Direct OT:
Representation as Pure Markedness

Chris Golston
Heinrich Heine UniversitŠt

This paper argues for a model of morphology and phonology in which phonological form is represented purely in terms of markedness rather than by strings of segments. The argument runs as follows: (i) underlying representations are prosodified; (ii) this makes them relatively ill-formed; (iii) so representations can be represented purely in terms of ill-formedness. Step (i) brings to its logical conclusion the incorporation of prosody into generative phonology begun in the 1970s and 1980s, making prosody relevant at all phonological levels of description. Step (ii) allows representation to be tied directly to markedness. Step (iii) makes it impossible to represent unmarked structure. I formalize representation as pure markedness in terms of intentional violation of phonological well-formedness constraints: the distintive features of representation are given in terms of constraint violations. The result is a theory in which representation and evaluation are completely conflated.s

Representation as pure markedness (RPM) has three important results for grammatical theory. First, it requires abandoning a central (if bizarre) claim implicit in most current work in phonology, which I'll call the theorem of Impossibility. Impossibility states that every underlying form is an impossible surface form; it keeps grammatical theory from having a unified theory of representation. Impossibility follows jointly from the assumption that underlying forms are not syllabified and that surface forms are. I'll argue that Impossibility is as theoretically and empirically unsupportable as it is counter-intuitive. In its place Direct OT offers a completely unified theory of phonological representation throughout the grammar.

Second, RPM requires rather than stipulates the underspecification of phonological representations (Kiparsky 1982; Archangeli 1984; Steriade 1987). Unmarked structure is underspecified in Direct OT because it is not representable. The permanence of underspecification (Steriade 1994) follows as well: unmarked structure is unspecified at all levels of phonology because it is not representable at any level.

Third, RPM makes the linear order of segments fully redundant. Linear representations like /pajama/ or [ba.na.na] are purely expository devices in Direct OT. Actual representation is completely non-linear.

The particular formalization of RPM proposed here, intentional constraint violation, is not statable in theories of phonology in which constraints are inviolable (Bird 1990, 1995; Goldsmith 1990, 1991; Lakoff 1993: Mohanan 1993, Paradis 1988; Scobbie 1991, 1992). As formulated here, at least, RPM requires that constraints be violable (Prince & Smolensky 1993). Hence Direct *OT*.

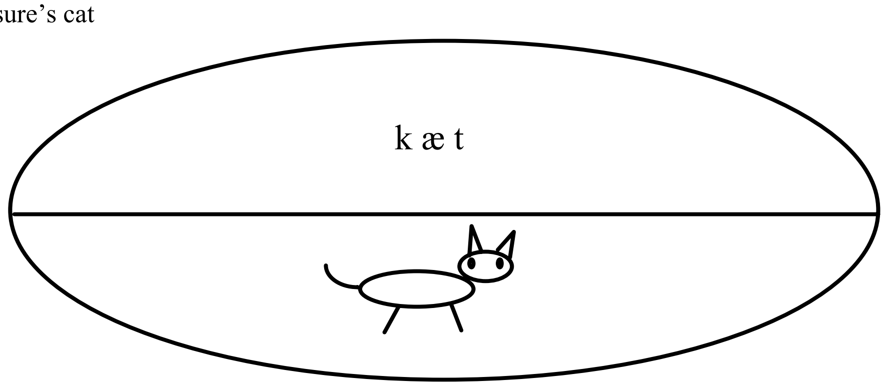
But RPM does not leave OT unchanged. The fourth major consequence of RPM is that it eliminates two of the five elements of standard OT ('UR' and 'SR') and diminishes the role of a third ('GEN'). Hence *Direct* OT. Where standard OT has unsyllabified segment strings (UR), syllabified segment strings (SR) and constraint violations (evaluation), Direct OT has only constraint violations. One of the major roles of GEN, producing a candidate set of syllabic parses for each unsyllabified input, is thus not needed. This reduction of theoretical machinery in OT is accomplished without adding anything to the model or reducing its ability to capture significant generalizations.

The paper is organized in the following way. Section 1 provides a brief overview of lexical representation in OT (1.1) and in Direct OT (1.2). Section 2 argues that underlying representation includes at least some prosodic information. Sections 3-6 elaborate the core of Direct OT: they show in detail how RPM works in underlying (3) and surface (4) representation; how it unifies the representation of all morphemes from roots and affixes to subtractive and templatic morphs (5); and how it informs speech production (6). Section 7 compares Direct OT to some recently proposed alternatives to lexical representation (Neef 1994, Hammond 1995, Russell 1995). Section 8 offers a short conclusion.

1. Lexical representation

Consider a traditional representation of the arbitrary pairing of form and meaning (1), according to which a phonological representation is paired with the semantic representation.

(1) Saussure's cat



A common understanding of an underlying phonological form is that it consists of an ordered string of segments (or sets of autosegments) with little or no information about prosody. A fully syllabified surface form, with onset [k] nucleus [¾] and coda [t], is the result of applying a set of phonological rules or well-formedness conditions to the unsyllabified string, deriving a syllabified string from an unsyllabified string

Such a view of things has some important implicit assumptions. Impossibility has already been mentioned in this regard: no underlying form is a possible surface form, because UR is stripped of prosody while SR is fully prosodified. This makes underlying representation necessarily unpronouncable, an unintuitive restriction at best. It also makes phonology necessarily derivational, since no underlying form can surface without being syllabified.

In the following section I'll look at how the theorem of Impossibility plays out in a particular theory of phonology, standard OT. But the bulk of the discussion holds for any theory in which underlying forms are not syllabified and surface forms are.

1. 1 Lexical representation in standard OT

Standard OT (Prince & Smolensky 1993) is a constraint-based theory of phonology that analyzes the phonologies of particular languages in terms of typologically well-founded well-formedness conditions. It thereby incorporates markedness theory directly into phonological derivations: only the least-marked structures make it to the surface. Standard OT is based on the three principles in (2):

(2) Principles of Optimality Theory

a. Violability
Constraints are violable; but violation is minimal

b. Ranking
Constraints are ranked on a language particular basis; the notion of minimal violation is defined in terms of this ranking.

c. Inclusiveness
The constraint hierarchy evaluates a set of candidate analyses that are admitted by very general considerations of structural well-formedness.

Nothing in (2) dictates the shapes of phonological representation, making it completely compatible with the possibility of RPM. As we will now see, however, this is not the way in which standard OT has developed. Rather, it incorporates prosodically deficient underlying representations, prosodically rich surface representations and thus fully embraces Impossibility. A brief overview of OT makes this clear.

OT makes use of five distinct components, briefly described in (3):

(3) Five components of standard Optimality Theory

UR	unprosodified string of segments
SR (candidate)	prosodified string of segments
Gen	function taking one UR to multiple SRs
Constraints	universal set of ranked, violable well-formedness conditions
Violations	indicators of ill-formedness in terms of constraints

The role of each of these components may be illustrated in a constraint-tableau (4), the central work-place of the theory, in which surface forms are derived from underlying representations. An underlying form, consisting of an unsyllabified string of segments, may be realized as a large number of possible output or surface forms (candidates), each of which contains the underlying form, but in a fully prosodified form, with stress, syllabification, stray and epenthetic segments and so on present. A number of ranked and violable constraints pick one of the candidates out as the best and this candidate is the surface form in the language.

(4) OT constraint tableaux

UR	CONSTRAINT	CONSTRAINT
SR		Violation
SR	Violation	

A simple example (from Prince 1993) illustrates how all of this works:

(5) An OT constraint tableau

/spatula/	NONFINAL	EDGEMOST
☞ (sp‡.tu) (la)		*
(spa) (tœ.la)	*!	

An input like /spatula/ consists of a linear string of segments and lacks any information that is fully predictable from that string, such as syllable structure. A number of possible output candidates is produced by a function GEN: these include in the present case the fully syllabified, footed and stressed candidates (sp‡.tu) (la) and (spa) (tœ.la). Ranked constraints (ranking shown by the solid vertical line; unranked constraints are separated by a dotted line) determine which candidate is the winner—in this case the candidates are evaluated by two constraints, NONFINAL and EDGEMOST. NONFINAL requires that the foot which contains the main stress not be the final foot of the word; the second candidate violates this constraint. EDGEMOST requires that the foot that contains the main stress occur at the (right) edge of the word; the first candidate violates this constraint because of the stress-foot (sp‡.tu). The constraints are ranked such that NONFINAL is more important than EDGEMOST, with the result that the first candidate wins, because it violates a less important constraint than its competitor.

Two additional elements in the tableau need to be mentioned here: the shading of the right-most column and the small hand pointing to the winning candidate. Both are merely expository in the theory. Shading merely shows us where not to look—the column headed by NONFINAL is enough to pick a winner and the column to its right plays no role in the selection of the grammatical form in this tableau. The small hand merely points out the winner and plays no role in the grammar.

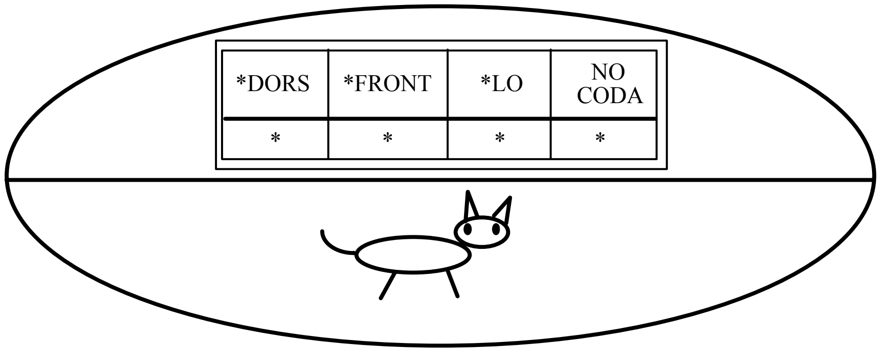
Impossibility has become firmly attached in standard OT, even though nothing in the theory requires it: the underlying forms used in OT analyses are invariably impossible surface forms. As we'll now see, this forces the function GEN to play a very large role in derivations. Many of the constraints used in OT evaluate candidates in terms of prosody: segments must belong to syllables (PARSE); syllables must have segments in them (FILL), must have onsets (ONSET) and must not have codas (NOCODA). It is GEN that creates representations which these constraints can evaluate. Unsyllabified strings like /k¾t/ have no onsets, nuclei, codas or syllables and thus are not evaluable in terms of prosodic constraints. The constraint grammar of OT is incapable of evaluating underlying forms and one of the main roles of GEN is to create forms which can be evaluated.

• But if syllable structure is so central to evaluation in OT, why is it left out of underlying representation? Generative phonology has kept syllable structure out of underlying reresentation for two closely related reasons. First, for any given string of segments syllable structure is predictable. Second, for any given string of segments syllable structure is not contrastive: we find no minimal pairs like [ta.tra] vs. [tar.ta]. What these (true) observations show is that a theory of grammar which posits both underlying syllable structure and underlying strings of segments is not free of redundancy. The usual solution abandons prosody in favor of linear order. RPM does the reverse. In the following section I sketch out how a minimal amount of prosodic information makes the linear order of segments fully predictable; the full picture is given in section 3.

1. 2 Lexical representation in Direct OT

I start with the possibility that the underlying form for CAT is not an unprosodified string, but a syllable, with onset, nucleus and coda. It is now apparent that the syllable [k¾t] is not perfect: the onset consonant [k] is relatively marked compared to, say, [t]; the low front vowel [¾] is relatively marked compared to [e] or [a]; the very presence of a coda consonant, especially an obstruent, is marked. All of this is the stuff of Optimality Theory proper, and it should now be clear that the constraint violations that are used to *evaluate* candidates can also be used to *represent* them:

(6) Saussure's cat



The constraint violations in (6) describe a string consisting of an unspecified consonant in coda position, a consonant [k] and a vowel [¾]. The most plausible candidate for such a string is [k¾t]. The syllable [k¾d] would have a violation of *VOICECODA; [kat] wouldn't have the *FRONT violation; [t¾k] would have a violation of *DORSCODA; and so on.

