2

The architecture of grammar and the division of labor in exponence

RICARDO BERMÚDEZ-OTERO

2.1 Introduction[1]

Languages commonly exhibit alternations governed by complex combinations of phonological, morphological, and lexical factors. An alternation of this sort will often admit a wide variety of analyses, each apportioning different roles to lexical storage and to morphological and phonological computation. Such analytic underdetermination poses a threat to falsifiability and to learnability: hypotheses can easily evade empirical disconfirmation if potential counterexamples can be redescribed in many different ways to suit the linguist's convenience, and so theories risk losing empirical content; by the same token, it becomes hard to explain how, among a plethora of choices, learners converge upon the target grammar (§2.2). To avert these dangers, the theory of grammar must set limits to the space of possible interactions between phonology, morphology, and the lexicon: in particular, it must ascertain the proper division of labor between storage and computation (§2.3), and it must constrain the ways in which morphological operations can manipulate phonological material and in which phonological processes can refer to morphosyntactic information (§2.4).

Concerning the question of storage vs. computation, this chapter pursues the hypothesis that different types of alternation reflect different modes of interaction between the lexicon and the grammar. This idea is fleshed out by means of a refined dual-route approach to exponence (§2.3.1, §2.3.5), in which the well-established distinction between explicit symbolic generalization and implicit pattern association (§2.3.1, §2.3.4) is supplemented with a further distinction between two types of lexical listing, analytic and nonanalytic (§2.3.1, §2.3.3.1), akin to Clahsen and Neubauer's (2010: 2634) contrast between 'combinatorial entries' and 'unanalyzed

[1] I am specially indebted to Sarah Collie, Jesse Saba Kirchner, Andrew Koontz-Garboden, Tobias Scheer, and Jochen Trommer for inspiration in the writing of this chapter: their influence will be readily apparent to all readers familiar with their work. Section 2.3.3 draws on research previously presented at meetings in Manchester, Leipzig, Groningen, and Warsaw: I am grateful to the audiences on all these occasions for their comments and suggestions. The chapter has benefited from careful scrutiny by Patrik Bye, Tobias Scheer, and Jochen Trommer; all errors and infelicities remain my own.

entries' (see also Stemberger and MacWhinney 1986, 1988). Assuming a stratal version of Optimality Theory (OT), I show that the peculiar syndrome of properties characteristic of stem-level morphophonology arises from the fact that stem-level forms are stored nonanalytically but stem-level processes are nonetheless explicitly represented in the grammar by means of symbolic generalizations, whose status resembles that of Jackendoff's (1975) lexical redundancy rules (§2.3.3.1). The model provides a highly explanatory account of internal cyclic effects in stem-level domains, which I illustrate with classic examples such as English *órigin* ~ *oríginal* ~ *orìginálity* (§2.3.3.2) and *còmp*[ə]*nsátion* vs. *cònd*[ɛ̀]*nsátion* (§2.3.3.3).

On the issue of the interaction between morphology and phonology, this chapter argues for the adoption of a restrictive stance based on general cognitive principles of modularity and locality (§2.4.1). A program is proposed consisting of four hypotheses (37): that morphology selects and concatenates morphs without ever altering their phonological content (§2.4.2); that phonological constraints other than those on prosodic alignment may not refer to morphosyntactic information (§2.4.3); that output phonological representations do not contain diacritics of morphosyntactic affiliation (§2.4.4); and that morphosyntactic conditioning in phonology is subject to cyclic locality (§2.4.4). These hypotheses will provide the guiding thread for an evaluation of several mechanisms currently used to describe morphologically conditioned phonological processes, including construction-specific cophonologies (§2.4.2.3), indexed constraints (§2.4.3), and readjustment rules (§2.4.3). The balance of argument supports a stratal-cyclic architecture for phonology—one, however, in which neither cyclicity nor stratification are innately stipulated, but both emerge from fundamental storage and processing mechanisms (§2.3.3.2, §2.3.3.3) and from timing effects in the child's linguistic development (§2.4.2.3).

2.2 Analytic underdetermination

A morphologically conditioned phonological alternation of no more than ordinary complexity will often be compatible with several sharply different grammatical descriptions, all of which may succeed in covering at least the central facts. As a representative example, this section examines a relatively straightforward alternation affecting second-conjugation theme vowels in Spanish deverbal derivation. We shall see that, although the phenomena look simple, current linguistic theory offers a surprisingly wide range of alternative analyses. The existence of such a large space of conceivable grammars forces us to ask what analytic biases guide the learner's choice of generalizations.

