Surface Constraints versus Underlying **Representation** 

# Luigi Burzio

The Johns Hopkins University

# **1** Introduction

The shift from derivational rules to representational constraints that characterizes much contemporary work in phonology calls for a serious reevaluation of the notion of Underlying Representation (UR). The reason is that the ability of constraints to apply "in parallel" systematically obviates the need for a derivation. The purpose of this paper is to argue, consistently with my (1994) Principles of English Stress (PES), that the notion of UR is neither conceptually necessary nor empirically supported, and should be dispensed with.

The paper is structured as follows. In the next two sections I consider the two main motivations for postulating a UR: calculation of regularities of sound structure such as prosody, and calculation of allomorphy. I argue that neither calculation makes UR necessary in a constraint-based framework, and outline a fully "parallel" account of both. I then turn to the empirical evidence supporting the proposed account of allomorphy, an account which links surface forms directly to other surface forms, without UR. I argue in section 4 that the phenomena traditionally attributed to the "cycle" are just this sort of evidence, since they follow directly from our approach, while the need for the "cycle" is an unexpected complication for the derivational alternative. In sections 5-8 I then discuss further evidence that also follows from surface-to-surface links, but fails to reduce to any "cyclic," or even derivational, account, and conclude in section 9.

#### **2 UR and Prosody**

The traditional notion of UR conflates two logically distinct notions. The first pertains to regularities of sound structure within individual words, the prime instance of which is prosody, namely syllable structure and metrical structure, which stand in a generally regular relation to segmental structure. Any theory recognizing such a regularity will by definition comprise a relevant calculation, namely it will postulate some mapping M from a substructure S (like the segments) to some other substructure P (like prosody), or equivalently a mapping M' from S to S&P. Succinctly then, if the statement in (1a) is an accurate observation, then the one in (1b) will have to hold in any theory.

- a. Within individual words, certain aspects of sound-structure "P"
  (e.g. prosody) are calculable from other aspects of sound-structure "S" (e.g. segments).
  - b. There exists a mapping M from S to P (or M' from S to S&P).

In Prince & Smolensky's (1993) "Optimality Theory," (OT) for instance, there is a mapping M' done by Gen and the evaluation procedure combined. Once we grant (1a,b), we can then surely refer to the input to M/M', namely S, as "UR". If we do so, then UR will exist, by definition. It remains an entirely open question, however, whether such UR, which we will now refer to as "UR-prosody" constitutes a separate level of mental representation. The traditional answer to this has of course been affirmative, but so far as we know that answer only rests on the premises in (2), which we need not accept.

- (2) a. M is a set of re-write rules.
  - b. UR-prosody corresponds to lexical storage.

The premise in (2a) is quite generally rejected within the constraint-based perspective of OT and other contemporary work, some of which is represented in this volume. Since we adopt that perspective, we will reject the premise in (2a) here. The premise in (2b) can be challenged as well. The latter merely reflects the presumption that lexical organization aims to minimize storage. If that were true, then indeed only the incalculable aspects of sound structure would presumably be stored, the calculable ones, like prosody, being computed on-line. That presumption, however, is purely conjectural. Given the fact that humans routinely remember countless details of their lives spanning many years, the argument that they are illequipped to remember, say, how to stress *america*, seems less than compelling. More concretely, the claim that storage is minimal entails that, while storage is costly, computation is free, surely an arbitrary claim, given our present state of knowledge. While the traditional hypothesis in linguistics is that storage and computation are organized "in series," as in (3a) below (the box being the storage), a perfectly plausible alternative hypothesis is that they are in fact organized "in parallel," as in (3b) (where the box is again the storage).



While in (3a) storage is limited to the unpredictable aspects of sound structure, in (3b) storage contains fully formed structures. Unlike the calculation M' of (a) that literally compiles prosodic structure, the calculation M of (b) would merely "check" it, ensuring proper relation between the two subparts. As noted, the organization in (3a) does indeed meet the criterion of optimal, i.e. minimal, storage, but now the one in (3b) meets the alternative criterion of optimal "retrieval" in the sense that the computation M is not required every time a lexical item is accessed for use, but essentially only every time an item is stored. See Derwing (1990) and references therein for further relevant discussion. There is therefore no valid conceptual basis for choosing between (3a) and (3b), so that an empirical one will be needed. While we will turn to that empirical basis below, let us consider in the meantime that whenever M/M' has some degree of indeterminacy, that indeterminacy fails to be registered within individual items, as for example in the pairs in (4).

- (4) a. hónest/ \*honést; robúst/ \*róbust
  - b. órchestra/ \*orchéstra; asbéstos/ \*ásbestos

Clearly, for purposes of stress, the items *honest/ robust* of (4a) have non-distinct segmental structures, and yet their stresses differ. This means that, in English, the calculation of stress from segmental structure (via syllable structure) is partly indeterminate. In the analysis of PES, that indeterminacy consists in parsing a final null syllable in  $ro(búst\phi)$ , but not in  $(hónes)t\phi$ . Whatever the exact analysis, however, the point is that there is indeterminacy in the calculation. In the serial system of (3a), that indeterminacy should give variability over different uses of the same item, clearly not the case.<sup>1</sup> A similar conclusion follows from (4b), where the syllable closed by *s* acts like a heavy syllable in *asbéstos*, but not in *órchestra*, and where one would thus expect to find \**orchéstra*, \**ásbestos* just as well, freely. On the other hand, if words are stored fully formed with their prosodic structure, the invariability of each item is obviously accounted for, while the variability across otherwise comparable structures will follow from supposing that the "checking" procedure M of (3b) has the appropriate degree of freedom (See PES, pp. 165ff).