The driving principle behind this form of representation is markedness. Take the syllables [ti:] and [i:t]. They are distinguished from one another by the fact that one lacks an onset and has a coda while the other lacks a coda and has an onset. Given a language in which [t] and [i] are the unmarked consonant and vowel, we may distinguish them as follows:

(7) [i:t]

ONS	COMPLX NUC	NO CODA
*	*	*

(8) [ti:]

ONS	COMPLX NUC	NO CODA
	*	

(Note that the ordering/ranking of these constraints is completely irrelevant here.) The presence of an onset and absence of a coda in (8) is not represented because it represents the unmarked case.

This relative ill-formedness of underlying forms, such that [i:t] is less well-formed than [ti:], cannot be expressed if one characterizes underlying representations as unsyllabified strings of segments: the *segment strings* [ti:] and [i:t] are equally marked phonologically—it is only the *syllables* [ti:] and [i:t] that differ in this respect. Thus in a language in which (7) and (8) contrast, standard OT encodes both the marked case (coda and no onset in /i:t/) and the unmarked case (onset and no coda in /ti:/) in underlying representation. The nature of representation that standard OT analyses adopt keeps them from expressing obvious markedness differences among forms. These differences are captured straightforwardly in Direct OT because all and only marked structure is expressible as constraint violations: the relative ill-formedness of [i:t] and [ti:] is evident from the number of asterisks used to represent them.

The general point is that there is a general equivalence between a prosodified representation on the one hand and a set of constraint violations on the other. A simple Gedankenexperiment makes this clear: imagine two minimally distinct candidates, A and B, only one of which is grammatical in the language at hand. If standard OT can rank A and B with respect to each other (and it must if only one of them is grammatical), there must be some non-empty set of constraints C that A and B violate to different degrees. Then violations of C are capable of distinguishing any two underlying representations α and β, where α has the same shape as A and β has that of B. That is, standard OT assumes already that any phonologically relevant distinction can be expressed in terms of violations of one or more constraints—Direct OT uses this necessary assumption to cut some of the fat ('UR' and 'SR' segment strings) out of the theory by representing morphemes *directly* as constraint violations. Constraint violations are the distinctive features of Direct OT; they alone are what represent phonologically and morphologically relevant contrasts.

The upshot of RPM is that representation and evaluation are cut from the same cloth, the cloth of evaluation. This makes the UR and SR segment strings of standard OT mere expository devices, akin to shaded columns and the pointing finger. Given constraints and distinctive patterns of violations, there is no need for unsyllabified strings of segments in UR: constraint violations suffice as lexical representations. Thus the central claim of Direct OT: Meaning is paired directly with constraint violations. This means that no derivation is required to get from an underlying representation to its evaluation in terms of constraint violations since the two are indistinct; a significant part of the work done by the function GEN in standard OT is thus not required. It also means that URs need not be impossible SRs: Impossibility does not arise under RPM.

Before fleshing out the proposal in its entirety (section 3-6), I'd like to show that underlying prosody is already part of current phonology and morphology; thus the claim that some prosodic structure is underlying does not significantly enrich any part of OT.

2. UR is prosodified

In this section I'll consider arguments for the underlyingness of three levels of prosodic organization: moraic structure, syllable structure and foot structure. Linguistic data (2.1), psycholinguistic data (2.2), and theory-internal considerations

(2.3) all call for the same conclusion: some amount of prosody is required in the underlying structure of words. The question is: How much and of what type (2.4).

2.1 Linguistic evidence for underlying prosody

Moras

A number of phonologists have argued for representing consonant and vowel length in terms of underlying prosody (Van der Hulst 1984; Hyman 1985; McCarthy & Prince 1986, 1988; Hayes 1989).[1] Long vowels and geminate consonants may be represented as in (9), with a rule that adds a single mora to every vowel, or as in (10) with no such rule and some redundancy:

(9) Minimal moraic specification

μ		μ	
a	a	k	k
(long V)	(short V)	(long C)	(short C)

(10) Maximal moraic specification

μμ ∨	μ	μ	
a	a	k	k
(long V)	(short V)	(long C)	(short C)

Either way, *some* prosodic information is treated as underlying. Any theory of phonology that embraces (9) or (10), as standard OT and most of contemporary phonological theory do, already makes use of prosody in underlying representation.

One could of course argue that length is contrastive while syllable affiliation is not and thus have a principled reason for including moras in UR but excluding other prosody. But minimal pairs like *eat* and *tea* can be taken as evidence that syllable affiliation is indeed contrastive: the former has an onset [t] while the latter has a coda [t], otherwise the words are identical. This fact is of course blurred if underlying representation is given as an unsyllabified segment string, in which case *tea* and *eat* seem to differ in terms of the order of segments, not in terms of their prosodic affiliation. We most likely do not need both linear order *and* syllable affiliation of segments; but it does not follow from this alone which of the two is required.

[1] See Noske (1992) for critical assessment of the moraic account of compensatory lengthening; I am concerned here with the representation of contrastive length.

More direct evidence for the underlying status of moras comes from minimal requirements on lexical formatives. Extending work on word-minimality (McCarthy & Prince 1986) to the level of roots and affixes, Golston (1991) has argued that English, Latin and Classical Greek require derivational affixes to be minimally monomoraic. As with contrastive length in moraic theory, this is only statable in a grammar which allows moraic structure in underlying representations.

Syllables

There is ample evidence that syllable structure is underlying in many languages. Two general classes of phenomena can be distinguished here: minimality requirements on roots and prosodically defined templatic morphology, both ultimately due to the pioneering work of McCarthy (1979, 1981), Marantz (1982) and McCarthy & Prince (1986, 1990).

Let us take minimality requirements first. Golston (1995) notes that the well-established shape of Proto-Indo-European roots is exactly the type of evidence one needs for claiming that syllable structure is underlying in a language. PIE had the following types of roots (where R denotes the class of sonorants):

(11) Proto-Indo-European root shapes

(C)VC
(C)VR
(C)VRC

The generalization behind (11) is that PIE roots are all single closed syllables: this and nothing else captures the fact that CV, CVV, CVCR and CVCV are impossible PIE roots. A grammar of PIE which does not countenance underlying syllable structure or constraints that regulate it in underlying representation cannot capture this.

Sanskrit provides a similar case. Sanskrit inherited roots from PIE but lost its laryngeal consonants, with compensatory lengthening in coda position. The result, as Steriade (1988) has shown, is that Sanskrit roots are both monosyllabic and bimoraic. Again, this is simply not statable without underlying syllables: if roots are stored simply as segment strings, the fact that the strings all happen to constitute exactly one heavy syllable is completely accidental.

A similar case can be made for reduplicative morphology. Work by Marantz (1982), McCarthy & Prince (1986), Steriade (1988) and others has shown that reduplication is inherently prosodic. Reduplicative morphemes which consist solely of a syllable template straightforwardly depend on syllable structure being part of lexical representation. Any theory of morphology which countenances prosodic templates has underlying prosodic structure.

Feet

Entirely parallel facts obtain for feet. A number of languages require lexical words to be minimally a foot (McCarthy & Prince 1986; Crowhurst 1991). Golston's (1991) observation that this applies to roots as well shows that foot structure must be available underlyingly. Kager (1995) presents similar data from a number of Australian languages, showing that the shapes of roots are often based on the feet used elsewhere in the phonology. Again, such restrictions on underlying forms do not seem to be statable without underlying prosody.[2]

Languages that express morphological category by shape, such as Classical Arabic (McCarthy 1979; McCarthy & Prince 1990), Yawelmani (Archangeli 1991) and Choctaw (Ulrich 1986, 1994; Lombardi & McCarthy 1991; Hammond 1993) also show the necessity of underlying prosody. The point has been made repeatedly in the literature that morphemes in such languages may consist solely of a certain prosodic shape: this *is* underlying foot structure.

2.2 Pyscho-linguistic evidence for underlying prosody

Psycho-linguistic evidence offers an important check on linguistic theory by casting light on the psychological reality of different grammatical models. The best models of grammar are compatible with both linguistic and with psycho-linguistic data. As we'll see here, two types of evidence bear on the mental representations speakers store and it is significant that they agree with one another point for point (Cutler 1986, 173; Levelt 1989, 355).

Tip of the Tongue (TOT) states

Brown & McNeill (1966) showed that speakers who cannot think of a word tend to know three things about it: the initial segment or onset, the number of syllables and the stress pattern. When a speaker tries to access the phonological form for *sextant*, for instance, words like *secant* and *sextet* come to mind. The results have been confirmed by much subsequent research (see Levelt 1989, 320). This strongly suggests that speakers store words as syllables and do not store them merely as strings of segments.

Consider the alternative. If a speaker stores a word as a string of segments and cannot access (part of) that string, she should not be able to compute the number of syllables or location of stress. For to do so would require access to the string of segments she (ex hypothesi) has no access to. If, on the other hand, prosodic structure is stored (perhaps in conjunction with a string of segments), we expect it to be accessible even if the full form of the word is not.

---

[2] Steriade (1988) and Kager are non-committal about how to guarantee prosodic shapes of roots.

Malapropisms

Classifications of speech errors include a category of sound-related substitutions (Fromkin 1973) or malapropisms (Fay & Cutler 1977), involving mis-selection of a word that is phonologically but not semantically similar to the intended word. Typical cases include ('F' from Fromkin; 'FC' from Fay & Cutler):

(12) Sound-related substitutions

Intended		Spoken	
white Anglo-Saxon Protestant	>	white Anglo-Saxon prostitute	F
a routine proposal	>	a routine promotion	F
the conquest of Peru	>	the conquest of Purdue	F
prohibition against incest	>	prohibition against insects	F
week	>	work	FC
open	>	over	FC
constructed	>	corrected	FC

As these cases illustrate, the overall prosody of the target is matched by the overall prosody of the error:

(13)

Intended	Spoken	stress pattern	syllable count
Protestant	prostitute	(x..)	3
proposal	promotion	(.x.)	3
Peru	Perdue	(.x)	2
incest	insects	(xx)	2
week	work	(x)	1
open	over	(x.)	2
constructed	corrected	(.x.)	3

What we do not generally find in sound-related substitutions are cases like *protester* (xx.) for *Protestant*; *propinquity* (.x.x) for *proposal*; *perdition* (.x.) for *Peru*; or *insecticide* (.x.x) for *insects*—forms we would expect if words were stored merely as segment strings.

The criteria for phonological similarity here are identical to those found in TOT states: same onset, same stress pattern, same number of syllables. Data like this has led researchers like Crompton (1982), Fromkin (1985) and Butterworth (1989) to posit a phonological sub-lexicon within the mental lexicon. *White Anglo-Saxon prostitute* is produced when *prostitute* is mis-selected because of its proximity to *Protestant* in the phonological sub-lexicon. None of this makes any sense if syllable count and stress pattern are not somehow stored.

Again, it is most significant that two quite different sources of evidence converge on the same criteria: word-onsets, stress pattern and number of syllables are used in accessing the phonological forms of words. Any psychologically plausible model of grammar must come to terms with this and admitting prosody into underlying representation seems like the necessary first step.

2.3 Lexicon Optimization

Inkelas (1994) has recently argued that the internal logic of OT requires non-alternating structure to be underlying. Her point of departure is Prince & Smolensky's notion of Lexicon Optimization (1993, 192), which she paraphrases as follows: "of all possible underlying representations that could generate the attested phonetic form of a given morpheme, that particular underlying representation is chosen whose mapping to phonetic form incurs the fewest violations of highly ranked grammatical constraints" (p. 6).

