LEXICON VERSUS GRAMMAR IN ENGLISH MORPHO-PHONOLOGY: MODULARITY REVISITED*

Luigi Burzio
Department of Cognitive Science
Johns Hopkins University
Baltimore, MD 21218
USA
burzio@jhu.edu

Lexicon and grammar are different by definition, but there is no sense in which they constitute different 'modules'. Phenomena like lexically-controlled-phonology and phonologically-controlled-morphology reveal mutual dependencies of resources that call for a single parallel architecture rather than a serial arrangement. In addition, the phenomenon of unproductive morphology, which is neither pure lexicon nor pure grammar but a mixture of the two, requires a framework in which the respective resources can interact additively rather than in a modular fashion. This article examines English evidence that challenges the traditional modularity, and proposes a parallel framework using distributed representations.

1. Introduction

The first part of the paper argues that that Lexicon, Morphology and Phonology are part of a single parallel architecture along the lines of Optimality Theory (OT). The second part of the paper, however, challenges some of OT's discrete resources like strict ranking of constraints, and argues that capturing the interaction between components requires subsymbolic representations.

2. Parallel Phonology

The most compelling argument for a parallel organization of grammatical

* This work was presented at the ***Error! Main Document Only.****KALS-KASELL International Conference on English and Linguistics*, held at Pusan National University, June 29-30, 2006. I am grateful to that audience for valuable comments, and to two anonymous referees for this journal, for suggestions that led to significant improvements.

resources is provided by mutual dependencies. Some structural property A depends on some other property B, but also vice-versa. In this situation, one cannot first calculate A and then move on to B, nor vice versa, but must rather calculate both A and B simultaneously, using a system of parallel constraints as in Optimality Theory. In his *Thematic Guide to Optimality Theory* J. McCarthy (2002, 144, 182) refers to this critical criterion as 'chicken and egg'. One of the first chicken-and-egg arguments in phonology was presented in Burzio (1991), based on the interaction of stress and vowel length in English. In the 'Level 1' or 'Latinate' lexicon, vowel length is controlled by the simple generalization in (1).

(1) Vowels are short unless length is needed to keep the stress of the base.

This generalization is illustrated in (2), which also provides the constraint-based analysis of Burzio (1991), (1994, 10.3), (2000).

(2) **Stress and Vowel Length**			*V̲ː	Preserve Stress
a.	Short:	dèfa̲m-á̲ːtion	✔	—
b.	Variable:	blásphe̲m-ous	✔	*
b'.		desí̲ːr-ous	*	✔
c.	Short:	divín-ity	✔	✔

In (2c) a short vowel does not preclude stressing *divín-ity* like its base *divi̲ne*, because English freely allows stress on light antepenultimate syllables, as in *América*. In (2a), there is also no interference by stress, but for a different reason. Adjacent stresses are quite generally ruled out, hence vowel length can be of no assistance in maintaining the stress of *defá̲ːme* as in **defà̲ːm-átion.* Interaction occurs only in the cases in (b, b'), which implicate syllables that -in the analysis of Burzio (1994)- are penultimate in the accentual foot. English stresses such syllable only when they are heavy: *Ari(zó̲ːna)*, *a(génda*). The short vowel of *blásphe̲m-ous* thus precludes the stress of *blasphé̲ːme*, while the long one of *desí̲ːr-ous* is needed to keep the stress of *desí̲ːre*. The variation of (2b, b') can be accounted for by taking the two constraints of (2) to be even ranked, yielding equally optimal outcomes, a point to which I will return.

In this maximally simple account of the generalization in (1), vowel length in the Latinate lexicon is effectively allophonic and driven by stress rather than contrastive as in the underived lexicon. Because of the chicken-and-egg relation between stress and vowel length, however, this account is unavailable to a serial approach. The reason is that, in order to decide if a vowel has a chance of being long (foot penultimate), one needs the sequential order in (3a) below. However, to decide where stress actually is in the cases in (2b, b'), one needs the sequential order in (3b), resulting in a paradox.

(3) **Sequential paradox**:
a. Calculate foot structure; then calculate vowel length.
b. Calculate vowel length; then calculate foot structure (stress).

A serial account is never outright impossible in the face of chicken-and-egg

evidence, of course. One can always be mechanically constructed under some loss of generalization (e.g. by postulating multiple shortening rules). On the other hand, theoretical arguments never come in a stronger form.

3. Parallel Morpho-phonology

The introduction of Optimality Theory (Prince and Smolensky 1993, 2004) has brought in a parallel architecture within the Phonology, but has by and large kept a serial organization between modules, as illustrated in (4), with the Lexicon and Morphology providing barebone assemblies of morphemes called 'Underlying Representations', which are then fed to the Phonology.

(4) **Lexicon/ Morphology** Underl.Repr. **Phonology** Surface Repr.

A main goal of the present paper is to show that mutual dependencies such as the one just discussed obtain not only within Phonology, but also across the different components, leading to the more radical parallel architecture of (5).