In sum, UR-prosody exists as the input to some calculating function, but there is

little reason to suppose it exists as a specific level of mental representation, since there is neither any reason to suppose the computing function is a set of re-write rules, nor that the lexicon is organized for minimal storage.

# **3 UR and Allomorphy**

The second notion of UR employed by past tradition is the one motivated by the phenomenon of allomorphy. Here, the relevant factual observations are the ones in (5).

- (5) a. There are sublexical structures that share properties of sound and meaning while falling short of identity.
  - b. The divergence between/ among them is (generally) calculable from context.

The "structures" of (5a) are the "allomorphs" of the same morpheme (coallomorphs), such as the bracketed portions of dog[z], cat[s], and those of [*ilektrik*], [*ilektris*]*ity*, [*ilektriš*]*ian*. The traditional approach has been to factor out the aspects common to such co-allomorphs, and ascribe them to a separate level of UR, which we refer to here as "UR-allomorphy". From this, the contextually determined variants would be derived, schematically, and for example, as in (6).

| (6) | a. | Underlying: | /M/<br>/ \ | /ilektrik/                |
|-----|----|-------------|------------|---------------------------|
|     |    |             | / \        | / \                       |
|     | b. | Derived:    | M' M''     | [ilektrik] [ilektris] ity |

The conclusion that (6) is the correct organization, however, turns on the premises in (7), both of which are once again not necessary.

- (7) a. The calculation of (5b) is due to a set of rules.
  - b. Storing only URs gives optimal storage.

That is to say, UR-allomorphy comes in part for the obsolete and already rejected assumption that all regularities are the result of re-write rules. Here the reasoning would be that if a surface form is calculable from context and the calculation is done by re-write rules, then there must be another representation beside the surface, to serve as the input to the rules. In part, UR-allomorphy comes from the assumption, that we also already challenged in connection with UR-prosody, that the lexicon is organized for optimal storage. Storing a single UR would obviously be more economical than storing all of its allomorphs. But, once again, this position is arbitrary since it implies that, while storage is costly, computing allomorphs is free. Assuming storage of full forms instead will just have the complementary virtue of minimizing computation, and will thus need to be seriously considered.

A constraint-based framework such as that of Prince & Smolensky (1993), or that of PES, permits a conception of the lexicon that is fully parallel, not only relative to prosody, but also relative to allomorphy, in a manner that can be illustrated as in (8).



In the schema in (8), analogous to the one proposed in Derwing (1990) and with transparent "connectionist" implications, the traditional notion of "morpheme" is being expressed by the connections, each representing a set of similarities in sound/meaning. Allomorphy, as in the case of *electri[k]/electri[s]/electri[š]* will result from resolution of a conflict between an identity constraint imposed by the connection, which we discuss below, and other constraints that impose adaptation of sound to context, here forcing the stem final consonant to be realized as [s] or [š] in certain environments. Morphological properties traditionally attributed to each affix, such as "attaches to adjectives; yields nouns" are in (8) the properties that define and in fact give rise to the specific connection implicating all instances of that affix. For instance, the connection tying the different instances of *-ity* will have the properties in (9).

- (9) a. There is a substructure *[ity]*, such that:
  - b. It occurs at the right edge of another substructure A, which is:
    - b'. an adjective;
  - c. And the structure [A-*ity*] is a noun whose meaning is:
    - c'. "the property of being A"

Now, in a system such as (8), there appears to be a redundancy between storage and computation, in the sense that, while each individual form is stored, its structure is

also often predictable from the connections. Yet, at least to some extent, that redundancy seems necessary. For instance, in *fidel-ity*, there is no corresponding adjective \*fidel. Analogously, in arbore-al, lun-ar, there are no corresponding \*arbore, \*lune, etc. --the phenomenon of "bound stems". Some storage is thus necessary simply to record this fact. The connection of *fidel-ity* to other items in -ity will still hold, with respect to a proper subset of the properties in (9), namely (a,b,c), but not (b', c'). Storage is also necessary to record the complementary fact that not all the items that would be possible given the connections actually exist, e.g. \*parental-ity, \*coastal-ity, etc. In sum, the "parallel" architecture in (8) does not meet the criterion of economy of storage, since it features storage of all allomorphs rather than of a single UR for each morpheme (the "Full Listing Hypothesis," see again Derwing 1990 and references therein for further discussion). But reliance on storage seems partly justified by the considerable degree of idiosyncrasy encountered. More importantly, the model in (8) achieves an alternative type of economy, namely economy of computation, in the sense that computation of each surface allomorph is not required every time it is used.<sup>2</sup>

The apparent redundancy between storage and computation in (8) need in fact not exist in any strong sense. That is, one can imagine a system of this sort in which the connections could serve as predictors of new forms, making the storage of all forms superfluous. We saw that this was not the case for the items in (8), but those were representative of the "level 1" type affixation, to use Lexical Phonology terms. In contrast, "level 2" affixation (i.e. affixation of -less, -ness, -ful, etc.), as well as compounding, would indeed seem to have those characteristics. This suggests that the connections can have different levels of activeness or strength, and in turn that the storage may in each case only have the complementary level of activeness, with little or no redundancy between the two. The less active storage and more active connections would now correspond to level 2 affixation and compounding, whence their greater productivity, or perhaps vice-versa. That is, the larger size of these classes would make the storage less effective, given the finite storage capacity. In either case, the more active connections would correspond to the relative lack of idiosyncrasy, and in particular to the well known lack of bound stems with those classes. Hence, the apparent gross redundancy between storage and computation in (8) can be overcome by supposing that those two aspects can in fact be "active" to complementary degrees, which has the desired effect of tying greater productivity to relative lack of idiosyncrasy, as seems correct.<sup>3</sup>.

