ultima. The proposal is that all syllables are heavy except those containing schwa and that feet are iambic, with the final foot extrametrical in clash: [(H)**(H)**⟨(H)⟩], [(H)**(H)**L], [(H)**(LH)**] and [(H)**(LL)**].

There are several reasons to doubt this analysis. Indonesian, a very close relative of Javanese and Malay, is clearly trochaic (Cohn 1989, Cohn & McCarthy 1994) and a variant of Sarangani Manobo is also trochaic (Hayes 1995: 178–180). A feature of Indonesian and presumably these other languages is a well-documented dispreference for stressed schwa (Urbanczyk 1996, de Lacy 2002). Final stress in words like Javanese [bənə́r] 'correct' and [gatələ́n] 'itch' may simply reflect a bias toward stressing closed syllables when stressed schwa is unavoidable.

Second, the assignment of primary stress in English should be mentioned as a potential counterexample to the claims made here, as Joe Pater has pointed out. According to 'Schane's Rule' (Schane 1972), primary stress goes on the rightmost non-final stressed syllable. This generalisation evolved into the Lexical Category Prominence Rule of Liberman & Prince (1977), which assigns main stress to the final foot unless it is monosyllabic (non-branching), in which case main stress goes on the penultimate foot: (Àga)(mém)(nòn) *vs.* (Hàli)(càr)(nássus).

This principle for assigning main stress falls well outside the typology documented in Hayes (1995), and it is not surprising that it has received little attention in Optimality Theory. Further problems, all well known, are presented by disyllables, which follow other generalisations, by outright exceptions like *Ládefòged*, and by the many forms that require a highly opaque derivation to conform to the generalisation: *cómmentàry*, *oblígatòry*, *nárcolèpsy*, *sálamànder*, *cáterpìllar*, *Áristòtle*, *púmpernìckel*, etc. On the whole, then, English primary stress seems much more like a research problem than a *prima facie* counterexample to categoricality.

6.4 Summary

Gradient alignment constraints are ubiquitous in analyses of metrical phonology within OT. Nonetheless, the case for gradient alignment of feet and word-heads is not persuasive. Directionally iterative foot-parsing has been persuasively argued by Kager (2001) to reflect constraints on lapses and clashes rather than gradient constraints on foot alignment. *LAPSE and its congeners provide a better match between prediction and observation than the gradient foot-parsing constraints ALLFTL/R. Nor does the evidence from one-foot-per-word and main-stress phenomena provide support for gradient constraints; the principal patterns can be analysed with simple categorical constraints modelled after the End Rule of Prince (1983).

7 Alignment in autosegmental phonology

7.1 Introduction

Perhaps the last bastion of gradient alignment is the autosegmental phonology of features and tones. Gradient alignment constraints have been applied to three kinds of autosegmental processes, docking, flop and spreading. In docking, a featural or tonal morpheme that is not associated with a segment or syllable in underlying representation becomes linked in the output. The gradience of standard alignment constraints has been used to produce docking that is near a word edge, even if it does not lie exactly at the edge. In flop, a feature or tone is reassociated onto a different segment or syllable. With gradient alignment, the flopped feature or tone may be attracted toward a word edge without actually reaching it. And in spreading, an element that is linked to one segment or syllable in underlying representation extends its domain in one or both directions, often over an unbounded distance. Gradient alignment has been identified as the constraint that compels spreading by forcing a feature or tone to extend its reach toward a word edge, even when the edge itself is unattainable for other reasons.

It is obviously impossible within the scope of this article to locate and reanalyse every known application of gradient alignment to docking, flop and spreading phenomena. Since that cannot be done, I will pursue the more modest goal of proposing alternatives, illustrating them with a few examples, and highlighting any empirical differences from gradient alignment.

7.2 Docking

Docking is what happens to a floating feature or tone. Floating elements are typically affixal; like other affixes, they tend toward one edge of the word or the other, but may be displaced from it for phonological reasons. Two examples are given in (48); see Akinlabi (1996), Zoll (1996) and Piggott (2000) for general discussion.

(48) *Feature docking effects*

a. *Japanese palatal prosody* (Hamano 1986, Mester & Itô 1989, Zoll 1997)

mimetic	*'uncontrolled'*	
meta-meta	meča-meča	'destroyed'
toko-toko	čoko-čoko	'childish small steps'
poko-poko	p^yoko-p^yoko	'jumping around imprudently'
dosa-dosa	doša-doša	'in large amounts'
	*ǰosa-ǰosa	

b. *Chaha labialisation* (Leslau 1950, 1967, McCarthy 1983, Banksira 1997, Rose 1997)

3SG PERF MASC	with 3SG MASC OBJ	
dänäg	dänäg^{w}	'hit'
näkäb	näkäb^{w}	'find'
näkäs	näk^{w}äs	'bite'
bäkär	bäk^{w}är	'lack'
mäsär	m^{w}äsär	'seem'

In Japanese mimetic words, palatalisation is used to mark 'uncontrolledness' (Hamano 1986). In a mimetic stem C_1VC_2V, C_2 is affected if it is a coronal (except *r*); otherwise C_1 is affected. When there are two coronals, as in [doša-doša], the second one is affected. In Chaha verbs, labialisation is the mark of a 3rd person masculine object. Labialisation falls on the rightmost labialisable consonant in the root. The labialisable consonants are the dorsals and primary labials, which have labialised counterparts in Chaha's consonant inventory. Coronals cannot be labialised, either morphologically or in the inventory as a whole. This shows that the relevant constraint, call it $*T^{w}$, is undominated.

Japanese palatalisation and Chaha labialisation both involve affixes consisting of just a single feature. As affixes, their position in the word is determined by the categorical PREFIX and SUFFIX constraints defined in (17). Consider Japanese first. The 'uncontrolled' affix [−anterior] is a formal suffix, so its position is determined by the ranking of SUFFIX([−ant]) and SUFFIX/σ([−ant]). Perfect satisfaction of SUFFIX([−ant]) is impossible unless a full segment is epenthesised finally to bear this feature; we may assume that DEP forecloses this possibility. Since [−anterior] must be realised on a consonant, and mimetics are all of the form C_1VC_2V, the best it can aspire to is docking on C_2. This satisfies SUFFIX/σ([−ant]), as shown in (49):

(49) *Role of* SUFFIX/σ([−ant]) *in Japanese*

/dosa+[−ant]/	SUFFIX([−ant])	SUFFIX/σ([−ant])
☞ a. doša	*	
b. joša	*	*!

Both candidates have an unaligned suffix, so they tie on the unmodified SUFFIX constraint. Wherever it is ranked, SUFFIX/σ([−ant]) correctly resolves this tie in favour of palatalising C_2.

When C_2 is not a coronal, however, violation of SUFFIX/σ([−ant]) is compelled, as shown by [p^{y}oko-p^{y}oko]. Zoll (1997) proposes that non-coronals are treated differently because, unlike palatalised coronals, palatalised non-coronals are complex segments p^{y}, k^{y}, etc., and complex segments are marked non-initially. The responsible constraint, in Zoll's

(1996) terms, in COINCIDE(Complex Seg, Wd, L).[29] It crucially dominates SUFFIX/σ([−ant]):

(50) COINCIDE(Comp Seg, Wd, L) ≫ SUFFIX/σ([−ant])

/poko+[−ant]/	COINCIDE	SUFFIX/σ([−ant])
☞ a. p^{y}oko		*
b. pok^{y}o	*!	