(5)

Surface Repr.	**IO-FAITH (Lexicon)**
	MARKEDNESS (Phonology)
	OO-FAITH (Morphology)

In (5), there are no Underlying Representations, but only stored surface forms. Such forms are evaluated by a constraint hierarchy consisting of the three types of constraints shown. Each surface form is both the input and the output of the evaluation, in conformity with Prince and Smolensky's 'Lexicon Optimization', to which I return below. A difference between an input and its output can exist in this conception only hypothetically or transitorily. Once the optimal output is calculated from some input, that output and not the original input is stored in the mental lexicon, becoming the steady-state input-output. The grammar in (5) thus 'checks' surface forms for well-formedness. The first two classes of constraints in (5), INPUT-OUTPUT FAITH and Markedness, representing essentially the Lexicon and the Phonology respectively, are the same as in mainstream OT, although IO-FAITH has a broader range of application, as we will see. The third class of constraints, OUTPUT-OUTPUT FAITH, an instance of which was 'Preserve Stress' of (2), has also been featured in much work within OT (see Downing, Hall, and Raffelsiefen 2005 and refs.). Unlike other work, however, in (5) OO-FAITH is taken to constitute the Morphology proper, in the form of constraints like the ones in (6).

(6) a. *-ive* ⇒ / V ___ 'The left context of *-ive* must match a verb in the lexicon' (BASE IDENTITY)

b. A_1 ⇒ /X ___ where X= /___ A_2 'The left context of affix A_1 must match the left context of affix A_2' (PARADIGM UNIFORMITY)

The statements in (6) express the selectional properties of affixes and are taken to be violable constraints in OT. The one in (6a) is violated for example by

compuls-ive, given *compel*, as well as by *gener[ə]t-ive*, given *gener[ey]te*. The one in (6b), pertaining in general to inflectional paradigms, will account for example for sameness of stress in *pervért* (involving a null affix) and *pervért-ing* (compare the noun *pérvert* that does not enter into such a paradigm).

In the next two sections I discuss two phenomena in particular that falsify the serial arrangement in (4) and call for (5) instead.

4. Phonologically controlled Morphology

Morphological selection does not always work purely by lexical category as in the case of *-ive* in (6a). Sometimes affixes require their bases to have specific phonological properties as well. The overall typology of cases is tabulated in (7), along with the theoretical tools that would be available to a serial approach.

(7)	**Types of Phonologically controlled Morphology**	SERIAL APPROACH
a.	Phonological structure of stem: *dismíss-al*	Cyclic application
b.	Full phonological structure of word/ phrase including added morpheme:	
i.	Allomorphy is suppletive: *a/ an*; Italian *vád-/ and-*	Multiple Inputs
ii.	Allomorphy is phonological: *sof(t)-en*	—no recourse —

The case in (7a) is instantiated by deverbal *-al*, which requires a base with final stress, as in *remóv-al, dismíss-al, conférr-al* (cf. **inhérit-al, *prómis-al, *énter-al*). In this type of case, the phonological structure of the base is sufficient for morphological selection, and could be provided by the model in (4) if this was applied cyclically as in Kiparsky's (2000) 'Lexical Phonology and Morphology OT'. The cases in (7b) are different, however, in that they require the fuller phonological structure that includes the morpheme being added. The English *a/ an* allomorphy for example requires full syllabification as in *a.pear*, *a.n ap.ple*, each being optimal compared with **an. pear*, **a. ap.ple*. Similarly, the Italian case requires stress of the full word, then selecting *vád-* when stress falls on the stem, as in *vád-o* 'I go', but *and-* when it falls on the affix, as in *and-iámo*. The potential paradox in these cases is that the affix or determiner must first be added for the full phonological structure to be calculated, but selection of the latter morpheme also depends on that same structure. For the cases in (7bi), however, a technique has been developed within OT that averts the paradox by postulating multiple inputs, like the set {a, an} for the English case (see McCarthy 2002, 183 for references). Outputs from either input, e.g. *a.n ap.ple*/ *a. ap.ple* are then evaluated in parallel. This enables allomorph selection to be done retroactively, as it were, still within the basic organization in (4).This technique is not applicable to the cases in (7bii), however, instantiated by English de-adjectival *-en*, whose behavior is illustrated in (8).

(8) SINGLE OBSTRUENT (Morphology):
$\text{-en}_V \Rightarrow$ / [+sonor][-sonorant] $]_A$ ___

a. quick-en, redd-en, fatt-en, rough-en, short-en, sharp-en

b. *slow-en, *green-en,
c. *apt-en, *lax-en, *strict-en
d. moist-en, fast-en, soft-en, hast-en, chast-en

As shown, *-en*: [ṇ] (syllabic *n*) attaches to adjectives that end in a single obstruent. Yet in (8d), such condition is met only after the stem-final *t*'s are deleted, a deletion that is in turn due to the presence of the suffix [ṇ], evidently as a result of a constraint along the lines of (9).

(9) NO GEMINATES (Phonology): *[-son. +cont.] Cx Cx ; x: [+cor. -cont]
(A sequence of coronal stops is excluded after a fricative)

The case in (8) is thus outright paradoxical for the model in (4), manifesting a mutual dependency between morphology and phonology just like the phonology-internal one of section 2. Here, affixation of *-en* to *sof(t)* depends on presence of a single obstruent (morphology); but that in turn depends on the presence of *-en* (phonology). The phonological character of the [soft]/ [sof] allomorphy precludes the multiple input approach of (7bi) which does not relate members of the input set to one-another, and is thus only suited to *suppletive* allomorphy. In contrast, the model in (5) yields the straightforward analysis in (10) below.