Turning now to the account of contextually determined allomorphy, we take it to arise from appropriate context-based constraints, as in any OT-type framework The distinctive characteristic of the conception in (8) will be, however, that the "base" for the calculation of each allomorph is not a single "UR," but rather the set of all the other co-allomorphs to which the allomorph in question is connected. It will be useful in this regard to consider the characterization of allomorphy from the OT perspective that has UR, as in the diagram in (10).



In the system in (10), co-allomorphs M', M" are derived from a common underlying /M/. In this system, it is necessary to ensure that each allomorph deviate from the UR only minimally, and no more than required by the constraints at work, i.e. not freely. Free deviation would result in the *cat*-means-"dog" syndrome. That is, the surface *[cat]*, unquestionably an optimal structure in itself, could come from the UR /dog/ or some other UR, if free UR-to-surface deviation was allowed. In Prince & Smolensky (1993), this fallacy is avoided by postulating a set of constraints, collectively referred to as "faithfulness," which impose adherence to UR, preventing deletions or additions of segments, features, etc. Surface representation then will typically represent the proper resolution of the conflict between the faithfulness constraints, that impose identity to UR, and other constraints that may impose deviations from it. Hence, given any two allomorphs M', M'', faithfulness will apply between each and the single UR /M/, as in (10). Now, however, it is easy to see that the notion of faithfulness involved in (10) is both symmetric and transitive, since it is essentially the notion of identity. But then the organization in (11), in which there is no UR, and each allomorph is required to be directly faithful to the other must surely be equivalent to the one in (10), at least as a first approximation.



If we now adopt (11), nothing will change when the number of allomorphs is greater than 2. The calculation of each will simply invoke multiple instances of faithfulness, one for each of the other allomorphs, and UR will continue to be superfluous. If we consider in this connection that UR is just an optimal interpolation among surface forms, there indeed seems little point in first calculating a UR, only to use it to recalculate the surface forms back.

In our conception, we will thus take "faithfulness" to be at work directly among surface allomorphs as in (11). We will refer to this use of faithfulness as "antiallomorphy" (AA), and presume it to consist of a family of constraints, concerning whichever aspects of sound structure will seem relevant, such as segments, features, prosodic structure, etc., much as with the UR-based faithfulness of other versions of OT. The way to incorporate AA into the schema in (8) above is now to take the connections to be essentially self-sustaining, that is, as induced by patterns of similarity, but then requiring identity once they obtain. That identity will be our AA. Note now that, if, as suggested above, "level 2" affixation is to be understood in terms of more active connections than "level 1," then the prediction, again correct, is that level 2 affixes should induce less allomorphic variation in their stems than level 1 (since the stronger connections would imply a stronger AA). The cluster of properties we are thus able to express is then: higher productivity; absence of bound stems; relative absence of allomorphy, which seems correct. From a more traditional perspective, we find no particular reason for this cluster.

Summing up, we have argued that there are two notions of UR, one relative to prosody, and one relative to allomorphy. UR-prosody exists in any theory to the extent that prosody is calculable --it will be the input to the calculation. However, it need not exist as a specific level of mental representation. As for UR-allomorphy, it need not exist in any sense, since any surface allomorph can be calculated directly from its surface co-allomorphs. Note that such a many-to-one calculation is possible by means of constraints, which can apply in parallel, collectively, but it would not be possible by means of rules, since the rules can only map one-to-one. Note also that some many-to-one mapping would be necessary in any event to calculate a unique UR from multiple surface forms, a problem rule-based theories overlook.

We will now turn to specific evidence for the proposed conception in which words are stored as fully formed (Full Listing Hypothesis) and both prosody and allomorphy are calculated in parallel by means of appropriate sets of constraints.

#### 4 The Cycle = Anti-allomorphy

One major piece of evidence against UR is that the framework that employs it requires the principle of the phonological cycle. Consider the prototypical example illustrating the need for the cycle in (12), from Chomsky & Halle (1968).