The argument is essentially that non-alternating structure is best put into UR even if it is predictable, because this allows the shortest (most harmonic) way from UR to SR. Consider a word like *do*: the syllable structure is invariably CV, even if suffixes (*doable*) or prefixes (*undo*) are added, so why not store the syllable structure as part of the underlying form? If adding structure violates the constraint FILL, the best lexical entry will have all of the non-alternating structure of the output form. This includes most of the syllable structure of course, since only the syllable affiliation at the edges of lexical items ever alternates. This is similar to Vennemann's (1973) claim that lexical representations of non-alternating parts of morphemes are identical to their phonetic representations.

The general point is that nothing is gained, and in fact a great deal is lost, by assuming that no syllable affiliation is underlying: given a string of segments /k¾t/ as the underlying representation for CAT, the optimal candidate must violate a number of FILL constraints in order to create onset, nucleus and coda positions. Given a fully syllabified underlying representation, the optimal candidate violates no FILL violations. Put another way, if the best OT grammar has the simplest derivations, then all non-alternating prosody is underlying.

2.4 The nature of underlying prosody in Direct OT

RPM allows only a small part of prosody to be underlying, the marked part. As the representations in (6) - (8) make clear, RPM provides no means of representing (i) a simple onset, (ii) a simple nucleus, or (iii) the lack of a coda. It is both unnecessary and impossible to encode such structure in Direct OT. RPM is only able to represent (i) the lack or complexity or high sonority of an onset, (ii) the lack or complexity or low sonority of a nucleus and (iii) the presence or complexity or low sonority of a coda. Unmarked prosodic structure can't be represented as constraint violation because it violates no constraints.

This is in sharp contrast to the situation found in standard OT, where any limitations on the type and amount of prosodic structure must be stipulated. Indeed, we have just seen that the internal logic of standard OT seems to *require* underlying specification of most prosodic structure. Closer consideration only makes things worse: lexicon optimization requires underlying representaion of *unmarked* prosody. Onset and nucleus nodes are put into UR to minimize violations of FILL, but no such impending violations make it better to put coda nodes into UR since such nodes are banned. Thus lexicon optimization requires either that all prosody be underlying or that only unmarked prosody be underlying.

Thus, a major difference between standard OT and Direct OT is that the former requires underlying prosody that the latter rules out in principle.

3. Representation as violation

We've seen that most work in generative phonology tacitly assumes Impossibility, repeated here as (14):

(14) Theorem of Impossibility: Every underlying form is an impossible surface form

This follows from two claims: (i) that underlying forms are not syllabified and (ii) that surface forms are always syllabified. We've seen that there are both empirical and theoretical grounds for abandoning (i); and it is clear that (14) does not follow from (ii) alone, or from any known principle of grammar, or even from common sense. We may therefore abandon it and the problems that come with it: different kinds of representation for UR and SR; URs that cannot be evaluated in terms of markedness; the absolute necessity of derivations.

This is what paves the way for representing lexical items in terms of pure markedness. We will now see in detail how this is to be done. For monosyllables violations of syllable well-formedness constraints suffice to make linear order redundant (3.1). For polysyllabic morphemes linear order is predictable once we recognize violations of alignment constraints (3.2).

3.1 Monosyllables and syllable well-formedness

Above I tried to show that [ti:] and [i:t] can be represented as in (7) and (8), repeated below:

(15) [i:t]

ONS	CMPLX NUC	NO CODA
*	*	*

[ti:]

ONS	CMPLX NUC	NO CODA
	*	

The unviolated constraints for [ti:] are here for comparison only; they need not be part of the underlying form since they contain no unpredictable information. These patterns of constraint violations are the distinctive features that distinguish [ti:] and [i:t] from one another and from all other syllables in a language, given a reasonable theory of underspecification (eg, Steriade 1994). In principle, at least, they suffice as underlying forms and render segment strings superfluous as a means of representation. Thus the surface linear order of [i:] and [t] is a function of the constraint-violations that define them and need not be otherwise encoded.

The immediate advantage of RPM to Optimality Theory is that it eliminates the unsyllabified segment string ('UR'). In its place, Direct OT posits what is needed in standard OT anyway: evaluation in terms of constraint violations. The savings come from the fact that one does not need both a segment string and constraint violations, the first as underlying representation and the second as a means of carrying out the evaluation metric: constraint violations do both.

The amount and type of underlying prosodic structure allowed in Direct OT follows directly from RPM: only marked prosodic structure is allowed underlyingly because unmarked structure cannot be represented. It is then clear how morphemes like *Pat, tap* and *apt* (16-18) can be distinguished. (16) and (17) differ only in that the latter has a place feature in the coda:

(16) [p¾t]

*STOP	*LAB	*FRONT	*LO	LAB CODA	COMPLXCODA	ONS
**	*	*	*			

(17) [t¾p]

*STOP	*LAB	*FRONT	*LO	LAB CODA	COMPLXCODA	ONS
**	*	*	*	*		

(18) [¾pt]

*STOP	*LAB	*FRONT	*LO	LAB CODA	COMPLXC ODA	ONS
**	*	*	*	*	*	*

The difference between (17) and (18) reduces to violation of COMPLEXCODA and ONSET in the latter. (16) and (18) differ the most, with (18) violating three constraints that (16) respects. (Recall that the unviolated constraints are included here only for comparison; non-violation is not distinctive.)

Looking now at the representation of a more complex root like *tramp* (19), we see how efficient representation-as-violation can be. Nothing needs to be said about the linear order of *any* of the features because their order within onset, nucleus and coda is given by sonority sequencing.

(19) [tr¾mp]

*STOP	*LIQ	*FRONT	*LO	*NAS CODA	*LAB CODA	*STOP CODA
**	*	*	*	*	*	*

(19) contains all of the information of the unsyllabified string /tramp/ and renders it obsolete; the linear order of the segments in *tramp* is redundant given their syllable affiliation.

3.2 Polysyllables and alignment

Polysyllabic forms can also be represented without directly specifying the linear order of segments. This may be done with violations of alignment constraints (McCarthy & Prince 1993c). Take a minimal pair like *able* [e.bl] and *bail* [bel]. The crucial difference in terms of alignment is that the [b] in *bail* occurs at the left edge of the word while in *able* it is separated from that edge by the syllable peak [e]. The alignment of [e] and [l] is identical for both words: no syllable peak separates [e] from the left edge and one syllable peak separates [l] is separated from that edge by one syllable peak.

The plausibility of using featural alignment constraints for deriving linear order rests of course with the independent need for such constraints in grammar. Three sources of evidence secure this: tonal alignment, vocalic alignment and consonantal alignment.

Autosegmental studies of tone (Leben 1973, Goldsmith 1979, Pulleyblank 1986) have shown that tonal melodies in many languages are mapped onto tone-bearing units (TBUs) directionally, usually from left to right. Given a tonal melody LH, for instance, a three-syllable word in Mende surfaces as LHH, with one-to-one association of tones to TBUs, followed by spread of H to the final TBU. Translating this into alignment terms, we see that the tones are all as far to the left as they can be without violating the ban on contour tones.

Evidence for alignment of vocalic features is found in Tiv (Pulleyblank 1988). In Tiv verbal roots, "no features are actually assigned to vowels underlyingly; surface forms result from the interaction of morpheme-level specifications with rules of spreading and redundancy" (p. 299). According to Pulleyblank's analysis, vowel features are mapped onto syllables from left to right, with the result that the underspecified vowel [e] generally occurs only to the right of other vowels; conversely, high vowels do not occur to the right of non-high vowels, round vowels do not occur to the right of non-round vowels, and low vowels do not occur to the right of non-low vowels. Again, this translates into alignment constraints that require vocalic features to align with the left edges of verbal roots. A few exceptional cases violate these constraints: eg, *dz□nda* 'drive away' is one of four verbal roots that violates ALIGN-L (LOW; ROOT). For full discussion and analysis the reader is referred to the original article.

Evidence for the alignment of consonantal features comes from McCarthy's (1979) work on Classical Arabic. Greenberg (1978) observed that Classical Arabic had no roots of the shape X-X-Y; McCarthy explained this in terms of left-to-right association of melody to skeletal slots, in conjunction with an OCP-driven ban on adjacent identical autosegments. An alignment constraint requiring consonantal features to be root-initial has the same effect in a non-derivational theory. Given features defining two consonants [s] and [m], *sasam* is worse than *samam* because in it the features NASAL and LABIAL are further removed from the left edge of the word.

Consider now the minimal pairs *slander* [sl¾n.dr] and *Sandler* [s¾nd.lr] which differ in the linear order of [l]. I assume that a feature incurs a violation of ALIGN-L (FEATURE, ROOT) for every sonority peak that separates it from the edge of the root:

(20) *slander* *Sandler*

```
     Pk   Pk         Pk      Pk
     |    |          |       |
  [s l æ n d r     [s æ n d l  r
```

So the features that specify [n, d, r] in *slander* and those for [n, d, l, r] in *Sandler* each incur one violation mark.

Lexical representations for *slander* and *Sandler* are then as follows:

(21) *slander*

*CONT	*LIQ	*LAT	*LO	*FRNT	*NAS CODA	ALN-L NAS	ALN-L VOI	*LIQ NUC	ALN-L LIQ
*	**	*	*	*	*	*	*	*	*

Sandler

*CONT	*LIQ	*LAT	*LO	*FRNT	*NAS CODA	ALN-L NAS	ALN-L VOI	*LIQ NUC	ALN-L LIQ
*	**	*	*	*	*	*	*	*	**

The tableaux are identical except for the last column, which shows that both liquids are mis-aligned from the left edge of the root in *Sandler.* The first tableau can be read as follows: *slander* is a two-syllable root (at most one violation per feature-alignment constraint) with the segments {s, l, r, ¾, n, d}; the {n} occurs in the coda; {n, d, r} are not part of the first onset or nucleus; {r} is syllabic. The second tableau reads: *Sandler* is a two-syllable root with the segments {s, l, r, ¾, n, d}; the {n} occurs in the coda; {n, d, l, r} are not part of the first onset or nucleus; {r} is syllabic. It is optimally realized as [s¾nd.lr] rather than as [s¾n.drl] because the latter further mis-aligns the feature Lateral (which now occurs two syllable peaks from the left edge of the root).

Note that violations of Align do not simply stipulate the linear order of segments in the way that segments strings do. A couple of examples may make this clearer. The difference between *poultry* and *portly* is expressed by *LATERALCODA, which *poultry* violates and *portly* respects. These words do not differ in terms of lateral alignment since [l] is in both cases misaligned from the left edge of the root by exactly one peak. Similarly for a pair like *blurr* [blr] and *burl* [brl], whose representations are identical except that the latter has a mis-aligned feature Lateral. Alignment, sonority sequencing and syllable well-formedness encode linear order in Direct OT without stipulating it.

There is nothing in an unsyllabified segmental string of segments that cannot be represented as a violation of constraints that are independently needed to evaluate candidates in OT. We are left then with two theories: standard OT, which uses constraint violations for evaluation and a string of segments for underlying representation and Direct OT, which uses constraint violations for both. Simplicity dictates the latter.

4. Surface representation as constraint violation

Two issues need to be addressed here. First, there is the adequacy of constraint violations as surface representations in non-alternating morphemes (4.1). Second comes the question of alternation in polymorphemic words (4.2).