(10)	/soft-ṇ/	NO GEMINATES (Phonology)	MAX	SINGLE OBSTRUENT (Morphology)
a.	soft-ṇ	*		*
b.	☞ sof- ṇ		*	

Mutual dependencies of this sort exist also with Armenian plural allomorphs *-er/ -ner*, and Czech Prepositional Case allomorphs *-ích, -ech* . See Burzio (2005b) for references and discussion.

Unlike serial approaches, which can handle only one of the subcases in (7) given any particular enrichment (cyclicity or Multiple Inputs), the parallel framework deals successfully with all as would be easy to show, aside from the case in (7bi) requiring some discussion of suppletion, to which I return below.

5. Lexically controlled Phonology

The output of the phonology is sometimes variable given a certain type of input, and the variation is typically under lexical control, each lexical item exhibiting a consistent form. English shows variation in each of the cases in (11)-(13).

(11) **Word stress** (Burzio 1994, 67-75)
a. cárpenter, cýlinder, áncestor, mínister, cálendar, sínister, báluster, ...
b. seméster, decémber, philánder, disáster, piláster, evángel, exámple, ...

(12) **Vowel reduction in syllables closed by sonorants** (Burzio to appear)

a. Agamémnon, márathon támpon, péon, sámpan, cáravan, méteor, ...
b. Wiscónsin, ápron, cústom, bósom, phántom, ránsom, bácon, uténsil, ...

(13) **Penultimate vowel shortening** (Burzio 1994, 10.3, 2000)
a. blásphemous, ádmirable, áspirant, résident, lénitive, sátirist, bícycle, ...
b. desí:rous, oppó:sable, excí:tant, antecé:dent, divís:ive, escá:pist, ...

Such lexically controlled variation is potentially troublesome for the model in (4), where the lexicon is upstream of the phonology, with no obvious way to control its output. With morphologically underived items such as those in (11) and (12), however, that model turns out to be perfectly adequate, by way of the simple ranking schema in (14).

(14) **Lexical control of variation:**

/b/	C1	C2	IO-FAITH
a		*	*
☞ b	*		

In (14), the two top-ranked constraints are unranked relative to one-another, thus allowing for either output (a) or (b), leaving it to lower-ranked IO-FAITH to decide –the lexical control. This simple solution is applicable to (11) as shown in (15), as well as to the vowel reduction cases in (12) (see Burzio to appear).

(15)

/(cárpen)ter/	ALIGN FOOT R	ALIGN FOOT L	IO-FAITH
a. car(pénter)		*	*
b. ☞ (cárpen)ter	*		

This analysis follows Burzio (1994) and employs the notion that certain types of syllables can remain extrametrical at the right edge. In words such as those in (11), the left and the right edges of the word compete for alignment with foot boundaries. The respective alignment constraints instantiate C1, C2 of (14), with the input then acting as the arbitrator. Inputting (a) would give the alternative outcome, indeed as in *se(méster)*.

In this analysis, the item *(cárpen)ter* of (15) has its stress on the first syllable for the same reason that it begins with the sound [k], namely because this information is part of the input. Unlike in rule-based systems, placing stress information in the input this way is not tantamount to treating stress as just lexical or irregular, however. Rather, OT's Richness of the Base allows any information to appear in the input, hence also stress. Exactly which part of that input will survive in the output will then depend on the constraint ranking. In the regime of (14), (15), where IO-FAITH is fully dominated, input information will have no effect except in one case, namely when the dominant constraints tie over a conflict, as in the case at hand. While the input is thus 'rich' or free in general, the present approach adopts in full Prince and Smolensky's notion of Lexicon Optimization, according to which the stored lexicon corresponds to calculated outputs. Hence, the form *(cárpen)ter* will be the steady-state input-output. In this conception, the grammar must be thought of as performing a 'check' on such forms, rather than literally constructing them. When presented

with /(cárpen)ter/ as an input, the grammar will produce it unchanged as an output, thus satisfying the check. Had the input been /carpen(tér)/, the grammar would have produced a different output, maybe [car(pénter)], in which case that would have been stored as the steady-state input-output. Note that there is no circularity in going from input to output and back, since an input is taken to equal the output only in the steady-state, lexical storage, sense of input, hence only after the grammatical calculation, not a-priori.

The ability to provide such a straightforward account is one of the virtues of Optimality Theory and its parallel architecture. Unlike constraint hierarchies, rule systems can only either override or accept input information (by being either feature changing or feature filling, respectively). They are unable to accept such information just in case it is 'close enough', i.e. within the tolerance limits of the grammar in the way of (15). With rules, one could thus not place stress in the input without incorrectly making English stress fully irregular and unpredictable. Rather, one would have to tag individual morphemes with some grammatical information such as '*carpenter*: final syllable must be extrametrical', or '*Wisconsin*: undergoes vowel reduction in syllables closed by sonorants'. While some of these diacritic marks may be more problematic than others, all constitute a complication and a duplication of grammatical resources in the lexicon. The input information in (15) crucially differs from such diacritic marks by not containing any grammatical vocabulary, like 'Rule 22' or 'Ranking C1 >> C2', but rather only vocabulary out of which input-output representations are naturally constructed (features, syllables, feet, stresses, etc.).