- (12) a. cond[e]nsation b.  $comp[\partial]nsation$ 
  - a'. cond[é]nse b'. cómp[ $\partial$ ]nsàte

The problem posed by (12) is that there is a difference in vowel reduction between (a) and (b), despite the identical stresses. That difference requires relating the items in (a,b) to the ones in (a',b'), respectively, where a difference in stress exists. Now if surface forms are related directly to other surface forms as proposed here, the noted effect is accounted for directly: the failure of vowel reduction in (a) is an AA effect due the connection between that item and the one in (a') (see PES, p. 333f. for slightly more detailed discussion). In contrast, if surface forms were derived from a UR, vowel reduction in (a) and (b) should be identical, given that the latter depends on stress and that the stresses are identical (which in turn reflects the parallel

segmental structures). UR-based accounts have thus been forced to assume that, in the derivation of condensation, the "inner" word condénse is derived first -- the phonological "cycle," vowel reduction applying precisely at that stage, hence being blocked by stress. In contrast, in the derivation of *compensation*, the first cycle would only build *cómpensàte*, with no stress on *e*, hence permitting the reduction. While the "cyclic" hypothesis has played a very important role in helping linguists describe and understand phonological phenomena, it is not difficult to see, about 30 years since it was first introduced, that it lacks any conceptual justification. The reason is that a derivation that did not obey the cycle would be unquestionably simpler. This is clear especially in the domain of stress, where later cycles need to be massively devoted to undoing what earlier ones did (see PES, pp. 187ff.). From our perspective, the need to postulate a cycle is just the admission that words are related to other words, not to UR's. Note that while AA does the work of the cycle, it does not lack independent motivation, unlike the cycle. The reason is that AA is simply our reinterpretation of "faithfulness," which other theories, at least constraintbased ones, also need in some form. In short, cyclic effects follow directly from the organization in (8) under AA, without requiring any further complication.

Certain instances of counterbleeding also reduce to AA, such as the one in (13), found in some Canadian dialects (as discussed in Bromberger & Halle 1989, among others).

(13) a.  $[r \land yt]$  "write" b. [rayd] "ride" a'.  $[r \land yDin]$  "writing" b'. [rayDin] "riding"

In (13), the diphthong *ay* apparently raises to Ay before a voiceless consonant, as in (a), but not before a voiced one, as in (b,b'). In (a'), however, raising is not "bled" by the *t* turning into the voiced flap *D*. A derivational approach would of course simply order raising before flapping. Instead, we can take the raised diphthong of (a') to be anti-allomorphic to that of (a), in turn due to the *t*. That is, we take (13a) to satisfy the constraint-equivalent of the raising rule, say "\* [+low] y [-voice]," and (13a') to satisfy AA, which bars contrasts like rAy.../ray.... On this view, it is evidently the case that either there is no constraint blocking raised Ay before voiced *D* or, if there is one, it must be ranked lower than the AA constraint. On the other hand, there must be a constraint excluding flapping in non-intervocalic positions higher-ranked than AA, lest the latter give \*rAyD for (13a).

#### **5** Countercyclic Anti-allomorphy

One case of AA studied in detail in PES is metrical AA, yielding similarity of stress

in the manner of *medícinal/ medìcinálity* and many other cases, formerly also attributed to the phonological cycle. There are a number of metrical AA effects that the cycle cannot deal with, however. One of these concerns the pairs in (14).

- (14) a. prevénting b. académical
  - a'. prevént b'. académic

As discussed in PES and elsewhere, bare verbs and adjectives in *-ic*, as in (a',b'), stress quite generally as if they had one more syllable. That is, postulating one invisible but metrically parsed syllable for these two classes, along the lines suggested above for  $ro(búst\phi)$ , would reduce their stress patterns to that of the majority of items. The question, however, will be why this should be a systematic property of these two classes, while being only a sporadic one of others. The answer given in PES (pp. 244ff.) is that bare verbs like (a') are being anti-allomorphic to their inflected variants like (a), in which the "missing" syllable is in fact present as -ing. Similarly, adjectives in -ic like (b') are taken to mirror their variants in -ical like (b) (as originally proposed in Chomsky & Halle, 1968), which again realize the missing syllable. Now the cycle will be of little help here since the anti-allomorphy effect obtains derivationally in reverse. That is, a serial account would have to take the forms in (14a,b) as basic, assigning stress to them first, and then de-affix them, to obtain the ones in (14a',b') while allowing the earlier stress to remain. This was actually proposed in Chomsky & Halle (1968, p. 88) for (14b,b'), but has limited initial plausibility, and none when generalized to (14a,a'). We also note cases like Bérnardine/ Bèrnardína and other similar pairs noted in Burzio (1987), where the stress of the first member can only be interpreted in relation to (i.e. as antiallomorphic with) that of the second (the expected stress of the former in isolation being \*Bernárdine). Again a serial account would presuppose a rather implausible truncation ordered after stress.

Thus, we take the fact that consistency of form obtains sometimes "forward" into a morphological derivation, as in *cond[é]nse/ cònd[e]nsátion* of (12) above, but sometimes "backwards," as in the cases in (14), to argue against the traditional derivational model and for the declarative one in (8) above. It seems, however, that some of the asymmetry of the derivational model must be re-introduced into (8). A totally asymmetrical hypothesis would incorrectly predict, e.g., unreduced *[e]* in *\*cómp[e]nsàte* by anti-allomorphy with *comp[é]nsatory*, just as it predicts *cònd[e]nsátion* from *cond[é]nse*. The relevantly asymmetrical notion here is obviously that of "containment," the word *condensation* containing *condense*, while the word *compensate* does not contain *compensatory*. To be more explicit, let us take the connections in (8) to be congruous with the psycholinguistic notion of "priming," that is the notion that retrieval of one lexical item facilitates retrieval of other lexical items similar to the former along some dimension. We would then say that *condensation* primes *condense*, while *compensate* does not prime *compensatory*, at least not equally, which I suspect s psycholinguisticaly correct. The connections in (8) are thus taken to be essentially one-way. The cases in (14) would be exceptions to this, however, since the items in (a',b') appear to be connected with, or "prime," those in (a,b), while not containing them. We can take this fact to stem from the particularly close relationship within each pair, the items in (a,a') being members of the same verbal paradigm, while those in (b,b') are virtual synonyms. For present purposes we may simply suppose that under categorial identity (V/V; A/A; etc.) the connections in (8) are symmetrical, while being asymmetrical as per containment otherwise (i.e.  $\alpha$  is connected with  $\beta$  only if  $\alpha$  contains  $\beta$ ). Assuming something along these lines, the cases in (14) will then continue to follow from the parallel model in (8), while a derivational-type account would require an implausible reversal of the cycle as noted.