Japanese mimetic stems are never longer than CVCV, so SUFFIX/σ([−ant]) is sufficient to account for the right-edge bias in where to dock the palatalisation morpheme.

Chaha is similar to Japanese. The labial morpheme [+round] is a formal suffix, so it likewise is governed by SUFFIX([+rd]) and SUFFIX/σ([+rd]). Both constraints are satisfied if the final consonant is labialised, as in [näkäb^{w}]. Since [+round] coronals are excluded by undominated *T^{w}, a final coronal will force violation of SUFFIX([+rd]), though SUFFIX/σ([+rd]) is still satisfied: [bäk^{w}är]. Crucially, SUFFIX/σ([+rd]) prefers [bäk^{w}är] to *[b^{w}äkär]. When both of the last two consonants are coronals, violation of both SUFFIX constraints is unavoidable: [m^{w}äsär]. The distribution of morphological labialisation in these triliteral roots has been accounted for with categorical constraints.

Quadriliteral roots challenge this result, however. In quadriliteral stems of the form $L_1VL_2VN_3VN_4$, where *L* stands for any labialisable consonant and *N* for any non-labialisable one, the categorical constraints SUFFIX([+rd]) and SUFFIX/σ([+rd]) do not determine whether L_1 or L_2 should be labialised. Previous analyses, relying on the example [kəb^{w}äšäš] 'entangle the fibre' (Leslau 1967),[30] conclude that the attraction of [+round] to the rightmost labialisable consonant persists even with greater distance from the right edge of the stem.

There are two alternative interpretations of this fact. First, SUFFIX/Ft([+rd]) might be decisive, since Chaha reportedly has penult stress (Li 2002, citing a personal communication from Degif Petros Banksira). This constraint favours [kə(b^{w}äšäš)$_{Ft}$] over [k^{w}ə(bäšäš)$_{Ft}$] under the indicated foot-parsing. A complication: the position of labialisation is determined relative to the stem edge, while stress is presumably located relative to the word edge, and these will not always coincide, because of suffixation, so an analysis requiring OO-faith or strata would be required. Second, it is conceivable that segmental markedness favours [kəb^{w}äšäš]. When two candidates tie on SUFFIX([+rd]) and SUFFIX/σ([+rd]), even low-ranking segmental markedness constraints can emerge to settle the dispute. If such

[29] COINCIDE constraints are described by Zoll as the result of conjoining (in the sense of Smolensky 1995) ALIGN and a markedness constraint, such as the one against complex segments. Since the conjunction of two constraints, even if one is gradient, is necessarily a categorical constraint (see §2), COINCIDE is categorical.

[30] This verb form is an impersonal, not a 3rd masculine singular object, but the distribution of labialisation is the same.

a constraint (or constraint ranking) regards labialised velars as more marked than labialised labials, then there is no argument for alignment: [kəb^{w}äšäš] is superior to *[k^{w}əbäšäš] for segmental markedness reasons, not because of the position of labialisation.[31] The test of this analysis is to find a similar root, but with the dorsal and labial reversed. An exhaustive search of Leslau's (1979) Chaha dictionary has unearthed five other $L_1L_2N_3N_4$ words, and all of them have the same dorsal–labial order as [kəb^{w}äšäš], so they neither favour nor disfavour the categorical approach.

The Chaha example, though not entirely probative in itself, highlights an empirical difference between the categorical PREFIX and SUFFIX constraints in (17) and gradient alignment. Affixes that consist of floating features or tones are typically docked onto the adjoining root. When the affix is pushed away from its preferred edge for phonological reasons, if the root is big enough, then eventually the PREFIX or SUFFIX constraints will fail to distinguish among the viable candidates. The prediction of the categorical theory is that other markedness constraints, even low-ranking ones, will make the decision instead. The prediction of the gradient theory is that ALIGN's edge bias will be felt no matter how big the root gets. (As we will see in the next section, a similar prediction is made for flop processes, with dubious results following from gradience.)

The Chaha example also suggests another empirical difference between the categorical and gradient approaches. Under the categorical theory, there can in principle be featural or tonal morphemes that work like the infixes in Tagalog and Nakanai (§5): they are permitted close to an edge, but forbidden to migrate further. Both the gradient and categorical approaches can deal with the most common situation, where the only licit docking site is right at the edge. (All of the mutation systems discussed in Lieber 1987: ch. 2 are of this type.) But only the categorical approach can analyse a language that is like Chaha in all relevant respects except that labialisation never gets past the final syllable, so only stems like [mäsär] and [kəbäšäš] could not be labialised in the relevant morphological category. I have not encountered an example of this type, though it should be noted that there are very few cases of docking that resemble Chaha at all.

7.3 Flop

Flop is the reassociation of a feature or tone from one segment or syllable to another. In the flop processes of interest here, the feature or tone is attracted toward the word edge or the head foot, though it may not always make it there.

In Cuzco Quechua (Parker & Weber 1996, Parker 1997, MacEachern 1999), the three-way lexical contrast plain/aspirated/glottalised is possible only in the leftmost non-coda obstruent stop of the root.

[31] In their analysis of Chaha's close relative Ennemor, Hetzron & Habte (1966: 28) observe: 'la seule chose concrete qu'on peut noter au sujet des verbes quadrilitères est que *g* y semble résister à la labialisation plus que les autres consonnes'.

(51) *Cuzco Quechua laryngeal contrasts*

q^{h}ata	'mountainside'
hap^{ʔ}iy	'to light (a fire)'
lap^{ʔ}ay	'to lick up (said of dog)'
warak^{ʔ}a	'sling made of wood'

The restriction of glottalisation and aspiration to obstruent stop onsets is unremarkable; the responsible markedness constraints are undominated in Cuzco Quechua. The important thing for present purposes is that words like Parker & Weber's hypothetical *[poq^{ʔ}a] are prohibited; it would have to be realised as [p^{ʔ}oqa] instead.

This looks like an obvious case for gradient alignment (Parker 1997). ALIGN(Laryngeal, L; Stem, L; Seg) will correctly favour [p^{ʔ}oqa] over *[poq^{ʔ}a]. There is an alternative, however: the same result can be obtained from a categorical COINCIDE constraint requiring aspirated and glottalised consonants to be literally initial in the root: COINCIDE (Laryngeal, Root, L). That constraint, ranked above the faithfulness constraint NOFLOP (McCarthy 2002a), which bans featural reassociation, will also correctly favour [p^{ʔ}oqa] over *[poq^{ʔ}a]:

(52) COINCIDE(Lar, Rt, L) ≫ NOFLOP

/poq^{ʔ}a/	COINCIDE	NOFLOP
☞ a. p^{ʔ}oqa		*
b. poq^{ʔ}a	*!	