The model in (5) is equally successful in dealing with (15). The reason is that, with morphologically simple cases, the two models in (4) and (5) are identical. With no morphological relations to deal with, the third set of constraints in (5) 'OO-FAITH (Morphology)' is moot, and with no morphological work being done upstream of the phonology in (4), the Underlying Representation need not be at all different from the Surface Representation given Prince and Smolensky's 'Lexicon Optimization'. It is only when morphology is introduced into the picture that the two models diverge. Then, only the one in (5) maintains the critical virtue of (15), as shown in (16).

(16)	/blásphem-ous/	*V:	OO-FAITH (stress)	IO-FAITH
a.	blasphé:m-ous	*		*
b.	☞ blásphem-ous		*	

In (16), the surface form *blásphem-ous* with a short vowel is both an input and the optimal output for that input, given the grammar shown. Had the vowel been long in the input, it would have been similarly reflected in the output, indeed as in *desí:rous*. This account, which updates that of (2) above, depends on the ability of input information to attach to morphologically complex forms independent of their bases. This ability is present in the fully parallel model in (5), but absent in the serial one in (4). By relying on the traditional 'Underlying Representation', the latter is committed to a single input form /blæsfe:m.../ for both *blasphemous* and its base *blasphé:me*. Presumably, this form would have a long vowel, but then this would be just like the input form /desi:r.../ of *desí:rous*, barring any account of the difference. By espousing serialism with

morphologically complex items, the model in (4) replicates the problems of rule-based systems that require diacritically marked morphemes. This kind of approach has in fact been pursued in otherwise insightful work by Anttila (2002). In Anttila's proposal, the ranking indeterminacy of (14) constitutes a 'Partial Order', defined as a disjunction of total orders or rankings, here the orders 'C1 >> C2'; 'C2 >> C1'. In the English case at hand, the morpheme /blæsfeːm.../ would be tagged with the order '* Vː >> OO-FAITH (stress)', while the morpheme /desiːr.../ would be tagged with the opposite order (see also Pater 2000 for a slightly different but equally problematic approach). Beside constituting a complication and a duplication of resources (grammar restated in the lexicon), such tagging does not appear empirically tenable, as control of the variation is only reliably attributable to words, not to morphemes, as the data in (17) show.

(17) ***-ative* shortening**

a. Short: appreciative, associative, collaborative, commemorative, cooperative, alterative, applicative, imaginative, ...

b. Long: accomodative, accumulative, agglutinative, aggregative, imitative, authoritative, connotative, multiplicative,...

The variation in (17) (once attributed to a special rule of '*-ative* shortening' by Halle and Vergnaud 1987) is of course just a subcase of that in (13) as the affected vowel is penultimate. The fact that the stress in some cases retracts two syllables away instead of just one is attributable to faithfulness to the base, e.g. *collábor**a**ːte*, and is to be analyzed as in *col(lábora)tive*, where the final syllable is extrametrical like that of *(ádjec)tive* or *(cárpen)ter* (Burzio 1994, 296). In these cases, attaching diacritics to morphemes will not help, since the variation affects just the same morpheme /aːt/ in all cases. Only an analysis that can associate idiosyncratic information to full words as in (16) will therefore correctly extend to these cases. For other cases comparable to (17) in Spanish , Zahao, and Tagalog, leading to the same conclusion, see Burzio (2005b) and references.

In sum, both phonological control of the morphology and lexical control of the phonology instantiate mutual dependencies across modules. In the first case, a phonological input: /soft-en/ depends on the morphology, but at the same time the morphological choice also depends on the phonological output: [sofn̩]. In the second case, some output: [blæsfəməs] depends on the grammar (shortening), but the exact grammar (shortening or not) also depends on the output itself. Such mutual dependencies yield exact replicas of one of the core arguments for parallel OT in phonology, pointing to a parallel organization of the components.

6. Dual Routes and Paradoxes of Strict Ranking

My next goal is to argue that Lexicon-Morphology and Phonology interact more closely than strictly-ranked constraints can express. I begin by noting that, even under strict-ranking of constraints, however, the model in (5) is more successful than alternatives in dealing with the important correlation described in (18).

(18)	Phonological regularity	Morphological regularity
a. L1	*parént-al* (Reg. stress) *natur-al* (Reg. V-shortening)	Irreg.: *compUlS-ive*; *ARBORE-al, crimIN-al*, ...
b. L2	*éffort-less* (Irreg. stress) *cri:me-less* (No V-shortening)	Reg.: *compell-ed, tree-less*, *cri:me-less*, ...

This table illustrates the fact, more carefully documented in Burzio (2002a), that one sector of the derived lexicon is simultaneously both morphologically irregular and phonologically regular, while the remainder of the derived lexicon is its mirror image: morphologically regular but phonologically irregular. In (18), the shading highlights regularity, phonological or morphological, and their complementary distributions. The lexicon produced by affixes that are roughly the etymologically Latinate ones, or 'Level 1' (L1) in the Lexical Phonology of Kiparsky (1982), exhibits regular syllabification, substantially regular stress, regular vowel shortening as discussed above, as well as velar softening, as in *criti[k]/ criti[s]ism*, arguably also a natural phonological process. At the same time, that same sector is riddled with morphological (and semantic) idiosyncrasy, such as the portions given in capitals, which have no correspondents in the respective base words. The lexicon produced by the Germanic or 'Level 2' (L2) affixes differs in both of these respects simultaneously, being both highly regular morphologically, and irregular phonologically in that stress, syllabification and other properties of the base are not recomputed to take account of the affix. The model in (5) accounts for this correlation by way of the simple assumption that morphological/ OO-F(AITH) constraints such as the one in (6a) above (the left context of *-ive* must be identical to a verb) are relatively low-ranked for L1 affixes like *-ive*, *-al*, etc., but high-ranked for L2 affixes like *-ness*, *-less*, *-ing*, etc. The L1 behavior is then characterizable as in (19a, b).