### 6 Anti-allomorphy of Affixes

A further piece of evidence for the organization in (8) is that AA obtains with stems and affixes alike. Consider (15).

(15) napoleónic/ \*napoléonic/ ?\*napoleonical

The items in (15) show that, in contrast to other phonetically monosyllabic suffixes like *-al* or *-ous*, *-ic* is metrically bisyllabic quite generally (yielding presuffixal stress), and not just in the cases in which there is an *-ical* variant as in (14b,b'). This fact follows from supposing that, beside satisfying stem AA over *-ic/ -ical* pairs as discussed, items in *-ic* also satisfy AA relative to all occurrences of the suffix itself, which will thus parse always as  $ic\phi$ ), in the analysis of PES. Note that *-ic* continues to be metrically anti-allomorphic even when this in fact violates AA for the stem. So, if *-ic* could parse as a single syllable, it would permit, for instance, *\*títanic*, consistent with *títan*, but that is not the case. On the reasons why affix AA prevails over stem AA here, see PES (p. 302ff.). The cases in (16) lead to a similar conclusion on the existence of affixal AA.

- (16) a. leave/left; keep/kept
  - b. weave/ weav[d]; seep/ seep[t]

While the "irregular" past tenses of (16a) are commonly regarded as quite different from the "regular" ones in (16b), the voicing assimilation affecting both cases is in fact transparently much the same phenomenon. As argued in PES (p. 275f.), the fact

that it is "progressive" in (b), but "regressive" in (a) follows from supposing that the regular affix imposes a strong form of AA on its stems, while the irregular one does not, as is independently clear from the difference in vowel shortening. Then, the different direction of voicing assimilation is interpretable as stem AA in (b), versus suffix AA in (a), implying that suffix AA therefore exists. That is, if the suffix in (16a) was itself totally free to alternate between [t] and [d], there would be no reason for the voicing alternation of *leave/left*. Rather, *leave/\*levd* should obtain instead. Derivational approaches are bound to miss such consistency of suffixes, because the latter obtains across independent derivations. For the similarity across, e.g., *academ-ic* and *napoleon-ic* there are palliative solutions in the derivational literature, in the form of metrical markings in UR. Suffixes -al, -ous, etc., but not *-ic* are taken to be "extrametrical" (in Hayes's sense). The introduction of metrical structure into UR, however, has the effect of making UR relevantly non-distinct from the surface, and is an implicit admission that the surface, not UR, is relevant to allomorphy --exactly our argument. As for the case of *leave/left*, derivational solutions will be forced to simply state that voicing assimilation is regressive, in contrast to that of *weaved*. That, however, will fail to capture the noted fact that stem allomorphy for voicing correlates with stem allomorphy for vowel shortening (which in turn results from resyllabification: PES, p. 63ff.).

### 7 More on Transderivational Anti-allomorphy

There are other cases in which AA obtains across what would be different derivations and which are thus inexpressible in a derivational framework. These cases continue to be expressed by the connections in (8), because the latter transcend derivational relations, affecting all pairs that share morphological material. Consider the Italian alternations in (17).

| (17) | a.  | comico/ comici;<br><i>comical-S/ -PL</i> | cattolico/ cattolici<br>catholic-S/ -PL |
|------|-----|--|---|
|      | a'. | comic-issimo;<br>extremely comical       | comic-ità<br>comicality                 |
|      | b.  | antico/ antichi;<br>ancient-S/ -PL       | etrusco/ etruschi<br>etruscan-S/ -PL    |
|      | b'. | antich-issimo<br>extremely ancient       | antich-ità<br>antiquity                 |

In (17a), stem-final [k] palatalizes to  $[\check{c}]$  in the (masculine) plural, while in (17b) it

remains [k]. From our perspective, this means that the conflict between stem AA, that would inhibit palatalization, and some appropriate constraint that would impose it before *i*, is resolved differently in the two cases. We may in fact take the two constraints to be essentially of equal rank, and the resolution of the conflict to be just idiosyncratic. The contrast between (a') and (b') now illustrates the further fact that, for each stem, that resolution is always the same. That is, if one stem palatalizes before one *i*-initial suffix, it will palatalize before all such suffixes. This follows from AA. Palatalization of stem-final k with any one suffix creates a palatalized allomorph of the stem. But once that allomorph exists, any other suffix will be able to utilize it without creating any *further* violation of AA, hence satisfying new instances of the palatalization constraint for free, and giving the observed "all-or-none" effect. For a derivational approach, there is no reason for this effect. Since it is clear that palatalization would need to obtain in some derivations but not others, there is no natural way to ensure the invariant outcome for each individual stem. One has to diacritically mark each stem in UR, indicating whether or not it undergoes palatalization, rather than just allowing the palatalization rule to be optional, which would be the conceptually natural solution. Once again, the ad-hoc encoding of surface properties into UR simply concedes the irrelevance of UR compared with surface representation.