4.1 Non-alternating morphemes

As we have seen, Impossibility insists that underlying forms not be pronounceable. This has especially counterintuitive results with non-alternating morphemes. Consider a non-alternating word like *but*: its surface representation contains an onset, a nucleus and a coda but underlyingly it must consist of an unspeakable string of segments. As should now be clear, Impossibility is not a fact about language but a choice on the part of linguists to insists on a certain type of representation at the expense of lengthy derivations. If we value shortness of derivation instead (Chomsky 1994), we arrive at a different type of grammar, one in which Impossibility does not arise.

There are of course underlying forms that cannot be spoken in isolation. Hooper (1975) cites the Spanish participial marker *-ndo* as a morpheme that cannot be assigned underlying syllable structure: [nd] is not a permissible onset in Spanish and [n] is not a permissible syllable. While it is true that *-ndo* is not a syllable or even two syllables, it is not true that it has no syllable structure. In fact, it has invariant syllable structure, with [n] always occurring in coda position and [do] always forming a following syllable. It can be represented as in (22) and contrasted with Spanish *Don* in (23).

(22) PARTICIPLE

*VOI	*NAS	*STOP	*RND	*NAS COD	ALN-L RD	ALN-L STOP
*	*	*	*	*	*	*

(23) 'Sir'

*VOI	*NAS	*STOP	*RND	*NAS COD	ALN-L NAS
*	*	*	*	*	*

The suffix *-ndo* has misaligned ROUND and STOP features (separated from the left edge of the morpheme by the sonority peak [n]) while *Don* has a misaligned NASAL feature.

As is apparent with non-alternating forms like *but* or *-ndo*, Direct OT reduces both underlying and surface phonological representations to evaluation in terms of constraints. The equivalence of UR and SR in such forms eliminates the need for deriving one from the other, eliminating much of the processual component of OT. Similarly, the equivalence of SR and evaluation eliminates the need for evaluation as a process. Evaluation is not something the grammar does to representations in Direct OT: rather, evaluation and representation are non-distinct.

We see then that RPM makes no principled distinction between an underlying form and a surface form. The representation of a non-alternating form is thus the same at all levels of representation, just as the term *non-alternating* suggests. This is in sharp contrast to standard OT, where non-alternating forms have one shape in UR and another in SR.

4.2 Alternation

In this section I'll look at two types of alternation, prosodic and featural. Resyllabification is a case of prosodic alternation; it involves alternation of the syllabic affiliation of some set of features. For instance, German *Bund* 'club' has as a plural [bun.d-«]. with the final stop resyllabifying into the onset of the syllable headed by the plural suffix -«. Devoicing is a case of featural alternation. We can see a case of this by comparing the voiced [d] of the plural *Bunde* with that of the singular *Bund* [bunt].

The prosodic fact we need to account for is that the root-final consonant in *Bund* surfaces sometimes as a coda (singular) and sometimes as an onset (plural); the featural alternation involves the different values of the feature VOI in onset and coda position. The underlying form for *Bund* is as follows:

(24) *Bund* 'club'

*VOI	*LAB	*HI	*RND	*NAS COD	*STOP COD	ALN-L NAS	ALN-L VOI	ALN-L STOP
**	*	*	*	*	*	*	*	*

(24) specifies a sonority peak that is high and round, preceded by a trough containing a voiced labial and followed by a trough consisting of [nt] and the feature VOI. Note that one of the *VOI violations is essentially floating in this representation: it is not part of the coda (there is no **VoiCod* violation and could not be given German phonotactics) and there is no second syllable-nucleus for it to be an onset to.

The unsuffixed nominative form [bunt] surfaces when the violations in (24) are fed into a full tableau. Evaluation proceeds as in standard OT (irrelevant constraints omitted):

(25) *Bund* 'club'

*VOI COD	FILL	*VOI	*STOP COD	
		**	*	☞ bunT.voi
	*!	**	(*)	bun.dE
*!		**	*	bund

(Note that the segmental spell-outs in the final column are *purely expository*, just like the pointy finger and the shading; note also that there is no evidence for constraint ranking yet.) The first candidate [bunT.VOI] is identical with the underlying form: it has two instances of voice, neither of which is in the coda. The second candidate [bun.dE] fails because it introduces an empty segment (E) in violation of FILL (Prince & Smolensky 1993). Note that it also fails to respect the underlying *STOPCODA violation (parenthesized in the tableau). The third candidate [bund] violates an undominated constraint against voicing features in coda position.

We see then that (24) serves both as the underlying form of *Bund* and as its (unsuffixed) surface form. Phonetic evidence for (24) as a surface form comes from the phonetics of final-devoicing in German. A number of studies (eg, Port & Crawford 1989, Brockhaus 1992) have shown that final devoicing is not fully neutralizing: native speakers can usually tell the difference between *bunt* 'variegated' and *Bund* 'club'. Traditional accounts of devoicing have difficulty in accounting for this because the VOI feature is not part of the surface phonological representation. Standard OT comes a step closer since it does have VOI in surface representation; but the fact that is unparsed (bunt <voi>) makes it difficult to see how it could survive into phonetic implementation. The representation in (24) offers a solution: the feature Voi is part of the surface representation as the final part of the sonority trough following [u]—but it is not affiliated with the coda. Phonetic implementation of this representation is equivalent to tonal downstep: the phonetic realization is somewhere between a voiced coda consonant and a voiceless coda consonant.

Let us turn now to the suffixed plural *Bunde* [bun.d«]. The form of the plural suffix is given in (26) and requires some comment. Schwa is generally not counted among the phonemes in German because it does not contrast with other vowels (Wiese 1995). Standard OT can deal with this as an underlying root node devoid of features, incurring a FILL violation in the winning candidate. Direct OT offers a simpler solution, bypassing the empty root node: this plural is marked directly as a segmental violation of FILL, the sole phonological reflex of the morpheme.

(26) PLURAL

FILLSEG
*

A word about the solid border around the tableau in (26). Polymorphemic words require some graphic way of representing which distinctive features (constraint violations) go with which morphemes. In segmental theories of representation this is usually done with a dash (-) or a plus (+) within the string: eg, dog-s. In Direct OT, where there are no strings, the morphemic affiliation of distinctive features must be shown non-linearly. In what follows I designate morphemeic affiliation by means of different borders: solid for affixes, plain for roots.

Concatenation of morphemes is represented merely by feeding the tableaux of different morphemes into a single tableau. The set of constraint violations that represents *Bund-e* is just the union of the set of violations for *Bund* and the set of violations for *-e*. It is exactly the same set of constraint violations used in standard OT to evaluate the two-syllable candidate form (bund)-(e). From here Direct OT is about like standard OT (irrelevant constraints omitted here for simplicity):

(27) ‘clubs’

FILL	*VOI	*STOP COD	ALN-L VOI	ALN-L STOP	
*	**	(*)	*	*	☞ bun.d«
(*)!	**	*	*	*	bunt

GEN gives us a number of candidates, two of which need to be considered here: those corresponding to the syllabified strings [bun.d«] and [bunt].[3]

The first candidate in (27) fails to respect the distinctive *STOPCODA violation of the root, a violation of faithfulness. The breach of faithfulness is indicated by parentheses around the asterisk in the appropriate cell (*). These parentheses do not indicate optionality but the fact that a distinctive feature of the root (a violation of *STOPCODA) is not realized; they graphically represent the less marked status of realizing all of the distinctive features of a morpheme. Not to realize a distinctive feature (an *) is more marked (‘(*)’) than realizing it.

The second candidate in (27) fails to manifest a distinctive feature of the input, the underlying FILL violation of the suffix, and receives a faithfulness violation as well. The ranking of FILL above *STOPCODA correctly selects the first candidate as the winner.[4]

5. Problematic Morphology

So far we have seen that it is possible to represent roots (*tea, eat*) and suffixes (*-ndo, -«*) in terms of underlying constraint violations. In this section I show that representation as constraint violation generalizes naturally to prefixes (5.1), infixes

[3] A couple of other candidates, [bund?] and [bunt<voi>?] are also possible given that glottal stop is the epenthetic consonant in German (Wiese 1995, ch. 3). Both entail gratuitous violations of *COMPLEXCODA that [bun.d«] avoids and will not be further considered.

[4] Failing to respect the *FILL violation of the plural morpheme results in a violation of *Exponence* as well, since [bunt] has no reflex for the plural.

(5.2) and circumfixes (5.3) as well as to zero morphology (5.4) subtractive morphology (5.5) reduplication (5.6) and templatic morphology (5.7). Thus the central claim of Direct OT, that meaning is paired directly with constraint violations, seems to generalize to all types of morphology. Put another way, there would seem to be no morpheme whose representation (UR, SR) and evaluation (tableau) cannot be conflated.

This section simultaneously serves another purpose. RPM is compatible with a number of a priori highly unlikely underlying forms because it allows any constraint violation to be distinctive. It is thus compatible with the following scenarios: a morpheme is marked by a distinctive violation of ALIGN; or of CONTIGUITY; or PARSE; or NOCODA. If constraint violations are the distinctive features of Direct OT, such morphemes are predicted to occur. The second task of this section is to show that these types of morphemes do occur.

RPM represents things directly in terms of observable surface patterns. Most roots and affixes, as we have seen, are represented purely in terms of featural and syllabic constraint violations. We will now see how this strategy generalizes to other types of morphology: prefixation intentionally violates ALIGNMENT; infixation and circumfixation violate CONTIGUITY; subtractive morphology violates PARSE; reduplication violates FILL; and so on.

5.1 Prefixes

Prefixation is obviously not on every morphologist's list of difficult morphological processes. But it does cause problems in standard OT, in that prefixation (and suffixation for that matter) requires invoking language-specific constraints, something not otherwise tolerated in the theory (Prince & Smolensky 1993).

Let me illustrate with McCarthy & Prince's (1993c) treatment of Tagalog *um-*. They claim that a prefix is placed before its host so as not to violate a constraint which demands that it be there. More precisely, it is claimed that the Alignment constraint in (28) is responsible for prefixation of *um-*:

(28) ALIGN-um[5]

Align-L ($[\text{um}]_{\text{Af}}$, Stem)

(Similarly for suffixes, with R instead of L). (28) is a really unlikely candidate for a linguistic universal. Noting, as McCarthy & Prince do, that part of the constraint (the Align part) is universal and that only the argument it takes

[5] I've simplified the form of the rule by factoring out the L: McCarthy & Prince's formulation reads Align ($[\text{um}]_{\text{Af}}$, L, Stem, L).

([POSS]$_{Af}$) is language specific does not make the constraint any less tied to Tagalog. Moreover, (28) misses an important generalization about Tagalog: prefixing is the norm, not the exception (Schachter & Otanes 1972; Lieber 1992).

Direct OT avoids both of these problems. First, the generalization problem. In a heavily prefixing language like Tagalog, we have a dominant constraint ALIGN-R (ROOT, WD); prefixes are unmarked, suffixes marked with underlying violations of this constraint.

(29) Tagalog *-um-* infixation: an alternative account

	NOCODA	ALIGN-R (RT, WD)	CONTIG
☞ u.m-a.ral	*		
a.r-u.m-al	*		*!
a.-um.-ral	**!		*
a.ral-um	*	*!	
um-grad.wet	***!		
g-um-rad.wet	***!		*
☞ gr-u.m-ad.wet	**		*

The selection of prefixal *um-aral* follows from constraints against discontinuous morphemes (**ar-um-al*), codas (**a-um-ral*) and misaligned roots (**aral-um*). Nothing like (28) is required because prefixes are unmarked in Tagalog. (Note that in a heavily suffixing language like English, ALIGN-L (ROOT, WD) is dominant; suffixes are the unmarked case and prefixes carry with them underlying violation of ALIGN.)