Spanish verbs are arbitrarily divided into three inflectional classes or 'conjugations', each associated with its own theme vowel: [-a-] for the first conjugation, [-e-] for the second, and [-i-] for the third; see (1a). In nouns and adjectives derived from verb stems, the theme vowel of the base undergoes deletion before vowel-initial suffixes, but remains if the following suffix begins with a consonant (Rainer 1993: 95, 96, Pena 1999: 4337, Bermúdez-Otero 2007b). For our purposes, the interesting fact is that second-conjugation verb stems surface with [-i-], instead of the expected [-e-], before certain suffixes like [-β̞l-e], cognate with English *-able*/*-ible*, and [-ˈmjen̪t̪-o], cognate with English-*ment* (Rainer 1993: 95, 1999: 4609–10). This substitution of [-i-] for [-e-] also occurs in certain inflectional forms, such as the past participle.[2]

(1)

	1st-conjugation base		*2nd-conjugation base*		*3rd-conjugation base*	
a.	aᶞ̞miˈɾ-a-ɾ	‘admire.INF’	beˈβ̞-e-ɾ	‘drink.INF’	suˈfɾ-i-ɾ	‘suffer’
b.	aᶞ̞miɾ-a-ˈð̞oɾ-Ø	‘admirer’	beβ̞-e-ˈð̞oɾ-Ø	‘drinker’	sufɾ-i-ˈð̞oɾ-Ø	‘sufferer’
c.	aᶞ̞miˈɾ-a-β̞l-e	‘admirable’	beˈβ̞-**i**-β̞l-e	‘drinkable’	suˈfɾ-i-β̞l-e	‘sufferable’
d.	aᶞ̞miˈ ɾ-a-ð̞-o	‘admire.PTCP’	beˈβ̞-**i**-ð̞-o	‘drink.PTCP’	suˈfɾ-i-ð̞-o	‘suffer.PTCP’

The single fact of the alternation between [-e-] in *bebedor* ‘drinker’ and [-i-] in *bebible* ‘drinkable’ could conceivably be described in many different ways, surveyed in the following paragraphs.

One might begin by assuming that, at a relatively deep level of representation, the stem of *beber* bears the theme vowel /-e-/ in both *bebedor* and *bebible:* i.e. ⟦$_{\mathrm{V}}$⟦$_{\surd}$beb⟧⟦$_{\mathrm{Th}}$e⟧⟧ or, in “inferential” (Stump 2001: 1) or “amorphous” (Anderson 1992) theories, just /bebe/.[3] This representation could be the result of an operation of morph insertion, possibly conditioned by a conjugation diacritic (2a), as in Distributed Morphology (Halle and Marantz 1993); or it could be generated by a realizational rule of stem formation, again possibly referring to inflectional class features (2b), as in frameworks like Amorphous Morphology (Anderson 1992) or Paradigm Function Morphology (Stump 2001); or it could simply be stored in the lexical entry of *beber* (2c), as in lexicalist models like Lieber’s (1980: ch. 1).

(2)
a. *Th* ↔ -e- / [II]⌒__ (see Embick 2010: 76)
b. <[V, class II], (*V* → *Ve*)> (see Aronoff 1994: 68)
c. BEBER ↔ ⟦⟦$_{\surd}$ beb⟧ ⟦$_{\mathrm{Th}}$ e⟧⟧ (see Bermúdez-Otero 2013)

On this assumption, the theme vowel of the verbal base of *bebible* must at some point undergo a mapping that alters its phonological content along the lines of (3)—or its equivalent in the reader’s preferred theory of distinctive features for vowels, adjusted according to the reader’s preferred assumptions concerning featural (under)specification.

(3) [-high, –low, –back] → [+high, –low, –back]

Different frameworks would allow this process to take place in different places in the grammar. In inferential theories, where all exponence is driven by realization rules, the transformation of the theme vowel could be included in the statement of a realization rule for BLE (4a); in this scenario, the environment for (3) is defined morphologically. Similarly, Distributed Morphology, though a lexical rather than an

[2] Spanish verbs exhibit extensive theme-vowel allomorphy under inflection, especially in the second and third conjugations. Roca (2010) provides comprehensive coverage of the inflectional facts, but does not address derivation.