Another case of transderivational AA is that provided by the tendency for paradigms to be "uniform". Harris (1973) noted certain stress shifts that occurred between Latin and Spanish, such as the one in (18).

| b.   | Spanish: | cantába  | cantábamos   | cantábais    | cantában   |
|------|----------|----------|--------------|--------------|------------|
| a.   | Latin:   | cantá:ba | canta:bá:mus | canta:bá:tis | cantá:bant |
| (18) |          | 1SG      | 1PL          | 2PL          | 3PL 'sing  |

In (18), the rest of the singular is like the first person in having the stress on the thematic vowel (TV) in both Latin and Spanish. The third plural also has identical stresses in both languages. The question is why Spanish shifted the stress of the first and second plural, making the stress effectively uniform, placing it on the TV throughout. To deal with this shift and other related observations, Harris (1973) proposed that paradigm uniformity should be recognized as a principle of phonological theory. Since that proposal, several other accounts of Spanish verb stress have been attempted by Harris and others, but it would be easy to show that they are all merely descriptive, as they invariably attribute to the Spanish system properties which are not independently necessary. In contrast, Harris' (1973) solution would have been explanatory, had it turned out that paradigm uniformity was indeed a principle of UG. In the light of the present discussion it is now evident that Harris' paradigm uniformity is simply AA, indeed a principle of UG on our proposal. That

is, the paradigm in (18b) is uniform in the sense that the forms that make it up are metrically anti-allomorphic with each other, the stress being fixed on the same morpheme. The reason why such AA obtains only within paradigms is that members of a "paradigm" are tightly connected in the sense of (8) above (essentially by definition of paradigm), the relatively tight connections then inducing a relatively strong form of AA. The reason the same AA effect did not show up in Latin is simply that in Latin the penultimate syllables in all of (18a) were heavy, and in addition that the Latin stress system was entirely deterministic, penultimates attracting stress if and only if they were heavy, hence leaving no space to AA. In contrast, all of the penultimate syllables in the Spanish paradigm are light due to the Romance loss of vowel length, and it is independently clear that with light penultimates Spanish allows a choice, as in *pistóla* 'pistol,' versus *fábrica* 'factory'. It is that choice that is utilized by AA in (18b). There is therefore no reason to settle for the less than satisfactory conclusion of Roca (1992) that Spanish has two different stress systems: one for non-verbal categories, basically algorithmic though allowing some variation, and one for verbs, basically lexical, which associates stress with the TV. As argued in PES, (metrical) consistency, alias AA, will always obtain to the extent that it can. That extent corresponds to the degree of indeterminacy of the core stress principles in each language, a degree which is null in Latin, but not in Spanish. English is also like Spanish in this respect, its metrical indeterminacies (binary/ ternary feet; parsed/ unparsed final null syllables) allowing the two AA effects in (19), where the location of the consistent stress is in **boldface**.

- (19) a. medícinal/me(dìci)nálity
  - a'. phenòmenólogy/ phe(nòmeno)lógic
  - b. propagánda/ propa(gándis)t $\phi$
  - b'.  $américan/américa(n)st\phi$ )

The difference is therefore not between verbs and other categories, but rather between sets to which AA is relevant, such as the paradigms in (18) and the pairs in (19) (and (14) above), and those to which it is not, e.g. Spanish *pistóla/ fábrica*. Indeed, to the extent that Spanish nouns and adjectives also form paradigms, they also give rise to AA/ uniformity, as Harris (1973) had correctly noted, and as in (20).

| (20) | pa <b>pé</b> l/ pa <b>pé</b> les; | señór/ señóres; | útil/ útiles; |  |
|------|-----------------------------------|-----------------|---------------|--|
|      | paper-S/-PL                       | gentleman-S/-PL | useful-S/-PL  |  |

The apparent puzzle here is in part that C-ending nouns and adjectives stress as if they had one more syllable, i.e. on the final or penultimate rather than the

penultimate or antepenultimate, and in part that their plurals stress normally, despite their final C. From the AA/ uniformity perspective, the two parts of the puzzle solve each other, once we note that plurals of C-ending singulars have an extra syllable. It is then only if C-ending singulars parse if they had one more syllable (in the analysis of PES, as in  $pa(p \neq l \phi)$ ), that they can be consistent with their plurals (or their feminines, for that matter, e.g. señóra). The plurals, on the other hand, must not parse as if they had one more syllable to make that consistency possible. In other words, the pairs in (20) are much like the English ones in (14) above. The limits of Roca's lexical stress account are also evidenced by the fact that not all cases of paradigm uniformity cited by Harris (1973) are in fact metrical. An instance of this is the paradigm of verbs like *cocer*, in which the velar of Latin has undergone softening throughout, both where a front vowel motivates it, as in cos-er 'to cook,' *cos-emos* 'we cook,' and where there is no front vowel, as in *cuez-o* 'I cook' (all  $[\theta]$ or [s], dialectally). The latter form is thus only interpretable in terms of uniformity with the other forms. Note that the verb *cocer* is related to the noun *coc-ción* (*[ks]* or  $[k\theta]$ , dialectally) 'cooking,' which preserves the stem velar. This is consistent with AA, since this relation is more distant than those within the verb paradigm, thus enforcing a weaker (lower ranked) form of AA. In contrast, such noun would make it impossible for a traditional analysis to claim that the  $[\theta/s]$  of *cocer* is simply underlying, hence leaving the noted uniformity unresolved. (For further discussion, and critique of Roca 1992, see Vincent, 1995).