Second, the universality problem. The Direct OT analogue of (28) is required only for suffixes in a prefixing language and for prefixes in a suffixing language. A suffix in Tagalog is an affix that misaligns roots with the right eges of words; a prefix in English misaligns roots with the left edges of words. The constraint-violation that a Tagalog suffix or English prefix carries with it is exocentric, in that it is distinctive for the affix but takes roots as arguments: the marked case is the affix that misaligns roots with words. Such an approach does not require language specific alignment constraints like (28): the violated constraints, ALIGN-R (ROOT, WD) and ALIGN-L (ROOT, WD), contain no language specific information of any kind.

5.2 Infixes

A major result of Optimality Theory is its insightful handling of pseudo-infixes, prefixes or suffixes that are infixed under special circumstances. Pseudo-infixes are of two types. Tagalog *-um* infixation is driven by prosodic well-formedness

(Prince & Smolenksy 1993): a prefix is infixed into C-initial stems (*s-um-ulat* 'write') to avoid needless violation of NoCoda (**un-sulat*). We might call this prosodically induced infixation, due to constraints on syllable structure. Morphologically induced infixation is found in Dakota (Moravcsik 1977) in words like *la-ma-k'ota* 'I am a Lakota'. Here too the affixes in question also occur as prefixes. According to McCarthy & Prince (1993c), infixation in Dakota is driven by the desire of certain roots to left-align with a prosodic word; violation being minimal, the affix is infixed as close to the left edge as possible while allowing some material from the root to be word-initial. (The CV shape of the infix makes it clear that it is not infixed for prosodic reasons: **<u>ma</u>lak'ota* and *la<u>ma</u>k'ota* have identical syllable structure, showing that infixation is not always driven by prosody.)

But not all infixation is pseudo-infixation. This is most clear when infixation (i) does not alternate with prefixation or suffixation and (ii) isn't harmonic. English expletive-infixation (Siegel 1971, McCarthy 1982) is a good example: *fan-fuckin-tastic, Ala-fuckin-bama*, *Mononga-fuckin-hela.* As all of these cases make clear, expletive infixation invariably has detrimental affects. Most obviously, all cases of infixation have CONTIGUITY (McCarthy & Prince 1993b) violations that prefixed and suffixed words avoid. In addition, a word like *Ala-fuckin-bama* violates NOCODA and ONSET but is not for that reason realized as *fuckin-Alabama* which violates neither. Thus it is not the case that *fuckin* is a prefix or suffix that is infixed because of its shape.

But neither can infixation here be morphologically induced by the stem. Words like *hyper-fantastic*, (*fan-hyper-tastic), *pro-Alabama (*Ala-pro-bama)* and *anti-Monongahela* (**Mononga-anti-hela*) show that none of these roots has a high ranking Align (Rt, L, PrWd, L) requirement that forces infixation. The facts here support the native-speaker intuition that *fuckin* is inherently infixal—its status as such is not reducible to phonological details of its shape or to morphological requirements of its host.

Direct OT allows us to characterize true infixation as a violation of CONTIGUITY, its most salient surface cue. The relevant portion of the underlying form for *fuckin* is then:

(30) English expletive infix

CONTIG (ROOT)
*

As with the treatment of prefixes, the violation here is exocentric: it is a distinctive feature of the stem but takes a root as its argument. It straightforwardly captures what is distinctive about *Califuckinfornia* on the surface—the interruption of the root by the affix. Here as elsewhere in Direct OT, the violations that a morpheme brings with it are sufficient to represent it underlyingly.

To complete the analysis we need to deal with positioning. According to McCarthy (1982), *fuckin* precedes the first non-initial foot of the word. Interestingly, this is exactly the position in which we tend to get unfooted syllables in English. As McCarthy & Prince (1993c, 79) point out, the normal right-to-left alternation of stress in English is interrupted word-initially:

(31) (Tˆta) ma (g—uchee) *Ta (t‡ma) (g—uchee)

They derive this fact from the interaction of two constraints: one that requires all prosodic words to begin with feet, the other that requires all feet to occur at the right edge of the word. These two constraints, then, define a natural break in an English word before the first non-initial foot; infixation occurs here because of these constraints and the underlying violation of CONTIGUITY:

(32) Expletive infixation: *Tatamafuckingouchee*

CONTIG	ALGN-L (WD, FT)	ALIGN-R (FT, WD)	
(*)!	*	σσ	(fœckin)-Ta (t‡ma)(g—uchee)
	*!	σσ	Ta-(fœckin)-(t‡ma)(g—uchee)
		σσσσσσσσ!	(T‡ta)-(fœckin)-ma(gouchee)
		σσσσσσσ	☞ (T‡ta)ma-(fœckin)-(gouchee)
		σσσσσσσσσ!	(T‡ta)ma-(gou)(fœckin)-(chee)
(*)!		σσσσσσσ	(T‡ta)ma-(gouchee)-(fœckin)

As in standard OT, exact positioning of the infix within the stem is forced by other constraints; what Direct OT allows here is some way to force infixation when it is not a phonological property of the affix or a morphological property of the stem.

We are now in a position to sketch how the relevant parts of the grammars of Tagalog, English and Dakota differ. Tagalog ranks CONTIGUITY below NOCODA, with the result that infixes arise only under pressure from prosody. English ranks CONTIGUITY higher, with the result that it takes an underlying violation of it (part of the underlying form of *fuckin*) to yield an infix. Dakota is like English, but the morpheme that carries the violation of CONTIGUITY is the verb stem rather than the affix. The typology of infixation is thus filled out: it may be prosodically driven (Tagalog), in which case infixation is driven by the shape of affix and stem; affix-driven (English) in which case you must simply know whether an affix is infixed or not; or stem-driven (Dakota), in which case you must know whether a stem is infixing or not.

Note that there is no parallel in OT for an underlying violation of CONTIGUITY. Thus for Dakota, McCarthy & Prince (1993c, 113) cannot use the type of analysis I have sketched here. Rather they are forced to posit two grammars for Dakota. Infixing stems are subject to a grammar in which left-aligning roots is more important than left-aligning prefixes: ALIGN-L (ROOT, WD) >> ALIGN-L ([AGR]]AF, STEM). Prefixing stems are subject to a grammar with the opposite ranking. A parallel analysis of English expletive infixation would seem to require that affixes likewise be able to induce different grammars.

The Direct OT analysis sketched here claims that some morphemes carry with them a violation of CONTIGUITY and others do not. Analyses of Dakota and English are possible without any constraint re-ranking or proliferation of grammars: infixing stems (Dakota) and affixes (English) carry with them a violation of CONTIGUITY, prefixing stems (Dakota) and prefixes (English) do not.

5.3 Circumfixes

A number of languages have discontinuous affixes that both precede and follow the stem to which they attach (see Bauer 1988; Olson 1991). In terms of surface representation, circumfixation is to affixes what infixation is to roots: a blatant violation of contiguity. With infixes the violation is exocentric, realized as a discontinuous stem; with circumfixes the contiguity violation is endocentric, realized as a discontinuous affix.

Consider the following Tagalog data, from Schachter and Otanes 1972.

(33) ka...an 'the class or group of x'

ka-bukir-an	'fields'	bukid	'field'
ka-bunduk-an	'mountains'	bunduk	'mountainn'
ka-tagalug-an	'the Tagalogs'	Tagalog	'a Tagalog'

The circumfix may be represented as follows:

(34) [ka]...[an]

CONTIG (AFFIX)	*DORS	*LO	ALIGN-L (LO, AFF)	ALIGN-L (NAS, AFF)
*	*	**	*	*

The violation of CONTIGUITY is responsible for the discontinuity of the affix; the segmental content [ka.an] is given by the remaining constraint violations.

A tableau makes clear how the optimal form is circumfixal:

(35) kabukiran 'fields'

CONTIG (AFFIX)	CONTIG (ROOT)		
*		☞	ka-bukid-an
(*)!			ka.an-bukid
(*)!			bukid-ka.an
*	*!		ka-buk-an-id

The first candidate retains the distinctive discontinuity of the affix as well as the unmarked continuity of the root by circumfixing the affix around the root. The second and third candidates fail to violate Contiguity (Affix), a faithfulness (*) violation; the fourth retains the Contiguity (Affix) violation but violates Contiguity (Root) as well, again in violation of faithfulness. The ranking of constraints is irrelevant since only one of the candidates violates neither constraint. (Other possibilities such as [k-bukid-a.an] or [ka.a-bukid-n] lose out because of syllable well-formedness constraints.)

The analysis of circumfixes and infixes offered here makes direct reference to a surface property common to both: discontinuous morphemes. Circumfixes are treated as violations of CONTIGUITY (AFFIX), infixes as violations of CONTIGUITY (ROOT).

5.4 Zero Affixes

Morphological categories are not always overtly marked. Morpheme-based analyses often resort to zero-affixes to explain pairs like the following:[6]

(36) Zero-affixation

Singular	Plural
deer	deer-¯
fish	fish-¯
elk	elk-¯

The use of a zero affix is not totally ad hoc: its 'presence' may be used to block the addition of the elsewhere plural *-s*. The problem is that any use of a zero affix runs into an obvious dilemna: how do we kow that the ¯ is a suffix rather than a prefix (or infix or cirumfix)?

What is marked about the plurals in (36) is surely the lack of an affix. That is, zero-affixation violates EXPONENCE, the requirement that evey morpheme have a phonological reflex.[7] Direct OT represents such a morpheme as (37):

[6] Lieber (1992 and references therein) distinguishes zero-affixation from relisting. She analyzes English as a relisting language; the argument I wish to make here, that zero-affixation is simply violation of EXPONENCE, can also be made with her cases of zero-affixation.

(37) NOMINALIZER

EXPON
*

This avoids entirely the issue of where the morpheme is located (prefix, suffix, etc.) by directly recognizing the fact that the morpheme is not overtly represented. Note, however, that (37) blocks suffixation of default *-s*; suffixation of default *-s* would result in a faithfulness violation, since the EXPONENCE violation in (37) would not be realized.

5.5 Subtractive Morphology

Subtractive morphology, in which a grammatical category is marked by deletion of stem material, poses a major obstacle to morpheme-based theories of morphology, as Anderson (1992) has forcibly pointed out. The reason is that there seems to be no plausible way to characterize deletion as addition of an affix. Anderson takes this as evidence that at least some morphology is processual and thus 'a-morphous' as he puts it.

A paradigm case of subtractive morphology is found in Tohono O'odham (Zepeda 1983). (38) shows how the perfective form of verbs is formed by deleting the final syllable of the stem:

(38)	Stem	Perfective		
	cikapana	cikapa	'worked'	(surface [cik^{h}p^{h}])
	bisiceka	bisice	'sneezed'	(surface [bisc^{h}])

A number of analyses have been proposed (Hale 1965; Mathiot 1973; Martin 1988; Lombardi & McCarthy 1991; Hill & Zepeda 1992), all of them highly processual. The Lombardi & McCarthy analysis, for instance, treats subtraction as parsing out a light syllable at the right edge of the verb base and then deleting it. In that sense, each of these analyses is a-morphous and supports Anderson's claim that subtractive morphology is inherently processual. OT seems to be no better situated to handle the subtractive part of subtractive morphology. (Benua 1995 offers an OT analysis of the irregular phonology involved in subtractive morphology but leaves open the representation of subtraction morpheme.)

We saw above that the plural morpheme in German *Bunde* can be characterized simply as a violation of the faithfulness constraint FILL. In exactly parallel terms, subtractive morphology can be characterized as a violation of the other faithfulness constraint, PARSE. The underlying representation for the Tohono O'odham perfective is then merely (39):

---

[7] Alternatively, it violates a constraint that forces derived words to differ from their bases (Neef 1994).