(19) a. **Morphological irregularity**

Input: */compUlS-ive/* Base: */compel/*	IO-F	OO-F1
compell-ive	*	
☞ compuls-ive		*

b. **Phonological regularity**

Base: /na:ture/	* V:	OO-F1
☞ natur-al		*
na:tur-al	*	

The tableau in (19a) accounts for irregular or suppletive morphology by means of the same resources as were invoked above for lexically controlled phonology --the ability to associate input information with morphologically complex forms independently of their bases. In (19a) irregular *compulsive* is optimal because the irregularity is part of its input and IO-F is dominant over the source of morphological regularity, OO-F(AITH)1. Note that the ranking relation between IO-F(AITH) and OO-F1 in (19a) is different than I had proposed earlier, specifically in (16), but this is in fact the concern of this section, to be addressed

below. Aside from this, a low-ranked OO-F1 will be correctly expected to incur violations not only by IO-F as in (19a), but also by markedness constraints of the Phonology as in (19b). In turn, high-ranked OO-F2 for L2 affixes will predict that both effects in (19) could reverse simultaneously, indeed as is the case in (18b), thus accounting for the correlation tabulated in (18).

Note as well that since IO-F is higher-ranked than OO-F1, we expect that phonological effects that occur when OO-F1 is involved may be absent not only with OO-F2 as in (18b), but also when IO-F is involved, as with underived items. This is correct, as the latter fail to undergo either shortening: *di.nosaur*, or velar softening: *[k]ing*, *[k]itchen*, etc. Similarly, input specification for L1 items should also enable an override of the same phonological effects. This is also seemingly correct, as it gives a natural account of phonological exceptions to either shortening or velar softening, e.g. *obe.s-ity*, *anar[k]-ist*, and others. The table in (20) summarizes these results.

(20)

	OO-F2	IO-F	Markedness	OO-F1
a. Underived: di.nosaur, [k]ing		✔	*	
b. L1: n[æ]tur-al, criti[s]-ism			✔	*
L1: compuls-ive		✔/U..S/		*
L1: obe.s-ity		✔/..e.../	*	✔
c. L2: cri.me-less (*crimIN-less)	✔	*	*	

The correlation in (18) and the facts in (20) are intractable within the approach in (4). In that approach, morphology and phonology operate at different points in a sequential derivation, and no means are available to make the phonology regular or more active just in case the morphology exhibits suppletion or other idiosyncrasies. Consider specifically Kiparsky's (2000) 'Lexical Phonology and Morphology OT'. In that framework, as in its rule-based predecessor, the output of Level 1 morpho-phonology would be the input to Level 2, by iteration of the model in (4). In order to characterize their different phonological behaviors as in (18), Level 2 would have to be assigned a more restrictive grammar than Level 1. The most effective way to do this would be to rank IO-F2 higher than IO-F1, mirroring the difference between OO-F1 and OOF2 of the present approach. Morphological idiosyncrasies, however, must be part of some input by definition, and the high-ranked IO-F2 can surely not have the effect of suppressing such input idiosyncrasies (if anything, just the opposite). What will be needed instead is the requirement once stated for the rule-based Lexical Phonology that the output of Level 1/ input to Level 2 be pure lexical items, excluding bound stems and other idiosyncrasies (whence, e.g. *compell-ing*, but not **compuls-ing*, *cri.me-less*, not *crimin-less*, etc.). Note that constraining the input to Level 2 this way will just make IO-F2 a type of OO-F, exactly as in the present framework. In the context of (4), however, that restriction is stipulatory, and completely unrelated to what the phonology requires, which is just a higher rank for IO-FAITH.

The correlation in (18) is also intractable within the 'Dual Route' approach of Pinker (1991) and related work. That approach postulates a rule-based morphology reserved for regular phenomena like the English regular past

tense, and assigns irregular morphology to the lexicon. When applied to the facts in (18), this approach would in all presumption treat L2 formations as rule-based and L1 formations as lexicon. When the phonological facts are considered along with the purely morphological ones, however, this view proves sharply inadequate. First, there is no reason, if morphological regularity is due to 'rules', why phonological regularity should also not consist of rules, predicting coexistence of the two regularities on the same route rather than the exact complementarity observed in (18). Second, the view that the L1 formations are in fact not 'formed' but just lexicalized seems completely untenable. For one thing they do not behave like underived items, as we have seen in (20). L1 items shorten vowels (*divinity*), but underived ones do not (*di:nosaur*). Secondly, their stress patterns exhibit precise and predictable transfer effects from their bases, like that of *América*n/ *américanist* (cf. underived *antágonist*, not **ántagonist*), or *phenómenon*/ *phenòmenólogy* (cf. underived *àbracadábra*, not **abràcadábra*). See Burzio (1994) for the analysis, and Kim (2006) for a recent review of this issue.