Derivational accounts of such intra-paradigmatic consistencies thus seem precluded to the extent that the different items within a paradigm need to be derived independently from one-another. It is true, although it has never been proposed, that in (18b) one might attempt to postulate a stress "cycle" for the common structure *[cantaba]*, whose effects would be preserved upon affixation of the person-number marker in each case. Even aside from the *cocemos/ cuezo* case on which this would shed no light, however, the cycle would once again fail for the singular/ plural cases in (20), where it would again have to apply backwards, since it is the plurals rather than the singulars that have the regular pattern.

Our proposed AA account of Spanish (18b) will of course raise the question of why Italian, which also exhibits the penultimate/ antepenultimate stress indeterminacy, in fact maintains the Latin stress (*cantávo*, ... *cantavámo*, ...). While a full analysis of Italian is beyond the scope of this paper, one appears possible in terms of a different instantiation of AA than in Spanish. Plainly, in paradigms such as (18), as well as across related paradigms, there will be several morphemes in competition for metrical consistency/ AA: stem, TV, TA marker, PN marker, a competition always limited to the window of indeterminacy of the stress principles as noted. It is to be expected that the specific resolution may vary under slightly different circumstances. Now Italian appears never to deviate from the stress *-ámo*, *-áte* in the first and second plural, hence satisfying AA for those morphemes, which their Spanish counterparts violate. This seems predictable. In the present indicative,

Italian has the sequence *-iamo*, as well as both *-iamo* and *-iate* in the present subjunctive. Such *iV* sequences are apparently metrically bisyllabic (in Italian much as in English). This effectively puts all other morphemes beyond the reach of stress. So, compare Spanish *amámos/ vendémos/ vivímos/* 'we love/ sell/ live' where the stress is on the TV consistently with (18b), with the Italian counterparts *amiámo*, *vendiámo*, *viviámo*, where there is no TV and stress has no choice but to fall on the PN marker. In short, Italian is less able to satisfy consistency of stress on the TV than Spanish for largely independent reasons, and this is why it instantiates somewhat different patterns of consistency in its verbal paradigms. For further discussion of the Spanish/ Italian difference, generally compatible with the present perspective, see Vincent (1995).

In conclusion, the lack of allomorphic variation instantiated in the cases in (17), (18), (19), (20) above has no derivational expression, since it occurs across what would most plausibly be independent morphological derivations. In contrast, it reduces in each case to the AA induced by the proposed connections in (8), which stand for morphological relations more generally, rather than just those provided by derivations. Such uniformity is found most persistently within "paradigms" because paradigms are clusters of closely related items: the most immediate domain of application of AA.

#### **8 Metrically Conditioned Suppletion**

Our last argument for the organization in (8) above, in which AA is enforced over a network of interconnected surface forms, is given by certain patterns of metrically conditioned deletion and suppletion in Italian, illustrated in (21) (and discussed in Carstairs 1990; Di Fabio 1990; PES 10.2; Burzio & Di Fabio 1994).

| (21) | 1SG       | 1PL      | 2PL     | 3PL              |          |
|------|-----------|----------|---------|------------------|----------|
| a.   | fin-ísc-o | finiámo  | finíte  | fin-ísc-ono      | 'finish' |
| b.   | ésc-o     | usc-iámo | usc-íte | ésc-ono          | 'exit'   |
| c.   | vád-o     | and-iámo | and-áte | v ád-ono [vánno] | 'go'     |

What the examples in (21) illustrate is the general fact that the morphemes *isc*, *ésc*,  $v\dot{a}d$  occur only if stressed. In contexts in which they would not be able to receive stress, the first one --a semantically empty infix characteristic of most verbs in *-ire*, simply disappears, and the other two are replaced by their suppletive forms *usc*, *esc*, respectively. The distinction between suppression and suppletion seems straightforward: semantically empty elements can just be dropped, while contentful

ones must be replaced. This leaves us with the generalization in (22).

(22) \* unstressed -isk-, esc-, vad-

In theories in which stress is calculated serially from UR, (22) will be a complete mystery, requiring an ad-hoc surface filter to state it. In contrast, on the present proposal, (22) is just one instance of AA. That is, the three morphemes of (22), which occur stressed as in (21), resist restressing, because morphemes in general do, as we saw in the discussion of (14), (15), (18), (19), (20) above. As an instance of AA, (22) is surely a violable constraint. It happens to be unviolated because there are systematic ways available -- suppression/ suppletion, to avoid the violation. Note that we are not claiming that in (21) there is compliance with AA in a broad sense. Clearly, at least the alternations of (21b,c) will violate AA by alternating ésc with usc and vád with and. Our claim is rather only that there is no violation of metrical AA, and that is what we can predict. To see this, consider a segmental structure with a specific semantic content such as *vad*, or *part* which was forced into allomorphy by two different metrical environments. There will be two possible responses to this. One is to accept metrical allomorphy, as is the case with most items, e.g. párt-o/ part-iámo 'I leave/ we leave'. But the other is to create an alternative segmental structure with the same semantic content --a suppletive form, precisely as in vád-o/ and-iámo. This second choice will violate AA segmentally, but not metrically, since each segmental structure is associated with a unique metrical structure. Our approach entails exactly these two logical possibilities, and thus explains why metrically conditioned suppletion exists at all, although its exact distribution remains of course a matter of lexical idiosyncrasy. Metrically conditioned suppletion is a way to avoid metrical AA by creating segmental allomorphy instead. In frameworks that assign metrical structure derivationally, there can be no explanation for metrically conditioned suppletion, because there can be no notion of metrical AA, except for the limited one provided by the "cycle," which, however, is obviously of no assistance here.