(39) Tohono O'odham perfective[8]

PARSE SYLL
*

The fact that it is the last syllable of the word that is underparsed need not be stipulated as it is in the Lombardi & McCarthy analysis. Rather, it results from other constraints. First, an ALIGN-L (ROOT, WORD) constraint forces the left edge of the root to coincide with the left edge of a phonological word. Second, CONTIGUITY ensures against the underparsing of any medial syllable in the root.

This analysis is in fundamental agreement with Anderson's claim that subtractive morphology cannot be characterized as affixal. But it does not support his claim that (any) morphology is inherently processual. Rather, the analysis offered here supports the claim in (4) above: *Meaning is paired directly with constraint violations*. Morphology is not amorphous but ill-formed.

5.6 Reduplication

Reduplicative morphology has been elegantly analyzed as affixation of a prosodic template to a base (Marantz 1982; McCarthy & Prince 1986, 1988; Steriade 1988). The shape of the reduplicant in OT is regulated by constraints on templatic shape (McCarthy & Prince 1993b) or on alignment (McCathy & Prince 1993a).

Direct OT represents the introduction of empty prosodic structure directly as a violation of FILL. The syllable-sized reduplicant in Nootka (Stonham 1990), may be represented as in (40).

(40) Nootka reduplicant

FILL SYLL
*

The reprsentation of reduplication is thus entirely parallel with the representation of subtraction, as comparison of (40) and (39) shows: reduplication violates FILL, subtraction violates PARSE.

---

[8] Lombardi & McCarthy and Hill & Zepeda claim that only light syllables undergo truncation, but evidence for this (Hill & Zepeda ps. 384-385) seems to be rather weak. Hill & Zepeda base the claim on eleven words (their 36a-k) which, they claim "all exhibit bimoraic syllables at the right edge" (385). But the last three words (*e–ga* 'owning', *dada* 'arriving' and *cicwi* 'playing') would seem to end in short vowels and four of the remaining words are monosyllabic: *ka:* 'hearing', *ki:* 'living', *mu:* 'wounding by shooting' and bia 'dishing out food' and thus should not be subject to the rule anyway. I leave this all open for now and opt for the more general solution.

The general point is that violations of faithfulness (PARSE, FILL) can be distinctive. The fact that reduplication involves only authentic units of prosody (McCarthy & Prince 1986) falls out from the nature of representation in Direct OT: there are no constraints of the form FILLCC or FILLCVCC, so templates of the form CC or CVCC cannot be represented in terms of pure markedness.

5.7 Templatic morphology

A major challenge for Direct OT is how to model templatic morphology that is not reduplicative, such as that found in Classical Arabic (McCarthy 1979, McCarthy & Prince 1990), Yawelmani (Archangeli 1991) or Choctaw (Lombardi & McCarthy 1991; Hammond 1993; Ulrich 1994). There are two issues here. First, it is not always possible to translate the inherently positive notion of prosodic templates into constraint violations. Second, for Semitic languages at least, it is not possible to encode the linear order of putative roots like *k-t-b* 'write' or *d-r-b* 'beat,' for such roots are inherently not prosodified and thus do not violate the constraints necessary for encoding them. For such roots representation as pure markedness seems to be impossible.

Since the subject is fairly broad, I'll limit myself to the prosody of three areas of Classical Arabic: the binyanim, the masdars and the broken plurals. Again, the problems are (i) how to do without prosodic templates and (ii) how to encode unsyllabifiable roots. As a neutral way of indicating templatic shape I use the CV-notation developed in McCarthy 1979 (and later abandoned). As will be clear, this is merely expository; the lesson of Prosodic Morphology, that templates are *not* based on Cs and Vs, is not meant to be forgotten here but embraced.

Binyanim

The fifteen binyamin are based on the six prosodic shapes in (41).[9]

[9] I abstract away here from additional prefixal material: binyan 5 has the CVC.CVC shape of 2 plus a CV prefix; 6 has the CVV.CVC shape of 3 plus a CV prefix.

(41) Basic prosodic shapes of Classical Arabic verbs

Binyanim	Basic Shape
1	CV.CVC
2, 4, 5	CVC.CVC
3, 6	CVV.CVC
7, 8, 9	CCV.CVC
10, 12, 13, 14, 15	CCVC.CVC
11	CCVV.CVC

I take binyan 1 as basic and 'derive' the others from it, as it were, by means of intentional constraint violations. There are two reasons for choosing 1 as basic. Semantically, it tends to carry the least information (eg, 'write' as compared to 'cause to write', 'subscribe', etc); prosodically, it has the best shape, alternating single vowels with single consonants, making this the least marked choice for underlying representation. Thus the representation for the root 'write' is as in (42), corresponding to a linear reprsentation something like [kE.tEb]:

(42) 'write'

*DOR	*STOP	*LAB	ALN-L STOP	ALN-L LAB
*	***	*	***	**

This solves the problem of how to represent the triliteral root: it is treated exactly like a verbal root in Tiv, fully prosodified but lacking underlying vowel features.

The remaining prosodic shapes in (42) differ from the first in the structure of the initial syllable. The second and fifth shapes have a coda consonant. The third and sixth shapes have a long vowel. The fourth, fifth and sixth shapes have a complex onset. Given constraints on the three syllable positions, *COMPLEXONS, *COMPLEXNUC and NOCODA, we can characterize the differences between types as follows:

(43) How the types differ

Binyanim	Basic Shape	*CMPLX ONS	*CMPLX NUC	NO CODA
1	CV.CVC			
2, 4, 5	CVC.CVC			*
3, 6	CVV.CVC		*	
7, 8, 9	CCV.CVC	*		
10, 12, 13, 14, 15	CCVC.CVC	*		*
11	CCVV.CVC	*	*	

We can thus represent each of these oblique binyanim in terms of the violations they bring with them. Binyanim 2, 4 and 5 are marked underlyingly with a violation of NOCODA:

(44) Binyanim 2, 4, 5

NO CODA
*

Binyanim 3 and 6 bring with them a violation of *COMPLEXNUCLEUS, their defining surface property:

(45) Binyanim 3, 6

*CMPLXNUC
*

Binyanim 7, 8 and 9 are marked prosodically only by an underlying violation of *COMPLEXONSET:

(46) Binyanim 7, 8, 9

*CMPLX ONS
*

The remaining binyanim are doubly marked:

(47) Binyanim 10, 12, 13, 14, 15

NO CODA	*CMPLX ONS
*	*

(48) Binyan 11

*CMPLX ONS	*CMPLX NUC
*	*

The markedness of the types in (47) and (48) is partially reflected in the numbers of members in the binyanim: binyanim 12-15 are extremely rare (McCarthy 1979, 132).

The representations above do not say anything about the constraint violation being realized on the initial syllable of the root. Independent considerations guarantee this. To see how, we need to notice an important surface generalization that holds of all the binyanim: they all begin with a bimoraic foot (LL or H) —that is, they respect ALIGN-L (STEM, FT). Realizing a complex nucleus or an additional coda consonant on the second syllable of the root (kE.tEEb, ke.tEbb) would violate ALIGN by creating an unfooted single light syllable at the left edge of the stem; hence the first syllable is made heavy. This accounts for everything but the initial complex onset in binyanim 7-15, which is due to a separate consideration: syllable structure is stricter within a morpheme than at the edge (Steriade 1982, 1988). Again, this makes the best place to complicate syllable structure in a CV.CVC root the initial syllable, CCV.CVC being less marked than CV.CCVC.

Note that the symmetry of the system here is only possible if one treats complex onsets, complex nuclei and simplex codas as parallel. This is not straightforward in Prosodic Morphology (McCarthy & Prince 1986, 1991, 1993a) since onset complexity is irrelevant to the prosodic hierarchy. Within Direct OT, on the other hand, CCV, CVV and CVC form a natural class of deviations away from CV: each violates the basic constraints on onset, nucleus and coda exactly once.

The full set of fifteen triliteral binyanim for Classical Arabic is derived by adding prefixal material, vowel quality and binyan-specific consonant material. From here on, Direct OT mirrors McCarthy's original analysis, aligning consonantal and vocalic features with the left edge of a word and spreading them rightwards when necessary. I won't try and discuss this here due to limits on space and scope.[10]

The one aspect that does require further comment is the infixal material in binyan 8. Here the shape is CtV.CVC, with the second consonant prelinked (in the terminology of another day) to the melody: eg, [k-t-a.sab] 'to earn.' Focusing on the result of all this, which is a blatant violation of CONTIGUITY, we immediately see the analysis: binyan 8 has an infix. The representation in (49) both describes this and guarantees it:

(49) Binyan 8

*CONTIG	*STOP
*	*

The best place to add the stop is after the first consonant and before the first vowel—any later results in a fatal violation of ALIGN-STOP, since the first vowel will create a peak to the left of the inserted stop.

This essentially non-templatic approach has three advantages to it. First, as Mike Hammond (p.c.) points out, it solves the problem with medial gemination in forms like kat.tab (*kat.bab), problematic in the analyses of McCarthy 1979 and McCarthy & Prince 1986. The form kat.tab has identical feature-alignment violations as ka.tab, with [t] one peak and [b] two peaks from the left edge; *kat.bab adds additional and fatal violations of ALIGN-L (LAB, RT) and ALIGN-L (VOI, RT). Second, it greatly simplifies the representation of the various oblique binyanim: the prosodic shapes of binyanim 2-9 are each identified by a single distinctive feature, those of 10-15 by two. Finally, the non-templatic approach taken here makes

[10] Note that the representation *kEtEb* is completely compatible with McCarthy's (1979) claim that the OCP is responsible for non-occurrence of roots like *s-s-m*. As much work on the subject has shown (Mester 1986; McCarthy 1988, 1994; Yip 1988, 1989), constraints against partially or totally identical consonants need not be strictly local. Thus *s-s-m* and *sEsEm* behave identically with respect to the OCP since the intervening vowels in sEsEm have (by definition) no features which could interfere with the OCP. See Pierrehumbert (1994) for discussion.

Arabic look like other languages, reducing the number of basic morphological types from two (regular and interdigitating) to one.

Masdars

As McCarthy (1979, 180) points out, masdars (a type of deverbal noun) generally have the same shapes as their perfective verb counterparts but with the final syllable CVVC rather than CVC.[11]

(50) Masdars

Binyanim	Masdar
2, 4, 5	CVC.CVVC
3, 6	CVV.CVVC
7, 8, 9	CCV.CVVC
10, 12, 13, 14, 15	CCVC.CVVC
11	CCVV.CVVC

Looking at the surface forms in (50) it is immediately clear how the masdars differ from the verbs in (43): the masdars alone have a foot (the final CVVC) that is misaligned with the left edge of the stem. Taking this to be the defining characteristic of a masdar, we may represent them as their verbal counterparts in (44)- (48) plus the distinctive feature in (51):

(51) Masdar

ALIGN-L (FT, STEM)
*

Additional evidence for the distinctiveness of Align-L (Ft, Stem) comes from our final set of Arabic data, the broken plurals.

Broken plurals

As McCarthy & Prince (1990) make abundantly clear, the defining surface characteristic of broken plurals is their shape LHX, where X is L, H or nothing depending on the shape of the second syllable of the singular from which they are formed. Representative plurals and the singulars on which they are based are given in (51):

[11] The formation of Binyan 1 masdars is extremely idiosyncratic and will not be dealt with here; see McCarthy 1979 for discussion.