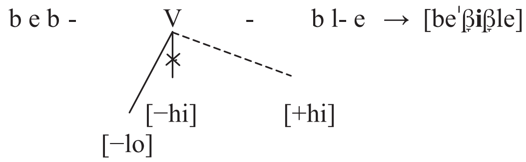
[3] I use hollow brackets to mark the edges of morphosyntactic constituents and of phonological cyclic domains. I reserve solid square brackets for phonological surface representations, feature matrices, and the edges of prosodic units.

inferential theory in Stump's (2001: 1) terms, countenances readjustment rules (Halle and Marantz 1993: 127–9, Embick and Halle 2005), which directly alter the phonological content of morphs in morphosyntactically defined contexts, as in (4b); see further §2.4.3 below.

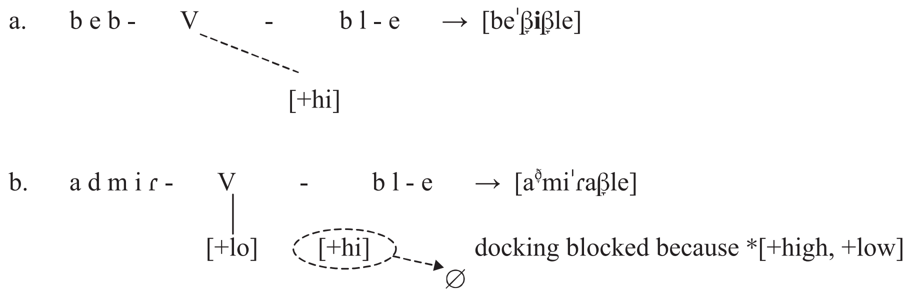
(4) a. <[V, BLE], (V*e* → V*ible*)>

b. /e/ → /i/ / ⟦*Th*,__⟧⌒⟦*a*,-bl-⟧

Alternatively, the effects of (3) could be achieved by phonological means. One possibility would be to muster the autosegmental machinery used to describe mutation in frameworks where exponence is strictly piece-based (e.g. Lieber 1992: 165–71, Stonham 1994: ch. 5, Wolf 2007, Hermans and van Oostendorp 2008, Trommer 2008c, 2011, Bye and Svenonius this volume, and see §2.4.2.2 below). For example, one could posit an initial floating [+high] feature in the underlying representation of BLE and MIENTO, docking onto the nearest vowel on the left.[4] Appropriate provisions in the phonology would prevent the floating autosegment from docking onto a low vowel, thus protecting first-conjugation stems; docking onto third-conjugation theme vowels would be implemented vacuously.

(5) 

In a slightly different version of this autosegmental analysis, the docking process in (5) could be recast as a feature-filling, rather than a feature-changing, operation by assuming underspecified underlying representations for verbal theme vowels in the style of Roca (2010: 413): [+low] in the first conjugation, a completely underspecified vowel in the second conjugation, and [+high] in the third.

(6) 

Yet another option would be to assume that the floating [+high] feature is packaged with the second-conjugation theme itself, rather than with BLE or MIENTO, so that these suffixes merely trigger an association rule that enables the floating feature to dock onto its sponsor: see Wiese's (1996) analysis of umlaut in present-day German.

[4] It is interesting to note that BLE has an allomorph [-β̞il-], with a high vowel, which occurs before other derivational suffixes: e.g. [sost̪eˈn-e-ɾ] 'sustain.INF' ∼ [sost̪eˈn-i-β̞l-e] 'sustainable' ∼ [sost̪en-i-β̞iˈl-ið̞a^ð̞-Ø] 'sustainability.' Independent evidence for the presence of the feature [+high] is however lacking in other environments where thematic [-e-] is replaced with [-i-]: see note 2.