While superior to alternatives, the model developed so far nonetheless suffers from its own limitations. These stem from OT's strict ranking of constraints, which must therefore be superseded. The latter impose an undesirable categorical bifurcation of effects which is not unlike that of the dual route model. I argued in regard to (19a) that morphological irregularity calls for the ranking 'IO-F >>OO-F1'. However, the discussion of stress transfers in connection with both (2) above and the critique of Dual Routes in this section entails the opposite ranking, for the following reasons. Like all phonological regularity, the relatively regular stress of English entails the OT ranking 'Markedness >> IO-F', specifically as in (21a) below, where {STRESS} is a set of constraints defining the stress grammar. Stress transfers, however, entail by definition the ranking (21b), where STRESS_i is some member of {STRESS}. That is why *phenómenon* is irregular compared with *àbracadábra* (STRESS_i is here in fact just ALIGN FOOT L of (15) above). By transitivity, however, OO-F1 (stress) will now dominate IO-F as in (21c), contradicting the ranking of (19a).

(21) a. {STRESS} >> IO-F (Regular stress)
b. OO-F1(STRESS) >> STRESS_i (Stress transfers)
c. OO-F1(STRESS) >> IO-F (Transitivity)

One might suggest that the ranking in (21c) concerns stress, while that of (19a) concerns segmental structure, and I will argue below that this is partly correct. However, the ranking in (19a) is in fact paradoxical in its own right. The reason is that strictly dominant IO-F will predict no role for OO-F1, thus treating the segmental parallelism of pairs like *parent/ parental*, *titan/ titanic*, etc. as synchronically accidental. On the other hand, the opposite ranking will leave no room for idiosyncrasy. This, of course, is just the 'dual route' syndrome.

Furthermore, paradoxical ranking is in fact found not only between FAITH constraints as in (21c) versus (19a), but also between FAITH and Markedness constrains, as shown in (22), raising another red flag.

(22) a. rémedy/ remę̇ːdiable *CiV*-LENGTH >> OO-F (V-length)
b. lévy/ léviable OO-F (V-length) >> *CiV*-LENGTH

The adjective in (22a) exhibits the regular lengthening of a vowel in the context '__ CiV', as in *Canaːdian*, *Mongoːlia*, etc., while the one in (22b) maintains the short vowel of its base *levy* (as in 'levy a tax'), the two cases thus suggesting the contradictory rankings shown. A solution to this emerges once we consider that there is an independent difference between the two cases. The adjective in (22a) has a different stress than its base, unlike the one in (22b) --a difference in turn attributable to the maximal foot size: *(lévia)ble* vs. **(rémedia)ble* (Burzio 1994). This suggests that the effects that the OT notion of FAITH aims to capture in fact consist of attraction –the closer or more similar the representations are, the stronger the effect, as if the rank of FAITH had been raised, just as in (22b).

The task at hand is therefore to devise an approach capable of yielding the attraction needed for (22) (and many similar cases, reviewed in Burzio 2002a). This implies the ability to gage overall similarity of representations. We will see that a framework that attains these goals also resolves the other paradoxes of strict ranking.

7. Enter Distributed Representations

In Burzio (2002a, b), (2005a), I propose the hypothesis in (23).

(23) **Representational Entailments Hypothesis** (REH): Mental representations of linguistic expressions are sets of entailments. E.g. a representation consisting of A and B corresponds to the entailments: A⇒B, B⇒A (if A then B; if B then A).

This hypothesis yields the attraction effect needed, in the way shown in (24).

(24)

		A	B	C	R1's entailments violated by -C
a. R1:					
b. R2:	A	B	-C	A ⇒ C; B ⇒ C
c. R2':	A	-B	-C	A ⇒ C; ~~B ⇒ C~~

Under the REH (23), a representation R1 consisting of A, B, C, as in (24a) will yield a certain number of entailments that can affect other representations. We consider in particular that occurrence of -C ('minus C', or 'not C') in a representation R2 which is otherwise identical to R1 will violate two of R1's entailments as shown in (24b). However, the same -C in a representation R2' which is more distant or dissimilar from R1 will violate only one entailment as shown in (24c). In particular, the entailment B⇒C violated earlier is now satisfied since R2' no longer instantiates B. In general, the number of entailments violated by any one difference between R2 and some R1 (such as C/ -C) will equal the number of components that R1 and R2 have in common –the shaded cells in (24). Hence the pressure for full identity in space (attraction) eases as overall distance increases.

A crude application of this model to (22) would consist of taking B and C of (24) to stand for the stress and the vowel length of each base in (22), each base playing the role of R1of (24a). When the derivative bears the same stress as the base (B), it will instantiate R2 of (24b), meeting maximal pressure against change of vowel length (-C). When the derivative is re-stressed (-B), however, the pressure against changing vowel length (-C) will be lessened.

The stress transfers obtaining with L1 formations can also be reconciled in the attraction model. Such transfers obtain under substantial segmental identity, making a relatively strong attraction (= high ranked OO-F1 of (21b, c)) predictable. The dependency of stress transfer on segmental identity or similarity is in fact independently established by such contrasts as *lárynx-es*/ *larýng-es* (although *-es* is technically a L2 affix), showing that the stress of *lárynx* is kept only under segmental identity.