Of course, examples like [lapt^{ʔ}ay] and [warak^{ʔ}a] show that perfect coincidence is not always achieved. These examples, though, present no better locus for the laryngeal contrast, because sounds like [l^{ʔ}] and [w^{ʔ}] are not found in this language. Since these words violate COINCIDE, it must be crucially dominated, as (53) shows.

(53) MAX(Lar), *[+cont, +glot] ≫ COINCIDE(Lar, Rt, L)

/warak^{ʔ}a/	MAX(Lar)	*[+cont,+gl]	COINCIDE
☞ a. warak^{ʔ}a			*
b. waraka	*!		
c. w^{ʔ}araka		*!	

Form (53b) vacuously satisfies COINCIDE by eliminating the offending feature. This breach of faithfulness is forbidden by high-ranking MAX (Laryngeal), however. Form (53c) non-vacuously satisfies COINCIDE, but at the intolerable cost of violating an undominated markedness constraint.

This analysis presupposes a description of Cuzco Quechua that goes something like this: glottalisation and aspiration fall on the root-initial consonant; if they cannot because the root-initial consonant is not an obstruent stop, they fall on any other non-coda obstruent in the root.

Gradient alignment, as in Parker's (1997) analysis, presupposes Parker & Weber's (1996) description: the laryngeally marked consonants 'are always the first syllable-initial stop of the root'. The difference turns, as in Chaha, on the prospect of finding longer roots with the right arrangement of consonants.

The crucial test-cases will have the form RV(C)SVSV, where *R* stands for any consonant other than an obstruent stop, and *S* is an obstruent stop, MacEachern cites no relevant data; Parker, though, has several possible examples:

(54) *Possible examples of gradient alignment in Cuzco Quechua*

moqčˀikuy	'to wash or rinse out the mouth'
musp^hapakuy	'to be delirious'
rankˀukuy	'to get twisted up'
ak^hakaw	'how hot it is!'

At first glance, these examples look like support for gradient alignment: when there are two or more non-initial stops, the first of them bears the laryngeal feature: [moqčˀikuy], *[moqčikˀuy]. Categorical COINCIDE will not make this distinction.

This argument omits a crucial step, however. For these examples to be convincing, it is necessary to show that they are unanalysable roots. Compounds are not useful evidence, because each member of the compound is a separate domain for the purposes of the assigning the laryngeal features (MacEachern 1999: 32). Nor are suffixed forms useful, because, according to Parker & Weber (1996), glottalised and aspirated stops 'occur only in roots, never in suffixes'. If [moqčˀikuy] includes a suffix *-kuy*, then this form is useless for testing gradient *vs.* categorical alignment; the non-glottalisation of *k* is adequately explained by the general ban against glottalisation in suffixes.

The fact that three of the four words in (54) end in *-kuy* should excite suspicion. The suspicion is confirmed by two observations. First, a machine-readable dictionary of Bolivian Cuzco (accessed 10 April 2003 at http://www.runasimi.de) provides transparently related forms that lack the putative *-kuy* suffix: [moqčˀiy] 'to fill one's mouth with water', [musp^hay] 'to be confused'. Second, in that same dictionary well over 600 verbs end in *-kuy*, a sound sequence that is otherwise unusual (e.g. no words begin in *kuy-*). Impressionistically, many of these 600 verbs have transparently related forms that lack *-kuy*. So it is surely a suffix.

I did a hand-search of all glottalised stops in the first half of the machine-readable dictionary, eliminating forms that, on semantic or morphological grounds, appeared to have only one stop in the root. This left just eight words with the requisite RV(C)SˀVSV pattern. Nearly all are flora and fauna terms, so they too may be morphologically complex in the same semantically opaque way as, say, *daddy longlegs*.

It seems clear that Cuzco Quechua is not going to offer decisive evidence either way. A particular difficulty is that this is a static condition, not a source of alternations, so the argument is perforce statistical. In

other words, it's not only necessary to display some morphologically simplex RV(C)S^{ʔ}VSV words; it is also necessary to display enough of them so that the absence of RV(C)SVS^{ʔ}V words is remarkable.

Cuzco Quechua does, however, reveal a typological prediction of gradient alignment theory. In Quechua, no root can contain more than one laryngeally marked consonant – *C^{ʔ}VC^{ʔ}V, *C^{h}VC^{h}V, *C^{ʔ}VC^{h}V, etc. But imagine a language that is identical to Quechua except that this restriction is not in force. Gradient alignment predicts that the glottalisation and aspiration will line up on onset obstruent stops as they occur in the root from left to right: [p^{ʔ}oq^{h}at^{ʔ}aka] *vs.* *[poq^{ʔ}at^{h}ak^{ʔ}a], *[p^{ʔ}oqat^{h}ak^{ʔ}a]. The COINCIDE analysis does not make this prediction, since it favours initial position but not close-to-initial position for laryngeal contrasts.

This hypothetical example may be hard to imagine for segmental contrasts, but it is easy to construct a tonal case. In Chichewa verb stems (Myers & Carleton 1996: 42–49), if any morpheme bears an underlying high tone, then that tone is flopped onto the ultima (e.g. /támbalal-a/ →[tambalalá] 'stretch out your legs!'). When there is more than one high tone in the input verb stem, all delete except for one: /támbalal-íts-a/ →[tambalalitsá], *[tambalalítsá] 'really stretch out your legs!'.

Myers & Carleton propose to derive the 'all H delete except one' generalisation by ranking the tonal alignment constraint above MAX(tone):

(55) ALIGN(H, R; St, R; σ) ≫ MAX(tone), NOFLOP(tone)

/támbalal-íts-a/	ALIGN(H,St)	MAX(tone)	NOFLOP(tone)
☞ a. tambalalitsá		*	*
b. tambalalítsá	*!		**
c. támbalalítsa	*****!		

The alternative advanced here says that H is attracted to the final syllable by COINCIDE(H, Stem, R). This constraint rules out the same candidates non-gradiently.[32]

Now, imagine a language identical to Chichewa except that the ranking of MAX(tone) and ALIGN(H, R; Stem, R; σ) is reversed. In this hypothetical language, all high tones should pile up at the right edge of the stem, so every stem will end in a sequence of high-toned syllables equal in length to the number of high tones in the input: /í-má-kú-támbalal-a/ →[imakutámbálálá]. Although this pattern is readily predicted by gradient alignment, there is no way to get it using categorical COINCIDE (H, Stem, R). No categorical constraint will, for example, induce the first H to move from the initial syllable to the preantepenult, simply because the preantepenult is closer to the final or head syllable.

[32] In certain tenses, Chichewa flops the high tone onto the penult. In that case, COINCIDE(H, Hd(Wd)) is the active constraint, under the assumption that Chichewa has a right-aligned trochaic foot, as suggested by reduplicated forms like [chikulupiriro-ríro] 'real faith' and parallels elsewhere in Bantu (Myers 1987).