In the various scenarios we have considered so far, the base of *bebible* is represented as /bebe/ at a relatively deep level in the grammar and is mapped onto /bebi/ either by a morphological realization rule or by a phonological process; but it is of course possible to imagine derivations of *bebible* in which /bebe/ never occurs as either an initial or an intermediate representation. For example, we could assume that, from the outset, the alternants /-e-/ and /-i-/ compete with each other in the morphology and that, in the case of *bebible*, /-i-/ wins, blocking /-e-/. As we have already seen, different theories of morphology will implement this competition in different ways, even if we agree that /-i-/ has the more specific distribution and /-e-/ is the default: piece-based theories in which lexical storage is restricted to roots and exponents of single functional heads, like mainstream Distributed Morphology (cf. Siddiqi 2009), will posit two underlying theme-vowel allomorphs, /-e-/ and /-i-/, competing for insertion in the *Th* position; theories allowing stem storage (Lieber 1980: ch. 1) might posit two listed stems for BEBER, /bebe-/ and /bebi-/, again competing for insertion; and process-based theories of morphology would set up two realization rules of second-conjugation stem formation, competing for application. Whether the contenders in this morphological competition are pieces (of whatever size) or rules, the contextual restrictions on /-i-/ could in principle be stated either morphologically or phonologically, for it is a well-established fact that allomorphs can arbitrarily subcategorize not only for morphological properties in their environment, but also for phonological ones (e.g. Kiparsky 1994: 19, Carstairs-McCarthy 1998b, Paster 2006, 2009, Bye 2007, Nevins 2011b, and see §2.4.2.1 below). In the case at hand, however, the option of imposing a phonological subcategorization frame on /-i-/ may seem a little far-fetched (though not impossible): borrowing the autosegmental machinery of (5), /-i-/ could be restricted to environments where a floating [+high] is present, but it would perhaps be unsatisfactory if this autosegment then passed inertly through the phonology on its way to Stray Erasure. In contrast, an analysis relying on morphologically defined subcategorization statements gives us yet another choice: whether to impose a contextual restriction on /-i-/ or rather on BLE and MIENTO.

The possibilities do not end there: notably, one could, if one wished to, find a way to prop up the suggestion that two second-conjugation theme-vowel allomorphs, /-e-/ and /-i-/, compete for selection and that, in the case of *bebible*, it is a floating [+high] feature affiliated with BLE that causes /-i-/ to win. This could be done simply by letting the choice of allomorph take place in the phonology, rather than in the morphology (e.g. Tranel 1996, 1998, Kager 1996, 2009: 420ff., Mascaró 1996, 2007, Rubach and Booij 2001, Bonet and Harbour this volume). In this scenario, the input to the phonology looks like the following:

(7) /beb-{**e,i**}-$^{[+\text{hi}]}$b1-e/

If a high-ranking featural faithfulness constraint indexed to stems, IDENT(stem)-[high], dominates the faithfulness constraint demanding the preservation of the floating autosegment affiliated with BLE, namely MAX(affix)-[high], then the floating feature will dock only when it can do so without changing the height of the theme vowel; cf. the ranking metaconstraint FAITH(root) $\gg$ FAITH(affix) postulated by McCarthy and Prince (1995a: 364). With a first-conjugation stem (8a), faithful docking is impossible. If the stem belongs to the second conjugation (8b), however, the floating feature of BLE can survive by coalescing with the input specification of the /-i-/ allomorph of the theme vowel. In the case of *bebedor* 'drinker,' we need to assume that the agentive suffix DOR does not carry a floating [+high] feature. This granted, /-e-/ could be made to win over /-i-/ in at least two different ways. Since [e] is the epenthetic vowel of Spanish, we can assume that it is the least marked under the phonological constraint hierarchy of the language; if so, the selection of /-e-/ in *bebedor* could be treated as an emergence-of-the-unmarked effect (McCarthy and Prince 1994a), as in (8c).

(8)

Input	Candidate	IDENT(stem)-[high]	MAX(affix)-[high]	*[+high]
(a) ⟦stem admiɾ-a ⟧ ⟦affix bl-e⟧ admiɾ-a: [−hi]1; affix: [+hi]2	admiɾible [+hi]2	*!		(*)*
	☞ admiɾable [−hi]1		*	(*)
(b) ⟦stem beb-{ e , i} ⟧ ⟦affix bl-e⟧ e: [−hi]1; i: [+hi]2; affix: [+hi]3	bebeble [−hi]1		*!	
	☞ bebible [+hi]2,3			*
(c) ⟦stem beb-{ e , i} ⟧ ⟦affix doɾ⟧ e: [−hi]1; i: [+hi]2	bebidoɾ [+hi]2			*!
	☞ bebedoɾ [−hi]1			

Alternatively, the theory of phonologically driven allomorph selection developed in Bonet (2004: 90ff.) and Bonet, Lloret, and Mascaró (2007) allows the input to the phonology to carry the stipulation that the /-e-/ allomorph is preferred to the /-i-/ allomorph: cf. (7) and (9a).

(9) a. /beb-{**e≻i**}-$^{[+hi]}$bl-e/

b. /beb-{**e≻i**}-doɾ/

In this approach, the right results follow from ranking Max(affix)-[high] above the constraint Priority, which favors obedience to the allomorphic preferences stipulated in the input.