Summation of 'entailments' thus replaces ranking of (at least FAITH) constraints in OT in this conception. The resulting ranking is not strict, because it depends on the specific circumstances. The entailments are atomic types of constraints taken to be ontologically inherent to mental representations rather than part of some extrinsic grammar, as in (23). We note that this approach effectively postulates that mental representations are distributed like those in a neural net. The reason is that, in order for an entailment A⇒B due to a representation R1 to be at all relevant to some other representation R2 as in (24), it must be the case that components A and B of R1 are the same as those of R2, namely that the two representations are 'distributed' over the same units.

We must note as well that the 'attraction' characterization of OO-F relations just deployed for (22) does not supersede the role of selectional constraints like '-ive ⇒ /V __' of (6). Rather, the latter are taken to be entailments as well, the overall OO-F effect then resulting from the additive combination of the two (selection + attraction). This assumption maintains the needed difference in (22) (different attraction) despite the same affix *-able* (same selection), while also yielding different ranks across L1 and L2 affixes in general due to the more aggressive selectional properties of the latter, giving the correlation in (18) as discussed. Entailments like '-ive ⇒ /V __' of (6) are in fact themselves the result of summation in the present approach. Consider that individual surface representations will yield entailments like those of (25).

(25) -ive ⇒ /prevent __; -ive ⇒ /compuls __; -ive ⇒ /gener[ə]t __

Each of these entailments abbreviates a cluster of entailments, from each property of the affix, to each property of the stem. When such entailments are all summed up together, the ones that refer to properties that all the stems share will gain strength, while discordant entailments will remain as background noise. What all the stems in (25) have in common is that they each correspond to a verb. Consider specifically that, beside *preventive*, *compulsive*, *generative*, the mental lexicon also contains the verbs *prevent*, *compel*, *gener[ey]te*. Each of these will yield an entailment like '[ʤenereyt] ⇒ V' (the sound structure [ʤenereyt] entails the category V). When this type of entailment is combined with that of (25), the selectional effect of (6a) above will result by transitivity, in the way described in (26).

(26) a. -ive ⇒ /[dʒenerəːt] __ (from adjective)
b. [dʒenereyt] ⇒ V (from verb)
c. -ive ⇒ / V __ (by transitivity)

The entailment in (26c) is invariant across stems, and thus gains strength by summation, resulting in a general constraint expressing the selectional properties of the affix as introduced in (6) above. The reason the transitivity effect in (26c) would obtain despite the lack of identity in [generəːt]/ [genereyt] is that each of these representations is a cluster of properties and that the two clusters intersect, such that the entailment conjunction *-ive* ⇒ *x* and *x* ⇒ V will be true of many *x*'s. At the same time, however, such conjunction will be false for some *x*'s given the noted discrepancies. One can therefore reason that emergence of categorical selection with L1 affixes as in (26c) is in fact 'stunted' not only in the sense that summation over instances of (26c) is modest/ low-ranked, but also in the sense that selection is effectively intermediate between the categorical character of (26c) and the list in (25). This sheds light on another important property of L1 formations, their limited productivity, illustrated in (27).

(27) impress-ive, direct-ive, abus-ive/ *confess-ive, *suspect-ive, *excusive, ...
parent-al, peripher-al, trib-al/ *student-al, *surger(i)-al, *brib-al, ...

This behavior indicates that L1 affixes, while selecting fairly homogeneously for the category of their stems, still select from a list of stems rather than from a pure category. The ungrammatical stems in (27) are excluded by simply not being in the list. Selectional statements for L2 affixes, on the other hand, are expected to be more sharply categorical given the closer word-stem matches, yielding the notoriously greater productivity of those affixes. While this discussion makes no pretense of offering proof that the circumstances described produce hybrid dependencies intermediate between a category and a list, this behavior is well attested of neural net models of morphology (McClelland and Patterson 2002), no doubt due to their use of distributed rather than symbolic representations, a property shared by the present approach.

Finally, just as it account for modulations of the rank of OO-F, the REH (23) can also yield modulations of IO-F. Recall that irregularities like *compuls-ive* or *arbore-al* require IO-F to trump OO-F1, but not in such a wholesale fashion as to exclude the many regularities like *prevent-ive* or *parent-al*. The desired property would now ensue from the following -arguably natural- interpretation of Prince and Smolensky's (1993, 2004) 'Richness of the Base'. We take the input (='base') to be unrestrained ('rich') in a way that includes dimensionality or size of representations. Consider that in a representation /A, B, C, D/, there will be three entailments of type x⇒A. Such entailments will instantiate the OT notion of IO-F in the present approach (attraction to an input). In a representation /A, B/, however, there will be only one such entailment, thus effectively tying the rank of IO-F to dimensionality. For further discussion, and evidence that IO-F responds to dimensionality of representations, see Burzio and Tantalou (in press). Assuming random distributions of inputs over dimensionality will now resolve our last residual paradox along the lines of (28).

(28)		IO-F @ P	OO-F1	IO-F @ (1-P)
a.	compuls-ive	✔	*	
b.	prevent-ive		✔	*

In the proposed conception, IO-F may well have the high rank needed for suppletion with L1 items as in *compulsive*, but only with some probability P, tied to the probability of a sufficiently high-dimensional input. It will have a lower rank with the rest of the probability distribution (1-P), thus accounting for the presence of many regular items like *preventive*. This approach thus features a stochastic component in the constraint ranking as in Boersma (1998) and related work. However, it is unique in deriving the latter from the notion of 'Richness of the Base', via the link between dimensionality and overall entailment structure.