I know of no language that displays the tonal or featural piling-up effect that is predicted by gradient alignment. If indeed no such language exists, then it must count against the gradient theory and in favour of categoricality, since the existence of such patterns is an unavoidable prediction of gradient constraints aligning tones and features. The absence of such pile-up effects in phonology is all the more striking because they have been well documented in work on OT syntax (Legendre 1999, 2000, Gouskova 2001, Grimshaw 2001).

There is a broader moral to be drawn here, one that emerges from Zoll's (1996: 141) discussion of a different but related range of examples: 'licensing of marked structure never involves an injunction to be as close to a strong position as possible. Rather, licensing *always* constitutes an all-or-nothing proposition whereby marked structures are licit in licensed positions but ill-formed everywhere else.' If something like this remark is correct, as I have argued here, then categorical constraints truly are the right way to deal with positional markedness restrictions on features and tones.

7.4 Spreading

In Kirchner (1993), Smolensky (1993), Archangeli & Pulleyblank (1994a), Cole & Kisseberth (1995), Pulleyblank (1996) and much other work, gradient alignment constraints are assigned primary responsibility for autosegmental spreading of features and tones. Imagine, for example, a language where all H tones spread rightward to the end of the word, except that the OCP prevents any H from spreading onto a syllable that precedes another H. (This is approximately Shona (Myers 1987, 1997), with low-toned syllables analysed as toneless.) The gradient constraint ALIGN(H, R; Wd, R; σ) ensures that each H tone maximises its spreading domain by counting the syllables between the right edge of each high-tone domain and the end of the word, as shown in (56).

(56) *Gradient* ALIGN(H, R; Wd, R; σ) *in tone spreading*

H H σσσσσσσ	OCP	ALIGN(H,Wd)
☞ a. H H σσσσσσσ		****
b. H H σσσσσσσ		********!
c. H H σσσσσσσ	*!	***
d. H H σσσσσσσ		******!

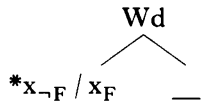
ALIGN's behaviour in this example is virtually identical to the behaviour of ALLFTR in (3). Both constraints gradiently evaluate each instance of a tone or foot for its distance in syllables from the right word edge. Not just the rightmost tone but all tones must be checked to ensure the victory of (56a) over (56d). Furthermore, just like (3), the winning candidate (56a) abundantly violates the responsible alignment constraint; it simply performs better on alignment than the viable alternatives.

The hypothetical example (56) was chosen because it fairly represents some of the problems that a theory of spreading processes must address. As I showed in the discussion of (3) in §2, there is no way to use categorical constraints to mimic directly the effects of gradient alignment in cases like this. I therefore explore a more distant alternative that is still capable of making the crucial distinctions in (56) and similar cases.

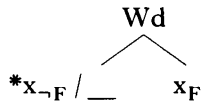
The constraints MATCH-R(F) and MATCH-L(F) demand agreement in F-value between a segment or syllable and any preceding/following segment (in the case of features) or syllable (in the case of tones). They are defined as follows, where F is a feature or tone and x is a segment or syllable:

(57) MATCH *constraints*

a. MATCH-R(F)



b. MATCH-L(F)



These constraint formulations fit the categorical markedness constraint scheme (1) and they observe the notational conventions introduced with (17). Accordingly, strict adjacency between x_F and $x_{\neg F}$ is not required, but shared membership in the prosodic word or some other constituent is required.

Applied to (56), MATCH-R(H) imposes the same harmonic ordering on candidates as ALIGN(H, R; Wd, R; σ) does:[33]

[33] MATCH evaluates the candidates in (58) under the assumption that a toneless syllable, which is effectively low-toned, mismatches a high-toned syllable. See Archangeli & Pulleyblank (1994b: 105–106) on why such an assumption is a necessary concomitant of underspecified representations.

(58) *Tableau* (56) *with categorical* MATCH-R(H)

H H σσσσσσσ	OCP	MATCH-R(H)
☞ a. H H σσσσσσσ		*
b. H H σσσσσσσ		*****!
c. H H σσσσσσσ	*!	
d. H H σσσσσσσ		***!

The absolute number of violations is different, but the relative harmony of candidates according to MATCH-R(H) is the same as their relative harmony according to ALIGN(H, R; Wd, R; σ): total spreading (58c) is best, then spreading up the OCP limit (58a), then spreading of only one tone (58d) and finally no spreading at all (58b).

The MATCH constraints are able to achieve this result because they combine the best properties of two other alternatives to gradient alignment in the literature. First, like the AGREE constraint in Baković (2000) (see also Eisner 1999, Lombardi 1999, 2001a), the MATCH constraints are not formulated autosegmentally. They require matching feature values, but not shared linkage to a single featural autosegment. The problem with a shared linkage condition is that it will not work unless the constraint quantifies universally over all instances of that autosegment and over all segments or syllables that ought to link to it. This double universal quantification is a hallmark of alignment (see (2)) and cannot be reconciled with the more restrictive constraint schema in (1).

Second, like the constraints variously named SPREAD(F) (Padgett 1995b), SPECIFY(T) (Myers 1997: 861–863) and EXTEND(F) (Kaun 1995: 98), the MATCH constraints are non-local: they do not require adjacency between the pair of segments or syllables being evaluated. The strictly local constraint AGREE is unable to favour spreading that does not fully succeed, since it assigns one mark for each pair of adjacent disagreeing segments or syllables. Thus, AGREE(H) wrongly assigns equal marks, one each, to (58a) and (58d).[34]

The MATCH constraints by themselves are not sufficient to capture all of the observed properties of autosegmental spreading, but neither are the

[34] To circumvent this problem, a strictly local agreement constraint must be supplemented with some mechanism to ensure that spreading iterates (as in Eisner 1999, Baković 2000, McCarthy 2002a).

alternatives. In particular, just like ALIGN and SPREAD, MATCH requires additional constraints to favour candidates where syllables or segments are not skipped over. These constraints have been well studied as part of an overall programme of reducing all phonological assimilation to strict locality (McCarthy 1994, Gafos 1996, 1998, Walker 1998, Ní Chiosáin & Padgett 2001 and others).

Needless to say, this brief review of feature and tone spreading is not sufficient to do justice to this broad and highly productive area of research. Nonetheless, it has proven possible to present a plausible categorical alternative to gradient alignment as the source of autosegmental spreading.

8 Conclusion

In this article, I have argued for a particular view of how OT constraints work. Constraints militate against structural configurations (markedness constraints) or non-identical mappings (faithfulness constraints). Constraints do so *categorically*: it is sufficient for any constraint to assign one violation-mark for each instance of the marked structure or unfaithful mapping in the candidate under evaluation. The definitional frame of a constraint, then, is 'Assign one violation-mark for every λ meeting condition C', where λ is an output structure or a non-identical mapping. No greater complexity of constraint definition is required or desirable.

This proposal stands at odds with a widely accepted view of OT constraints, that some are categorical and some are gradient. Gradient constraints assess goodness of fit over some range. In a review of the literature, two main types of gradience were identified, those constraints where the range is bounded and those where it is unbounded. Bounded gradience is met with sporadically in the OT literature, in certain constraints on hierarchies, scales and classes. Bounded gradience is unnecessary; any boundedly gradient constraint can, and in some cases must, be replaced by a set of categorical constraints (§3). Unbounded gradience is in all likelihood limited to alignment constraints, which have been important in analysing infixation, stress and various autosegmental processes.