(51) Singular Plural

mak.tab	HL	ma.kaa.tib	LHL	'office'
jun.dab	HL	ja.naa.dib	LHL	'locust'
sul.taan	HH	sa.laa.tiin	LHH	'sultan'
mif.taah	HH	ma.faa.tiih	LHH	'key'
nafs	H	nu.fuus	LH	'soul'
burd	H	bu.ruud	LH	'robe'

Whereas all of the singular in (51) begin with a foot (H), none of the plurals do, in violation of ALIGN-L (STEM, FT). What the broken plural brings to the singular is thus:

(52) Broken Plural

ALIGN-L (STEM, FT)
*

As the plurals in (53) make clear, the broken plural has no effect on singulars that already begin with LH:

(53) Singular Plural

simaal	LH	samaa?il	LHL	'left hand'
wasiid	LH	wasaa?id	LHL	'court'

This could not be otherwise: the singulars in (53) already violate ALIGN-L (STEM, FT) and it is not a constraint that can be violated more than once because there is only one left edge per stem.

We see then that the verbal and nominal systems of Classical Arabic make use of different types of distinctive constraint violations. The verbal paradigm uses violation of three basic syllable structure constraints (NOCODA, *CMPLXNUC, *CMPLXONS) to form the various binyanim. The nominal paradigm primarily uses two Align constraints to derive masdars from verbs and to derive broken plurals from singulars. The only remaining templates used in Arabic are CCVC and CVCC, each of which has a syllable less than the basic CV.CVC. They are straightforwardly analyzed as violations of PARSESYLLABLE, in conjunction with *COMPLEXONS (CCVC) and NOCODA (CVCC).

Rather than proving problematic for RPM, Arabic morphology seems to yield to it easily; the starting point is getting past the notion of triliteral roots. Once roots are given a prosodic shape (as they are in other languages) the generalizations upon which the morphology is based can be given in terms of pure markedness: syllable structure is complicated, alignment is violated, syllables aren't parsed.

This concludes the discussion of morphology. I've tried to show that most, perhaps all types of morphology can be represented in terms of pure markedness. I've simultaneously shown that violations of many types of constraint can be morphologically distinctive; this includes distinctive violation of constraints governing syllable structure, faithfulness and alignment. Whether every constraint can be distinctive in this way is an open question and will remain one at least until we know what the universal set of constraints is.

6. Reinforced representation and speech production

This section looks at a different line of evidence for Direct OT: speech errors. Unlike the psycholinguistic evidence presented earlier (TOT phenomena and malapropisms), this data tells us nothing direct about underlying representation. Still, it does tell us a great deal about surface representation, about what information is passed along in speech production and how. For this reason, it casts important light on the phonology-phonetics interface and is thus very relevant for linguistic theory.

I will claim that there is a significant relation between *how* features are stored in Direct OT and *which* features are most readily misspoken in running speech. Recall that much information in underlying representation in Direct OT is redundant: the existence of the final consonant in *cap* is signalled by violation of NOCODA, *LAB, *STOP, *LABCODA, *STOPCODA, *ALIGN-L (STOP) and *ALIGN-L (LAB). The existence of the initial consonant, on the other hand, is signalled only by violation of *DORSAL and *STOP.

Let us call the [p] in *cap* a (relatively) reinforced segment and the [k] an un-reinforced segment. Reinforcement comes from three sources: violations of sonority constraints, violations of alignment constraints and violations of coda-constraints.

Sonority restrictions on what can occur in an onset, nucleus or coda reinforce everything but place features in onsets and the vocalic feature BACK. This will be evident from any detailed version of the sonority hierarchy, eg the following taken from Steriade's (1982, 221) analysis of Greek syllable structure:

(43) Sonority Hierarchy: Consonants

[-son, -cont, -voice]:	p, t, k
[-son, -cont, +voice]:	b, d, g
[-son, +cont, -voice]:	s
[-son, +cont, +voice]:	z
[+son, -cont, +nas]:	m, n
[+son, -cont, -nas]:	l, r

Here and in most work on the subject, place-features plays no role in determining sonority (Jespersen 1904; Foley 1972; Selkirk 1982; Clements 1990). Similarly for backness among vowels.[12] The following hierarchy of vowel sonority (Selkirk 1982) is typical in ranking front and back vowels as identical in sonority:

[12] Likewise for (contrastive) rounding, I believe, although discussions of sonority I am aware of do not discuss it.

(44) Sonority Hierarchy: Vowels

Hi:	i, u
Mid:	e, o
Low:	a

One qualification needs to be added here: many languages (eg, English) restrict the number of possible vowel contrasts in stressless positions. Vocalic features are relatively unmarked (and thus un-reinforced) only in stressed syllables.

What all this means for Direct OT is that the place features for consonants and the backness feature for (stressed) vowels are less reinforced (less marked) in underlying and surface representations than other features.

Align-feature violations are the second source of reinforcement in Direct OT. All features except those of the first onset and nucleus violate feature alignment constraints—one per sonority peak that separates the feature from the left edge of the word.

The final source of reinforcement is the coda. Place features, although they are not marked in onsets, are highly marked in codas (Steriade 1982; It™ 1988; Yip 1991).

The least reinforced features according to Direct OT, then are: word-initial place features for consonants and backness features for (stressed) vowels. Remarkably, it is exactly these features which are most commonly involved in speech errors. The most common position error by far is the onset of a word; this is true both of natural corpora (eg, 82% of the consonant-interaction errors in the MIT corpus) and of elicited speech errors (Shattuck-Hufnagel 1987), making errors like the following (from Fromkin 1973) extremely common:

(45) Most common error locus: word-initial onsets

Locket or Ham	(Hockett or Lamb)
Zakoffs and Limmers	(Lakoffs and Zimmers)
Joman Rakobson	(Roman Jakobson)
Fats and Kodor	(Katz and Fodor)

As Van den Broecke & Goldstein (1980) and others have shown, the most common feature errors for consonants involve place:

(46) Most common feature errors for consonants: place

```
a [b]ut-[g]usting meal          (a gut-busting meal)
   |      |                        |    |
  lab    dor                      dor  lab
```

Shattuck-Hufnagel (1986) has shown that the most common vowel errors involve the feature back:

(47) Most common vowel errors: back

the deb[ou]t f[ei]c—debate focuses on
back back

The same study shows that most errors involving the nucleus (79% in her study of 300 such errors) occurred between vowels in stressed syllables.

What is surprising about these facts is (i) the linguistic salience of the *features* involved, place and backness, and (ii) the salience of the *positions* involved, word-initial onsets and stressed syllable-nuclei. It is unlikely that such types of speech error are to be ascribed to what is being encoded or where since both the what and the where are in this case central to the phonology of any language.

RPM provides a simple and independently motivated account of these facts in terms of *how* the features are encoded: the features that are least reinforced (least marked) in phonological representation are the ones that are most likely to run amock in speech errors. English places no constraints on place features in word-initial onsets in English and no constraints on BACK in stressed syllables. Since there is to my knowledge no other explanation of these facts, I take them as providing strong support of the model presented here.

7. Comparison with other proposals

Up to this point I have been at pains to compare Direct OT with standard OT and with the view that underlying representation consists of linear strings of unsyllabified segments. I'd now like to turn to three recent models of underlying representation that also reject central elements of traditional representation. Neef (1994), Hammond (1995) and Russell (1995) have independently proposed representing morphemes as constraints.

Neef argues that recurrent patterns in affixation are best captured using constraints. For instance, he demonstrates that a number of German verbal categories are marked by what Wiese (1995) calls schwa-syllables: prosodically weak syllables consisting of « or syllabic n. Neef proposes to capture this with a constraint (to which many word-classes may be subject) 'Word-form must end in exactly one schwa-syllable'. Additional constraints narrow this down to *«* or *n*, but the general pattern is captured by one constraint, not repeated by a number of affixes.

Russell reaches a similar conclusion on the basis of very different data. He demonstrates that in Nishga a single segment can represent a number of different affixes with similar forms, so that a word that we might expect to be *naks-t-s-t* on morphological grounds surfaces as *naks*, with the stem-final [s] doing the duty of four coronals as it were. Such data, he argues, suggests that affixes are better treated as alignment constraints which force a morphological category to align with a set of phonological features. In this way, any number of homophonous 'affixes' may be satisfied by a single string.

Hammond argues for representing all morphemes as constraints: eg, [k¾t] is suggested as a constraint to which only the phonological form of CAT is subject. On the basis of exceptional stress data in Spanish, he argues that lexical idiosyncracies (such as exceptional stress, or the fact that *cat* begins with a [k]) are best handled as constraints that can be ranked above general properties (eg, regular stress) of the language.

Space precludes an adequate analysis of these three proposals, but I would like to focus here on two things they share. The first is that they merely shift the burden of representation from one area to another: where other theories have underlying representations and no constraints these theories have constraints and no underlying representations. This is most clearly seen in Hammond's case: every UR is replaced by a corresponding constraint.

The second property these proposals share is that they abandon the idea that constraints are universal. This is in itself not objectionable, of course, if some other way of constraining the theory is in place. But it does introduce an odd discrepancy into constraints: there seem to be two kinds of constraints, morpheme-specific ones like [k¾t] (Hammond) and general ones like NoCoda and Parse. Calling both things constraints does not make them the same type of thing.

Whatever the faults of Direct OT, I believe it avoids these two. Rather than creating new constraints for old underlying representations, Direct OT makes constraint-violations do double-duty: they serve as both representations *and* evaluations. It thus *lessens* the burden of representation rather than merely shifting it. As for the second point, no new constraints or types of constraints are added—Direct OT uses a proper subset of the machinery of standard OT, dispensing with unsyllabified segment strings ('UR') and syllabified segment strings ('SR') altogether. Constraints remain universal; morphemes differ merely as to which constraints they violate and by how much.

It remains to be seen, of course, if Direct OT has the same empirical coverage as each of these proposals; this will no doubt be the deciding factor.

8. Conclusion

I've tried to show that constraint violations encode the same type of information as segments strings. If we grant that prosody is underlying, and it seems it must be, it becomes possible to represent lexical items directly in terms of pure markedness.

This has positive results in three areas. In terms of psycholinguistics, Direct OT helps bring some of what is known about language structure to problems in language storage and language processing. In this way it is part of a long line of work that tries to bridge the gap between linguistic and psycholinguistic theory.

In terms of morphological theory, Direct OT provides a principled and completely unified theory of representation. The set of distinctive features used to represent any type of morpheme—from roots and circumfixes to reduplication and templates—is simply the set of well-formedness constraints.

In terms of phonological theory, Direct OT has a number of positive consequences. In terms of OT, it provides an explicity theory of representation which allows two of the five elements of standard OT to be dispensed with. More generally, Direct OT abandons the stipulative use of linear order, allowing for a completely non-linear theory of representation; it derives undespecification from the theory of representation; and it literally reduces representation to markedness.