Note that, as in earlier discussion, two notions of input need to be distinguished here. On the one hand, the present approach combines Prince and Smolensky's Lexicon Optimization with the rejection of the underlying representations of (4) to reach the conclusion that inputs as stored forms are calculated outputs. On the other, however, this notion of input cannot be the one that distinguishes irregular *compulsive* from regular *preventive*, as those two representations will seem comparable in dimensionality. Rather, the relevant notion of input must refer to information, like the above /A, B, C, D/ versus /A, B/, as it *enters* (rather than exits) the calculation, in potential conflict with other information also entering the calculation, like the required faithfulness to *compel*, or *prevent*. The present proposal is thus that Richness of the Base should include freedom in the number of components that the pre-calculation input can recruit, representing the amount of memory that can be allocated to word-specific idiosyncrasy. Note in fact that it is only to this notion of pre-calculation or transitory input that the notion of Richness of the Base applies. The latter cannot apply to stored forms that are outputs, since by definition of calculation, outputs are not 'rich' or free.

In sum, while the parallel architecture in (5) is necessary, it is not sufficient. Attraction and other gradient effects require that lexicon and grammar -at least grammar qua morphology- also employ the same primitives so as to permit summation effects across components. This is because each of *arboreal*, *compulsive*, *parental*, *preventive*, etc. are simultaneously both lexical in certain ways and grammatical in others. It is also because FAITH constraints, purported to characterize the contribution of the Lexicon, have in fact ranks that depend both on distance from other representations and on dimensionality, as if lexical items had internal grammars.

8. Conclusion

The truism that our knowledge of language must consist of both a lexicon and a grammar has been arbitrarily construed in the generative tradition to mean that there are discrete modules organized in a sequence. The parallel approach of Optimality Theory has brought in a needed challenge to the traditional view, but it has also kept much of it intact. I have argued based on evidence from English

that parallelism does not stop at the boundary between phonology and morphology, and that lexicon and grammar consist of alternative uses of the same resources, and for this reason can yield phenomena that are not pure instantiations of either.

References

Anttila, Arto (2002) 'Morphologically Conditioned Phonological Alternations' *Natural Language and Linguistic Theory. 20.1*: 1-42

Boersma, Paul (1998) *Functional Phonology*, Holland Academic Graphics, The Hague.

Burzio, Luigi (1991)'Principles in Phonology' paper presented at the LSA winter meeting, Chicago (1/4/91).

Burzio, Luigi (1994) Principles of English Stress, Cambridge University Press.

Burzio, Luigi (2000) 'Cycles, Non-Derived-Environment Blocking, and Correspondence,' in Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer, eds. *Optimality Theory: Phonology, Syntax, and Acquisition.* Oxford University Press, 47-87.

Burzio, Luigi (2002a) 'Surface-to-Surface Morphology: when your Representations turn into Constraints' in P. Boucher (ed.) *Many Morphologies*, Cascadilla Press. 142-177.

Burzio, Luigi (2002b) 'Missing Players: Phonology and the Past-tense Debate,' *Lingua 112*, 157-199.

Burzio, Luigi (2005a) 'Sources of Paradigm Uniformity', in Laura J. Downing, T. A. Hall, Renate Raffelsiefen, eds.*,* 65-106.

Burzio, Luigi (2005b) 'Lexicon and Grammar: unequal but inseparable', ms. JHU. http://www.cog.jhu.edu/faculty/burzio/LexGramPaper05.pdf

Burzio, Luigi (to appear) 'Phonology and Phonetics of English Stress and Vowel Reduction' *Language Sciences.*

Burzio, Luigi and Niki Tantalou (in press) 'Modern Greek Accent and Faithfulness Constraints in OT', *Lingua.*

Downing, Laura J., T. A. Hall, Renate Raffelsiefen, eds. (2005). *Paradigms in Phonological Theory*. Oxford: Oxford University Press.

Halle, Morris, and Jean Roger Vergnaud (1987) *An Essay on Stress*. MIT Press.

Kim, Young-Seok. 2006. English Stress: Rules or Constraints? *Korean Journal of Linguistics* 31, 1-26.

Kiparsky, Paul (1982) 'Lexical Phonology and Morphology,' in I.S. Yang ed. *Linguistics in the Morning Calm*, 3-91. Seoul: Hanshin.

Kiparsky, Paul (2000) 'Opacity and Cyclicity', *The Linguistic Review* 17, 2-4, 351--365.

McCarthy, John (2002) *A Thematic Guide to Optimality Theory*, New York: Cambridge University Press.

McClelland, James, and Karalyn Patterson (2002) "'Words *or* Rules' cannot exploit the regularity in exceptions' *Trends in Cognitive Science 6.11*, 464-465.

Pater, Joe (2000) 'Nonuniformity in English stress: the role of ranked and lexically specific constraints'. *Phonology* 17:2, 237-274.

Pinker, Steven (1991) 'Rules of Language,' *Science, 253* (5019), 530-535.

Prince, Alan, and Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Blackwell. (2004).

Contact Information:

Luigi Burzio, Department of Cognitive Science, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218. USA

burzio@jhu.edu