I have argued (§§5–7) that gradient alignment constraints can be dispensed with because their effects are subsumed by other, categorical constraints, many of which have been previously proposed and independently motivated. Moreover, in some cases gradient alignment predicts patterns that are not observed, and these unwanted predictions can be avoided by adopting categorical constraints instead.

REFERENCES

Akinlabi, Akinbiyi (1996). Featural affixation. *JL* **32**. 239–289.

Alber, Birgit (1998). Stress preservation in German loan-words. In Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and morphology of the Germanic languages*. Tübingen: Niemeyer. 113–141.

Archangeli, Diana & Douglas Pulleyblank (1994a). Kinande vowel harmony: domains, grounded conditions, and one-sided alignment. Ms, University of Arizona & University of British Columbia.

Archangeli, Diana & Douglas Pulleyblank (1994b). *Grounded phonology*. Cambridge, Mass.: MIT Press.

Aronoff, Mark, Azhar Arsyad, Hasan Basri & Ellen Broselow (1987). Tier configuration in Makassarese reduplication. *CLS* **23:2**. 1–15.

Austin, Peter (1981). *A grammar of Diyari, South Australia*. Cambridge: Cambridge University Press.

Baković, Eric (1998). Unbounded stress and factorial typology. In Ron Artstein & Madeleine Holler (eds.) *RuLing Papers 1: Working Papers from Rutgers University*. New Brunswick: Department of Linguistics, Rutgers University. 15–28. Reprinted in McCarthy (2003b). 202–214.

Baković, Eric (2000). *Harmony, dominance, and control*. PhD dissertation, Rutgers University. Available as ROA-360 from the Rutgers Optimality Archive.

Baković, Eric & Colin Wilson (2000). Transparency, strict locality, and targeted constraints. *WCCFL* **19**. 43–56.

Banksira, Degif Petros (1997). *The sound system of Chaha*. PhD dissertation, UQAM.

Bat-El, Outi (1996). Selecting the best of the worst: the grammar of Hebrew blends. *Phonology* **13**. 283–328.

Beckman, Jill, Laura Walsh Dickey & Suzanne Urbanczyk (eds.) (1995). *Papers in Optimality Theory*. Amherst: GLSA.

Benua, Laura (1997). *Transderivational identity: phonological relations between words*. PhD dissertation, University of Massachusetts, Amherst. Published 2000 as *Phonological relations between words*. New York: Garland.

Blevins, Juliette (1992). Review of Halle & Vergnaud (1987). *Lg* **68**. 159–165.

Boersma, Paul (1998). *Functional phonology: formalizing the interactions between articulatory and perceptual drives*. The Hague: Holland Academic Graphics.

Bonet, Eulàlia, Maria-Rosa Lloret & Joan Mascaró (2003). Phonology–morphology conflicts in gender allomorphy: a unified approach. Handout of paper presented at the 2003 GLOW colloquium, Lund.

Broselow, Ellen (1982). On predicting the interaction of stress and epenthesis. *Glossa* **16**. 115–132.

Broselow, Ellen & John McCarthy (1983). A theory of internal reduplication. *The Linguistic Review* **3**. 25–88.

Buckley, Eugene (1994). Persistent and cumulative extrametricality in Kashaya. *NLLT* **12**. 423–464.

Buckley, Eugene (1997). Explaining Kashaya infixation. *BLS* **23**. 14–25.

Buckley, Eugene (1998). Alignment in Manam stress. *LI* **29**. 475–496.

Carlson, Katy (1998). Reduplication and sonority in Nakanai and Nuxalk. *Proceedings of the Eastern States Conference on Linguistics* **14**. 23–33.

Carpenter, Angela (2002). Noncontiguous metathesis and adjacency. In Carpenter *et al.* (2002). 1–26.

Carpenter, Angela, Andries Coetzee & Paul de Lacy (eds.) (2002). *Papers in Optimality Theory II*. Amherst: GLSA.

Chomsky, Noam (1965). *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Cohn, Abigail (1989). Stress in Indonesian and bracketing paradoxes. *NLLT* **7**. 167–216.

Cohn, Abigail (1992). The consequences of dissimilation in Sundanese. *Phonology* **9**. 199–220.

Cohn, Abigail & John J. McCarthy (1994). Alignment and parallelism in Indonesian phonology. Published 1998 in *Working Papers of the Cornell Phonetics Laboratory* **12**. 53–137.

Cole, Jennifer & Charles Kisseberth (1995). Nasal harmony in Optimal Domains Theory. Ms, University of Illinois. Available as ROA-49 from the Rutgers Optimality Archive.

Crowhurst, Megan (1996). An optimal alternative to Conflation. *Phonology* **13**. 409–424.

Crowhurst, Megan & Mark Hewitt (1995). Directional footing, degeneracy, and alignment. *NELS* **25**. 47–61.

Crowhurst, Megan & Mark Hewitt (1997). Boolean operations and constraint interactions in Optimality Theory. Ms, University of North Carolina & Brandeis University. Available as ROA-229 from the Rutgers Optimality Archive.

Dekkers, Joost, Frank van der Leeuw & Jeroen van de Weijer (eds.) (2000). *Optimality Theory: phonology, syntax, and acquisition*. Oxford: Oxford University Press.

de Lacy, Paul (1998). Sympathetic stress. Ms, University of Massachusetts, Amherst. Available as ROA-294 from the Rutgers Optimality Archive.

de Lacy, Paul (1999). Circumscriptive morphemes and haplologizing reduplicants. In Catherine Kitto & Carolyn Smallwood (eds.) *Proceedings of AFLA VI: The 6th Meeting of the Austronesian Formal Linguistics Association*. Toronto: Department of Linguistics, University of Toronto. 107–120.

de Lacy, Paul (2002). *The formal expression of markedness*. PhD dissertation, University of Massachusetts, Amherst. Available as ROA-542 from the Rutgers Optimality Archive.

Eisner, Jason (1999). Doing OT in a straitjacket. Ms, UCLA. Available July 2003 at http://www.cs.jhu.edu/~jason/papers/#ucla99.

Elenbaas, Nine & René Kager (1999). Ternary rhythm and the Lapse constraint. *Phonology* **16**. 273–329.

Ellison, Mark T. (1994). Phonological derivation in Optimality Theory. In *Proceedings of the 15th International Conference on Computational Linguistics* (*COLING*). Kyoto. 1007–1013.

Fitzgerald, Colleen M. (1997). *O'odham rhythms*. PhD dissertation, University of Arizona. Available as ROA-190 from the Rutgers Optimality Archive.

Franks, Steven (1989). The monosyllabic head effect. *NLLT* **7**. 551–563.

Fulmer, Lee S. (1997). *Parallelism and planes in Optimality Theory: evidence from Afar*. PhD dissertation, University of Arizona. Available as ROA-189 from the Rutgers Optimality Archive.