References

Al-Ani, S. 1978. *Readings in Arabic Linguistics*. Bloomington, Indidana: Indiana University Linguistics Club.
Anderson, Stepehn R. 1992. *A-Morphous Morphology*. Cambridge University Press.
Archangeli, Diana. 1991. Syllabification and Prosodic Templates in Yawelmani, *NLLT* **9**, 231-283.
Aronoff, Mark and Richard T. Oehrle (eds.). 1984. *Language sound structure: Studies in phonology presented to Morris Halle by his teacher and students.* Cambridge, MA: MIT Press.
Bauer, Laurie. 1978. A Descriptive Gap in Morphology. *Yearbook of Morphology* 1: 17-28.
Benua, Laura. (to appear). Identity Effects in Morphological Truncation. In J. Beckman, S. Urbanczyck and L. Walsh, eds. *University of Massachusetts Occasional Papers in Linguistics* 18: Papers in Optimality Theory, Graduate Linguistics Student Association, Amherst, Massachusetts.
Bird, Steven. 1990. *Constraint-Based Phonology*. Doctoal dissertation, University of Edinburgh.
Bird, Steven. 1995. *Computational Phonology*: *a constraint-based approach*. Cambridge University Press.
Brockhaus, Wiebke Gertrud. 1992. *Final Devoicing: Principles and Parameters*. PhD Dissertation, University of London.
Brown, R. and D. McNeill. 1966. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.
Butterworth, Brian L. 1989. Lexical access in speech production. In Marslen-Wilson (ed.).
Chomsky, N. 1994. A Minimalist Program. In Hale and Keyser (eds.), 1-52.
Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In Kingston and Beckman (eds.), 283-333.
Crompton, A. 1982. Syllables and segments in speech production. In A. Cutler (ed.), 109-162.
Crowhurst, Megan. 1991. *Minimality and Foot Structure in Metrical Phonology and Prosodic Morphology*. Ph.D. dissertation, University of Arizona, Tucson.
Cutler, Ann. (ed.) 1982. *Slips of the tongue and language production*. The Hague: Mouton.
Cutler, Ann. 1986. Phonological structure in speech recognition. *Phonology Yearbook* **3**, 161-178.
Eisenberg, Peter, Karl Heinz Ramers and Heinz Vater (eds.). 1992. *Silbenphonologie des Deutschen*. TŸbingen: Gunter Narr Verlag.
Fay, D. and Ann Cutler. 1977. Malapropisms and the structure of the mental lexicon. *LI* **8**, 505-520.
Foley, J. 1972. Rule precursors and phonological change by meta-rule. In Stockwell and Macaulay (eds).
Fromkin, Victoria A. (ed) 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
Fromkin, Victoria A. (ed). 1980. *Errors in Linguistic performance: slips of the tongue, ear, pen and hand.* New York: Academic Press.

Fromkin, Victoria A. 1985. Evidence in Linguistics. In Robbins and Fromkin (eds.).
Goldsmith, John. (ed.) 1993. *The Last Phonological Rule*. University of Chicago Press.
Goldsmith, John. 1979. *Autosegmental Phonology*. PhD dissertation, MIT, 1976. Distributed by IULC. New York: Garland Press.
Goldsmith, John. 1990. *Autosegmental and Metrical Phonology.* Basil Blackwell, Oxford.
Goldsmith, John. 1991. Phonology as an Intelligent System,. In Donna Jo Napoli and Judy Kegl, (eds), *Bridges between Psychology and Linguistics.* Lawrence Erlbaum Associates, Hillsdale, NY.
Goldsmith, John. 1994. *Handbook of Phonology*. Basil Blackwell.
Golston, Chris. 1991. *Both Lexicons*. Ph.D. dissertation, UCLA.
Golston, Chris. 1995. Prosodic constraints on roots and words. Paper presented at the *Berlin conferences on Interfaces in Phonology*.
Greenberg, Joseph. 1978. The patterning of root morphemes in Semitic. In Al-Ani (ed.).
Hale, Kenneth and Samuel Jay Keyser. 1993. *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger.* Cambridge, Mass: MIT Press.
Hale, Kenneth. 1965. Some preliminary observation on Papago morphophonemics. *IJAL* **31**:295-305.
Hammond, Michael. 1993. Heavy trochees in Choctaw morphology. *Phonology* **10**. 325-336.
Hammond, Michael. 1995. There is no lexicon! Ms. Univeristy of Arizona.
Hayes, Bruce. 1989. Compensatory Lengthening in Moraic Phonology. *Linguistic Inquiry* **20**, 253-306.
Hill, Jane H. and Ofelia Zepeda. 1992. Derived words in Tohono O'odham. *International Journal of American Linguistics*, **58**.4:355-404.
Hooper, Joan B. 1975. The Archi-segment in Natural Generative Phonology. *Language* **51**, 536-560.
Hulst, Harry van der and Norval Smith, (eds). 1982. *The Structure of Phonological Representations*, Parts 1 and 2. Dordrecht: Foris.
Hyman, Larry. 1985. *A Theory of Phonological Weight*. Foris, Dordrecht.
Inkelas, Sharon. 1994. The consequences of Optimization for Underspecification. Ms, University of California, Berkeley.
It™, Junko. 1988. *Syllable Theory in Prosodic Phonology*. PhD dissertation, University of Massachusetts, 1986. New York: Garland Press.
Jespersen, Otto. 1904. *Lehrbuch der Phonetik*. Leipzig and Berlin.
Keller, E. and M. Gopnik (eds.). 1987. *Motor and sensory processing in language.* Hillsdale, NJ: Lawrence Erlbaum.
Kingston, John and Mary E. Beckman, (eds.). 1990. Papers in Laboratory Phonology I. Cambridge University Press.
Lakoff, George. 1993. Cognitive Phonology. In Goldsmith (ed.), 117-145.
Leben, Will. 1973. *Suprasegmental Phonology*. PhD dissertation, MIT.
Levelt, W. J. M. 1989. *Speaking: from Intention to Articulation*. Cambridge, Mass.: MIT Press.
Lieber, Rochelle. 1992. *Deconstructing Morphology: Word Formation in Syntactic Theory.* Chicago: University of Chicago Press.
Lombardi, Linda and John McCarthy. 1991. Prosodic circumscription in Choctaw morphology. *Phonology* **8**:37-72.
Marantz, Alec. 1982. Re: reduplication. *LI* **13**, 435-482.
Marslen-Wilson, W. (ed.) 1989. *Lexical representation and process*. Cambridge, MA: MIT Press.
Martin, Jack. 1988. Subtractive morphology as disassociation. *West Coast Conference on Formal Linguistics* **7**:229-40.
Mathiot, Madeleine. 1973. *A Dictionary of Papago Usage*. Bloomington: Indiana University.
McCarthy, John J. 1979. *Formal problems in Semitic phonology and morphology*. MIT doctoral dissertation. (Reproduced 1982 by Indiana University Linguistics Club.)
McCarthy, John J. 1981. A prosodic theory of nonconcatenative morphology. *Linguistic Inquiry* **12**.373-418.
McCarthy, John J. 1982. Prosodic structure and expletive infixation. *Language* **58**, 2. 574-590.
McCarthy, John J. 1988. Feature Geometry and Dependency: A Review. *Phonetica* **43**, 84-108.
McCarthy, John J. 1994. The phonetics and phonology of Semitic pharyngeals. In P. A. Keating (ed.) *Papers in laboratory phonology III: phonological structure and phonetic form*. Cambridge: Cambridge University Press. 191-233.
McCarthy, John J. and Alan S. Prince. 1986. Prosodic Morphology. Ms, University of Massachusetts, Amherst and Brandeis University.
McCarthy, John J. and Alan S. Prince. 1988. Quantitative Transfer in Reduplicative and Templatic Morphology. In Linguistic Society of Korea, ed., *Linguistics in the Morning Calm* 2, Hanshin Publishing Co., Seoul. 3-35.
McCarthy, John J. and Alan S. Prince. 1990. Foot and Word in Prosodic Morphology: The Arabic Broken Plurals. *Natural Language and Linguistic Theory* 8, 209-282.
McCarthy, John J. and Alan S. Prince. 1993a. *Prosodic Morphology I: Constraint Interaction and Satisfaction.* Ms. University of Massachusetts, Amherst and Rutgers University.

McCarthy, John J. and Alan S. Prince. 1993b. The Emergence of the Unmarked: Optimality in Prosodic Morphology. *NELS* **23**.
McCarthy, John J. and Alan S. Prince. 1993c. Generalized alignment. *Yearbook of Morphology*, 79-153.
Mester, R. Armin. 1986. Studies in Tier Structure. Ph.D. dissertation, U Mass, Amherst.
Mohanan, K. P. 1993. Fields of attraction in phonology. In Goldsmith (ed.), 61-116.
Moravcski, Edith. 1977. On Rules of Infixing. Indiana University Linguistics Club.
Neef, Martin. 1994. *Wortdesign. Eine reprŠsentationelle Analyse morphologischer RegularitŠten*. PhD dissertation, UniversitŠt zu Kšln.
Noske, Roland. 1992. Moraic versus Constituent Syllables. In Eisenberg, Ramers and Vater (eds.), 284-328..
Olson, Susan.
Paradis, Carole and Jean-Fran□ois Prunet, (eds.). 1991. *The special status of coronals: internal and external evidence*. Academic Press.
Paradis, Carole. 1988. On constraints and repair strategies. *The Linguistic Review* 6: 71-97.
Pierrehumbert, Janet. 1994. Dissimilarity in the Arabic Verbal Roots. *NELS* **24**.
Port, Robert and Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, 17, 257-282.
Prince, Alan. 1993. Optimality in Theory and Practice. Paper presented at the Trilateral Phonology Weekend, UC Berkeley and UC Santa Cruz.
Prince, Alan. and Paul Smolensky. 1993. Optimality theory: Constraint interaction in generative grammar. Ms, Rutgers University and University of Colorado, Boulder.
Pulleyblank, Douglas. 1986. *Tone in Lexical Phonology*. Dordrecht: D. Reidel.
Pulleyblank, Douglas. 1988. Underspecification, the feature hierarchy and Tiv vowels. *Phonology* **5.2**, 299-326
Robbins, R. H. and Victoria A. Fromkin, (eds). 1985. *Linguistics and linguistic evidence*. Newcastle upon Tyne: Grevatt & Grevatt.
Russell, Kevin. 1995. Morphemes and candidates in Optimality Theory. Ms. University of Manitoba.
Schachter, Paul and F. Otanes. 1972. *Tagalog Reference Grammar.* Berkeley: University of California Press.
Scobbie, James. 1991. *Attributive value phonology.* Doctoral dissertation, University of Edinburgh.
Scobbie, James. 1992. Towards Declarative Phonology. Edinburgh working papers in cognitive science 7, 1-26. University of Edinburgh.
Selkirk, Elizabeth O. 1982. The syllable. In van der Hulst and Smith (eds.), part 2, 337-383.
Selkirk, Elizabeth O. 1984. On the major class features and syllable theory. In Aronoff and Oehrle (eds.).
Shattuck-Hufnagel, Stephanie. 1986. The representation of phonological information during speech production planning: evidence from vowel errors in spontaneous speech. *Phonology Yearbook* **3.** 117-149.
Shattuck-Hufnagel, Stephanie. 1987. The role of word onset consonants in speech production planning: New evidence from speech error patterns. In E. Keller & M. Gopnik (eds.).
Shuy, R. and C. J. Bailey (eds.). 1973. *Towards tomorrow's linguistics*. Washington, D.C.: Georgetown University Press.
Steriade, Donca. 1982. *Greek prosodies and the nature of syllabification*. PhD dissertation, MIT.
Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* **5**, 73-155.
Steriade, Donca. 1994. Underspecification and Markedness. In Goldsmith (ed.).
Stonham, John. 1990. *Current Issues in Morphological Theory.* PhD dissertation, Stanford University.
Ulrich, Charles H. 1994. A unified account of Choctaw intensives. *Phonology* **11**. 325-339.
Van den Broecke, M. P. R. and L. Goldstein. 1980. Consonant features in speech errors. In Fromkin 1980 (ed.).
Van der Hulst, Harry. 1984. *Syllable Structure and Stress in Dutch*. Foris, Dordrecht.
Vennemann. 1973. Phonological concreteness in natural generative grammar. In Shuy & Bailey (eds.)
Wiese, Richard. 1995. *The Phonology of German*. Oxford University Press.
Yip, Moira. 1988. The obligatory contour principle and phonological rules: a loss of identity. *LI* **19**, 65-100.
Yip, Moira. 1989. Feature geometry and cooccurrence restrictions. *Phonology* **6**, 349-374.
Yip, Moira. 1991. Coronals, Clusters, and the Coda Condition. In Paradis and Prunet (eds.), 49-78.
Zepeda, Ofelia. 1983. *A Papago Grammar.* Tucson: University of Arizona Press.