# 9 Conclusion

In this article, I have argued that the traditional notion of UR is an artifact of rule systems, and that it should be eliminated in favor of a fully parallel conception of the lexicon, possible in an OT-type framework. I have considered a number of cases in which the correct calculation of allomorphy phenomena requires postulating direct connections between surface forms with no intervening UR, and have argued that the notion of phonological cycle is itself an artifact of rule systems and their UR, the relevant phenomena being merely a subclass of those that result from the direct word-to-word connections. The above thesis that surface identity constraints and not

the cycle are at work in allomorphy is independently put forth in Kenstowicz & Flemming (this volume), who provide considerable additional evidence for it, although in that work the authors do not pursue what we have taken to be its major implication --the non-existence of UR.

We believe that the UR-less, parallel, conception proposed will provide for greater convergence with work in psychology, which, so far as we know, has as yet yielded no evidence for serial derivations. It will also make work in "connectionism" more directly relevant to the cognitive study of language, as noted by Derwing (1990). We may note as well that the proposed model has non-trivial consequences for the understanding of language acquisition. In essence, it would make language acquisition monotonic, avoiding the need for radical reanalyses in its course. Consider that, on the proposed organization, while learning morphologically complex words, e.g. *clar-ity*, the child could first enter them into the mental lexicon directly, and establish the proper connections (both to *clear* and to other items in *ity*) only later, as evidence for them becomes available, never having to change the lexical entry itself. In contrast, on the traditional conception, in the absence of morphological evidence, the child would presumably first have to enter *clarity* as a lexical item, only to have to discard that entry entirely later on, once having acquired the independent items *clear* and *-ity* and relevant morpho-phonological rules or principles. The "monotonic" hypothesis seems much more plausible.

#### NOTES

- 1. As J. Durand (p.c.) points out, there are cases of variability across dialects, such as *ábstract/ abstráct* (Adj.), and perhaps even variability within individual speakers. This is consistent with the text, the point only being that there is no massive variability of the kind that the serial approach would predict.
- 2. This point holds both for production and for recognition, as a lexical item would require neither composition from its parts (production), nor parsing into its parts (recognition) on the proposed view.
- 3. I must leave open the question of the exact formal expression of the different "activeness" of the storage relative to that of the connections. I argue below that the more active connections of level 2 affixes correspond to a higher ranked antiallomorphy constraint holding for their stems. However, the level of activeness must be the cause of the exact constraint ranking, and hence not reducible to it.

#### REFERENCES

Bromberger, S. & M. Halle (1989) Why Phonology Is Different. *Linguistic Inquiry* **20,1**. 51-70.

Burzio, L. (1987) English stress. In P.M. Bertinetto & M. Loporcaro (eds.) *Certamen Phonologicum. Papers from the 1987 Cortona Phonology Meeting*. 153-175. Turin: Rosenberg and Sellier.

Burzio, L. (1994) Principles of English stress. Cambridge: Cambridge University Press.

Burzio, L. & E. DiFabio (1994) Accentual stability. In M.L. Mazzola (ed.) *Issues and Theory in Romance Linguistics (LSRL XXIII)*. Washington, D.C.: Georgetown University Press.

Carstairs, A. (1988) Some implications of phonologically conditioned suppletion. In *Yearbook of Morphology 1*. 67-94.

Carstairs, A. (1990) Phonologically conditioned suppletion. In W. Dressler, H.C. Luscützky, O.E. Pfeiffer & J.R. Rennison, (eds.) *Contemporary Morphology*. New York: Mouton de Gruyter.

Chomsky, N. & M. Halle (1968) *The Sound pattern of English*. New York: Harper and Row.

DiFabio, E. (1990) *The Morphology of the verbal infix /-isk-/ in Italian and in Romance*. PhD Dissertation, Harvard.

Derwing, B. (1990) Morphology and the mental lexicon: psycholinguistic evidence. In W. Dressler, H. Luscützky, O. Pfeiffer & J. Rennison, (eds.) *Contemporary Morphology*. New York: Mouton de Gruyter.

Harris, J. (1973) On the ordering of certain phonological rules in Spanish. In S.R. Anderson & P. Kiparsky (eds.) *A Festschrift for Morris Halle*. New York: Holt Rinehart and Winston.

Kenstowicz, M. & E. Flemming (this volume) Base-identity and uniform exponence: alternatives to cyclicity.

Prince, A. & P. Smolensky (1993) *Optimality theory: constraint interaction in generative grammar*. Ms, Rutgers University, New Brunswick, and University of Colorado, Boulder. To appear, MIT Press.

Roca, I. (1992) Constraining extrametricality. In W. Dressler, H. Lüschutzky, O. Pfeiffer & J. Rennison (eds.) *Phonologica 1988*. 239-248. Cambridge: Cambridge University Press.

Vincent, N. (1995) On paradigmatic stress. Ms, Manchester.