Gafos, Diamandis (1996). *The articulatory basis of locality in phonology*. PhD dissertation, Johns Hopkins University.

Gafos, Diamandis (1998). Eliminating long-distance consonantal spreading. *NLLT* **16**. 223–278.

Goldsmith, John (1990). *Autosegmental and metrical phonology*. Oxford & Cambridge, Mass.: Blackwell.

Gordon, Matthew (2002). A factorial typology of quantity-insensitive stress. *NLLT* **20**. 491–552.

Gouskova, Maria (2001). Split scrambling: barriers as violable constraints. In Graham Horwood & Se-Kyung Kim (eds.) *RuLing Papers II: Working Papers from Rutgers University*. New Brunswick: Department of Linguistics, Rutgers University. 49–82.

Gouskova, Maria (2003). *Economy of representation in Optimality Theory*. PhD dissertation, University of Massachusetts, Amherst.

Green, Thomas & Michael Kenstowicz (1995). The Lapse constraint. Ms, MIT. Available as ROA-101 from the Rutgers Optimality Archive.

Grimshaw, Jane (2001). Optimal clitic positions and the lexicon in Romance clitic systems. In Géraldine Legendre, Jane Grimshaw & Sten Vikner (eds.) *Optimality-theoretic syntax*. Cambridge, Mass.: MIT Press. 205–240.

Hale, Kenneth & Abanel Lacayo Blanco (1989). *Diccionario elemental del Ulwa* (*Sumu meridional*). Cambridge, Mass.: Center for Cognitive Science, MIT.

Halle, Morris (1973). Stress rules in English: a new version. *LI* **4**. 451–464.

Halle, Morris (2001). Infixation versus onset metathesis in Tagalog, Chamorro, and Toba Batak. In Michael Kenstowicz (ed.) *Ken Hale : a life in language*. Cambridge, Mass.: MIT Press. 153–168.

Halle, Morris & Michael Kenstowicz (1991). The Free Element Condition and cyclic versus noncyclic stress. *LI* **22**. 457–501.

Halle, Morris & Jean-Roger Vergnaud (1978). Metrical structures in phonology. Ms, MIT.

Halle, Morris & Jean-Roger Vergnaud (1987). *An essay on stress*. Cambridge, Mass.: MIT Press.

Hamano, Shoko (1986). *The sound-symbolic system of Japanese*. PhD dissertation, University of Florida, Gainesville.

Hammond, Michael (1986). The obligatory-branching parameter in metrical theory. *NLLT* **4**. 185–228.

Hammond, Michael (1989). Lexical stresses in Macedonian and Polish. *Phonology* **6**. 19–38.

Harms, Robert T. (1981). A backwards metrical approach to Cairo Arabic stress. *Linguistic Analysis* **7**. 429–450.

Harrikari, Heli (1999). The gradient OCP – tonal evidence from Swedish. Ms, University of Helsinki. Available as ROA-355 from the Rutgers Optimality Archive.

Harris, James W. (1983). *Syllable structure and stress in Spanish: a nonlinear analysis*. Cambridge, Mass.: MIT Press.

Hayes, Bruce (1980). *A metrical theory of stress rules*. PhD dissertation, MIT. Published 1985, New York: Garland.

Hayes, Bruce (1995). *Metrical stress theory : principles and case studies*. Chicago: University of Chicago Press.

Hetzron, Robert & Habte Mariam Marcos (1966). Des traits pertinents superposés en ennemor. *Journal of Ethiopian Studies* **4**. 17–30.

Hooper [Bybee], Joan (1976). *An introduction to natural generative phonology*. New York: Academic Press.

Horwood, Graham (to appear). *Relational faithfulness*. PhD dissertation, Rutgers University.

Hudson, Grover (1974). The representation of non-productive alteration. In John M. Anderson & Charles Jones (eds.) *Historical linguistics*. Vol. 2. Amsterdam: North Holland. 203–229.

Hume, Elizabeth (1998). Metathesis in phonological theory: the case of Leti. *Lingua* **104**. 147–186.

Hung, Henrietta (1994). *The rhythmic and prosodic organisation of edge constituents*. PhD dissertation, Brandeis University. Available as ROA-24 from the Rutgers Optimality Archive.

Inkelas, Sharon (1989). *Prosodic constituency in the lexicon*. PhD dissertation, Stanford University. Published 1990, New York: Garland.

Johnston, Raymond Leslie (1980). *Nakanai of New Britain: the grammar of an Oceanic language*. Canberra: Australian National University.

Kager, René (1993). Alternatives to the iambic-trochaic law. *NLLT* **11**. 381–432.

Kager, René (1996). On affix allomorphy and syllable counting. In Ursula Kleinhenz (ed.) *Interfaces in phonology*. Berlin: Akademie Verlag. 155–171.

Kager, René (1999). *Optimality Theory*. Cambridge: Cambridge University Press.

Kager, René (2000). Stem stress and peak correspondence in Dutch. In Dekkers *et al.* (2000). 121–150.

Kager, René (2001). Rhythmic directionality by positional licensing. Paper presented at the 5th HIL Phonology Conference, University of Potsdam. Handout available as ROA-514 from the Rutgers Optimality Archive.

Karttunen, Lauri (1998). The proper treatment of optimality in computational phonology. In *Proceedings of the International Workshop on Finite State Methods in Natural Language Processing*. Ankara: Bilkent University. 1–12.
Kaun, Abigail (1995). *The typology of rounding harmony: an Optimality Theoretic approach*. PhD dissertation, UCLA. Available as ROA-227 from the Rutgers Optimality Archive.
Kenstowicz, Michael (1980). Notes on Cairene Arabic syncope. *Studies in the Linguistic Sciences* **10**. 39–54.
Kenstowicz, Michael (1994). Syllabification in Chukchee: a constraints-based analysis. In Alice Davison, Nicole Maier, Glaucia Silva & Wan Su Yan (eds.) *Proceedings of the Formal Linguistics Society of Mid-America* **4**. Iowa City: Department of Linguistics, University of Iowa. 160–181.
Kenstowicz, Michael (1996). Base-Identity and Uniform Exponence: alternatives to cyclicity. In Jacques Durand & Bernard Laks (eds.) *Current trends in phonology: models and methods*. Salford: ESRI. 363–393.
Kenstowicz, Michael (1997). Uniform exponence: exemplification and extension. In Viola Miglio & Bruce Morén (eds.) *University of Maryland Working Papers in Linguistics* **5**. *Selected Phonology Papers from the Hopkins Optimality Theory Workshop 1997/University of Maryland Mayfest 1997*. 139–155.
Kirchner, Robert (1993). Turkish vowel harmony and disharmony: an Optimality Theoretic account. Ms, UCLA. Available as ROA-4 from the Rutgers Optimality Archive.
Kirchner, Robert (1996). Synchronic chain shifts in Optimality Theory. *LI* **27**. 341–350.
Klein, Thomas (2002). Infixation and segmental constraint effects: UM and IN in Tagalog, Chamorro, and Toba Batak. Ms, University of Manchester. Available as ROA-535 from the Rutgers Optimality Archive.
Legendre, Géraldine (1999). Morphological and prosodic alignment at work: the case of South Slavic clitics. *WCCFL* **17**. 436–450.
Legendre, Géraldine (2000). Morphological and prosodic alignment of Bulgarian clitics. In Dekkers *et al.* (2000). 151–189.
Legendre, Géraldine, Paul Smolensky & Colin Wilson (1998). When is less more? Faithfulness and minimal links in *wh*-chains. In Pilar Barbosa, Danny Fox, Paul Hagstrom, Martha McGinnis & David Pesetsky (eds.) *Is the best good enough? Optimality and competition in syntax*. Cambridge, Mass.: MIT Press. 249–289.
Leslau, Wolf (1950). *Ethiopic documents: Gurage*. New York: Viking Fund.
Leslau, Wolf (1967). The impersonal in Chaha. In *To honor Roman Jakobson*. Vol. 2. The Hague: Mouton. 1150–1162.
Leslau, Wolf (1979). *Etymological dictionary of Gurage (Ethiopic)*. Wiesbaden: Harrassowitz.
Li, Zhiqiang (2002). Intonational structure in Chaha: declaratives, interrogatives and focus. Ms, MIT.
Liberman, Mark & Alan Prince (1977). On stress and linguistic rhythm. *LI* **8**. 249–336.
Lieber, Rochelle (1987). *An integrated theory of autosegmental processes*. Albany: SUNY Press.
Lombardi, Linda (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *NLLT* **17**. 267–302.
Lombardi, Linda (2001a). Why Place and Voice are different: constraint-specific alternations in Optimality Theory. In Lombardi (2001b). 13–45.
Lombardi, Linda (ed.) (2001b). *Segmental phonology in Optimality Theory: constraints and representations*. Cambridge: Cambridge University Press.
McCarthy, John J. (1979). On stress and syllabification. *LI* **10**. 443–465.
McCarthy, John J. (1983). Consonantal morphology in the Chaha verb. *WCCFL* **2**. 176–188.

McCarthy, John J. (1994). On coronal 'transparency'. Handout of paper presented to TREND, Santa Cruz.

McCarthy, John J. (1998). Morpheme structure constraints and paradigm occultation. *CLS* **32:2**. 123–150.

McCarthy, John J. (2000a). Faithfulness and prosodic circumscription. In Dekkers *et al.* (2000). 151–189.

McCarthy, John J. (2000b). The prosody of phase in Rotuman. *NLLT* **18**. 147–197.

McCarthy, John J. (2002a). Comparative markedness. In Carpenter *et al.* (2002). 171–246.

McCarthy, John J. (2002b). *A thematic guide to Optimality Theory*. Cambridge: Cambridge University Press.

McCarthy, John J. (2003a). Comparative markedness. *Theoretical Linguistics* **29**. 1–51.

McCarthy, John J. (ed.) (2003b). *Optimality Theory in phonology: a reader*. Malden, Mass.: Blackwell.

McCarthy, John & Alan Prince (1986). *Prosodic morphology*. Ms, University of Massachusetts, Amherst & Brandeis University. (See also McCarthy & Prince 1996.)

McCarthy, John J. & Alan Prince (1990). Foot and word in prosodic morphology: the Arabic broken plural. *NLLT* **8**. 209–283.

McCarthy, John J. & Alan Prince (1993a). Generalized alignment. *Yearbook of Morphology 1993*. 79–153.

McCarthy, John J. & Alan Prince (1993b). *Prosodic Morphology : constraint interaction and satisfaction*. Ms, University of Massachusetts, Amherst & Rutgers University. Available as ROA-482 from the Rutgers Optimality Archive.

McCarthy, John J. & Alan Prince (1994). The emergence of the unmarked: optimality in prosodic morphology. *NELS* **24**. 333–379.

McCarthy, John & Alan Prince (1995). Faithfulness and reduplicative identity. In Beckman *et al.* (1995). 249–384.

McCarthy, John & Alan Prince (1996). *Prosodic morphology 1986*. (Revised version of McCarthy & Prince 1986.) Ms, University of Massachusetts, Amherst & Brandeis University. Available July 2003 at http://ling.rutgers.edu/gamma/pm86all.pdf.

McCarthy, John J. & Alan Prince (1999). Faithfulness and identity in Prosodic Morphology. In René Kager, Harry van der Hulst & Wim Zonneveld (eds.) *The prosody–morphology interface*. Cambridge: Cambridge University Press. 218–309.

MacEachern, Margaret (1999). *Laryngeal cooccurrence restrictions*. New York: Garland. Revision of 1997 UCLA PhD dissertation.

Marsack, C. C. (1962). *Teach yourself Samoan*. London: Hodder & Stoughton.

Merchant, Jason (1994). A note on the typology of Alignment: gradient vs. categorical violation. Ms, UCSC. Available July 2003 at http://home.uchicago.edu/~merchant/manuscripts.html.

Merchant, Jason (1995). MCat–PCat alignment and cyclic stress domains. Ms, UCSC.

Mester, Armin (1994). The quantitative trochee in Latin. *NLLT* **12**. 1–61.

Mester, Armin & Junko Itô (1989). Feature predictability and underspecification: palatal prosody in Japanese mimetics. *Lg* **65**. 258–293.

Mester, Armin & Jaye Padgett (1994). Directional syllabification in Generalized Alignment. In Jason Merchant, Jaye Padgett & Rachel Walker (eds.) *Phonology at Santa Cruz 3*. Santa Cruz: Linguistics Research Center. 79–85.

Myers, Scott (1987). *Tone and the structure of words in Shona*. PhD dissertation, University of Massachusetts, Amherst.

Myers, Scott (1997). OCP effects in Optimality Theory. *NLLT* **15**. 847–892.

Myers, Scott & Troi Carleton (1996). Tonal transfer in Chichewa. *Phonology* **13**. 39–72.

Nelson, Diane & Ida Toivonen (2001). Counting and the grammar: case and numerals in Inari Sami. In Diane Nelson & Paul Foulkes (eds.) *Leeds Working Papers in Linguistics*. 179–192.

Nespor, Marina & Irene Vogel (1989). On clashes and lapses. *Phonology* **6**. 69–116.
Ní Chiosáin, Máire & Jaye Padgett (2001). Markedness, segment realization, and locality in spreading. In Lombardi (2001b). 118–156.
Noske, Manuela (1999). Deriving cyclicity: syllabification and final devoicing in German. *The Linguistic Review* **16**. 226–252.
Noyer, Rolf (1993). Mobile affixes in Huave: optimality and morphological well-formedness. *WCCFL* **12**. 67–82.
Ohala, Manjari (1977). Stress in Hindi. In Larry Hyman (ed.) *Studies in stress and accent*. Los Angeles: Department of Linguistics, University of Southern California. 327–338.
Orgun, C. Orhan & Ronald Sprouse (1999). From MPARSE to CONTROL: deriving ungrammaticality. *Phonology* **16**. 191–220.
Padgett, Jaye (1995a). Feature classes. In Beckman *et al.* (1995). 385–420.
Padgett, Jaye (1995b). Partial class behavior and nasal place assimilation. In Keiichiro Suzuki & Dirk Elzinga (eds.) *Proceedings of the 1995 Southwestern Workshop on Optimality Theory (SWOT)*. Tucson: Department of Linguistics, University of Arizona. 145–183. Reprinted in McCarthy (2003b). 379–393.
Padgett, Jaye (2002). Feature classes in phonology. *Lg* **78**. 81–110.
Parker, Steve (1997). An OT account of laryngealization in Cuzco Quechua. *Work Papers of the Summer Institute of Linguistics, University of North Dakota* **41**. 1–11. Available July 2003 at http://www.und.nodak.edu/dept/linguistics/wp/1997parker.pdf.
Parker, Steve & David Weber (1996). Glottalized and aspirated stops in Cuzco Quechua. *IJAL* **62**. 70–85.
Pater, Joe (2000). Non-uniformity in English secondary stress: the role of ranked and lexically specific constraints. *Phonology* **17**. 237–274.
Payne, David L. (1981). *The phonology and morphology of Axininca Campa*. Arlington, Texas: Summer Institute of Linguistics.
Peperkamp, Sharon (1997). *Prosodic words*. PhD dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.
Piggott, Glyne L. (2000). Against featural alignment. *JL* **36**. 85–129.
Poser, William (1982). Phonological representations and action-at-a-distance. In Harry van der Hulst & Norval Smith (eds.) *The structure of phonological representations*. Part 2. Dordrecht: Foris. 121–158.
Poser, William (1989). The metrical foot in Diyari. *Phonology* **6**. 117–148.
Potts, Christopher & Geoffrey K. Pullum (2002). Model theory and the content of OT constraints. *Phonology* **19**. 361–393.
Prince, Alan (1976). 'Applying' stress. Ms, University of Massachusetts, Amherst.
Prince, Alan (1980). A metrical theory for Estonian quantity. *LI* **11**. 511–562.
Prince, Alan (1983). Relating to the grid. *LI* **14**. 19–100.
Prince, Alan (1985). Improving tree theory. *BLS* **11**. 471–490.
Prince, Alan (1990). Quantitative consequences of rhythmic organization. *CLS* **26:2**. 355–398.
Prince, Alan (1998). Two lectures on Optimality Theory. Handout of paper presented at Phonology Forum 1998, Kobe University. Available July 2003 at http://ling.rutgers.edu/gamma/kobe-all.pdf.
Prince, Alan & Paul Smolensky (1991). Connectionism and Harmony Theory in linguistics. Ms, University of Colorado, Boulder. Report CU-CS-533-91.
Prince, Alan & Paul Smolensky (1993). *Optimality Theory: constraint interaction in generative grammar*. Ms, Rutgers University & University of Colorado, Boulder. Available as ROA-537 from the Rutgers Optimality Archive.
Prokosch, Eduard (1939). *A comparative Germanic grammar*. Baltimore: Linguistic Society of America.

Pulleyblank, Douglas (1996). Neutral vowels in Optimality Theory: a comparison of Yoruba and Wolof. *Canadian Journal of Linguistics* **41**. 295–347.

Raffelsiefen, Renate (1995). Conditions for stability: the case of schwa in German. *Arbeitspapiere des Sonderforschungsbereichs 282, 'Theorie des Lexikons'* **69**. University of Düsseldorf.

Raffelsiefen, Renate (1999). Constraints on schwa apocope in Middle High German. In Aditi Lahiri (ed.) *Analogy, levelling, markedness*. Berlin & New York: Mouton de Gruyter. 125–170.

Rose, Sharon (1997). *Theoretical issues in comparative Ethio-Semitic phonology and morphology*. PhD dissertation, McGill University.

Samek-Lodovici, Vieri (1993). A unified analysis of crosslinguistic morphological gemination. Ms, Rutgers University. Available as ROA-149 from the Rutgers Optimality Archive.

Schachter, Paul & Fe T. Otanes (1972). *Tagalog reference grammar*. Berkeley: University of California Press.

Schane, Sanford (1972). Noncyclic English word stress. Ms, UCSD. [*Not seen*. Cited by Liberman & Prince (1997) from Halle (1973).]

Selkirk, Elisabeth (1980). The role of prosodic categories in English word stress. *LI* **11**. 563–605.

Selkirk, Elisabeth (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, Mass.: MIT Press.

Selkirk, Elisabeth (1995). The prosodic structure of function words. In Beckman *et al.* (1995). 439–469.

Selkirk, Elisabeth (1996). The prosodic structure of function words. In James L. Morgan & Katherine Demuth (eds.) *Signal to syntax: bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum. 187–214.

Smith, Jennifer (2002). *Phonological augmentation in prominent positions*. PhD dissertation, University of Massachusetts.

Smolensky, Paul (1993). Harmony, markedness, and phonological activity. Paper presented at Rutgers Optimality Workshop 1, Rutgers University.

Smolensky, Paul (1995). On the internal structure of the constraint component *Con* of UG. Ms, UCLA. Available as ROA-86 from the Rutgers Optimality Archive.

Spaelti, Philip (1994). Weak edges and final geminates in Swiss German. *NELS* **24**. 573–588.

Spaelti, Philip (1997). *Dimensions of variation in multi-pattern reduplication*. PhD dissertation, UCSC. Available as ROA-311 from the Rutgers Optimality Archive.

Spring, Cari (1990). *Implications of Axininca Campa for prosodic morphology and reduplication*. PhD dissertation, University of Arizona.

Stemberger, Joseph P. & Barbara H. Bernhardt (1999). Contiguity, metathesis, and infixation. *WCCFL* **17**. 610–624.

Urbanczyk, Suzanne (1996). *Patterns of reduplication in Lushootseed*. PhD dissertation, University of Massachusetts, Amherst.

Vennemann, Theo (1972). On the theory of syllabic phonology. *Linguistische Berichte* **18**. 1–18.

Walker, Rachel (1998). *Nasalization, neutral segments, and opacity effects*. PhD dissertation, UCSC. Available as ROA-405 from the Rutgers Optimality Archive.

Walker, Rachel (2002). Yuhup prosodic morphology and a case of augmentation. *NELS* **32**. 551–562.

Welden, Ann (1980). Stress in Cairo Arabic. *Studies in the Linguistic Sciences* **10**. 99–120.

Wilson, Colin (2000). *Targeted constraints : an approach to contextual neutralization in Optimality Theory*. PhD dissertation, Johns Hopkins University.

Wilson, Colin (2001). Consonant cluster neutralisation and targeted constraints. *Phonology* **18**. 147–197.

Zhang, Jie (2000). Non-contrastive features and categorical patterning in Chinese diminutive suffixation: MAX[F] or IDENT[F]? *Phonology* **17**. 427–478.

Zoll, Cheryl (1996). *Parsing below the segment in a constraint-based framework*. PhD dissertation, University of California, Berkeley. Available as ROA-143 from the Rutgers Optimality Archive.

Zoll, Cheryl (1997). Conflicting directionality. *Phonology* **14**. 263–286.