This enumeration could easily be extended, but the point should by now be clear: morphologically conditioned phonological alternations pose a staggering problem of analytic underdetermination. I am of course not suggesting that all the analyses surveyed above are equally plausible, and I admit that, in at least a few cases, considering a wider range of Spanish data would in all likelihood tilt the scales in one direction or the other; but I shall not declare my own preferences here. For my current purposes, it is rather more significant to note that all these different descriptions of the same facts rely on ideas advocated by current schools of generative grammar and that many of the relevant devices are regarded by their proponents as mutually compatible. Indeed, looking at (1) from a nongenerativist viewpoint (e.g. connectionist or exemplar-based)[5] would open up yet another cornucopia of descriptions. The problem of underdetermination becomes further compounded if, as seems likely, several of the grammars we have sketched turn out to have exactly the same weak generative capacity (Chomsky 1963): presumably, each grammar will still have different consequences for acquisition and processing, and it will certainly instantiate a theory with its own typological implications, but a linguist will not be able to choose between these grammars by looking at the facts of Spanish alone.

Going one step further, we can take our own predicament as linguists as indicative of the immense obstacle that morphologically conditioned phonological alternations would pose for the learner if she had to contend with an equally vast and shapeless space of possible grammars.[6] In this sense, we have come up against an instance of the logical problem of language acquisition (Baker and McCarthy 1981), which is arguably nothing other than the problem of induction (Hume 1748[2000]: section 4, Popper 1935[1959]: ch. 1, §1) as faced by the language-learning child (Pinker 2004: 949, Bermúdez-Otero and McMahon 2006: 550). In this light it seems advisable to impose *a priori* biases on the space of possible hypotheses about morphology–phonology interactions: these are essential to guarantee learnability, for everyone agrees that language learning is impossible without some analytic bias (Lappin and Shieber 2007: 394–5); but, in addition, the postulation of analytic biases is one key source of typological predictions in linguistics, models of channel bias providing the other key source (Moreton 2008).

Taking a broad view of the problem, one should expect an adequate set of analytic biases to inform one's answers to two basic questions. First, in any instance of alternation, which aspects (if any) reflect lexical storage of allomorphs, and which (if any) involve the generation of positional variants by computational processes? Second, which computations take place in the morphological component of the grammar, and which are carried out in the phonology? In this chapter, section 2.3 addresses the first question; section 2.4, the second. In section 2.3, I survey the debate on

[5] For the distinction between connectionist and exemplar-based models, see Chandler (2010: 375–6).

[6] Shapelessness, rather than size, is the major problem: when the space of possible grammars grows too large for exhaustive search, learnability becomes dependent on the **structure** of the space, rather than on its size (Tesar and Smolensky 2000: 2–3).

the division of labor between storage and computation since *SPE* (Chomsky and Halle 1968). In section 2.4, I consider a set of hypotheses concerning the ability of morphology to manipulate phonological exponents and the ability of phonology to refer to morphosyntactic information. Throughout the discussion we shall see that linguists' choices of analytic biases are typically guided by very general programmatic assumptions about human cognition on key questions such as the nature of memory, the modularity of mind, and the locality of computation.

2.3 Storage vs. computation

2.3.1 Preview: modes of interaction between the lexicon and the grammar

We begin by considering the relative roles of the lexicon and of the grammar (particularly the phonology) in determining the form of linguistic exponents. If linguistics is a branch of cognitive science, then this inquiry need concern itself only with those patterns of exponence that show evidence of psychological reality. To beg as few questions as possible, however, our operational definition of cognitive reality should be appropriately liberal: we shall say, therefore, that a morphophonological pattern is psychologically real if speakers extend it to new items—even if only sporadically—whether in language change or in experimental tasks such as wug tests (Berko 1958). The application of such tests has reached an impressive level of refinement in recent years, enabling us to distinguish with growing precision between accidental patterns and linguistically significant generalizations (e.g. Becker, Ketrez, and Nevins 2011). This research has thrown up some challenging results, which suggest that native speakers can acquire knowledge of remarkably subtle and apparently unmotivated statistical trends (e.g. Hayes et al. 2009).

By our operational definition of psychological reality, then, linguistic theory must answer for a highly heterogeneous set of morphophonological phenomena differing very widely in productivity. The well-studied case of Modern English past-tense marking provides examples of generalizations lying far apart on this spectrum: notably, the alternation shown in (10a) sustains no exceptions (Albright and Hayes 2003: 151) and applies automatically to any new verb inflected by means of the suffix /-d/; in contrast, ablaut patterns like those in (10b) are seldom extended to new verbs, though their psychological reality cannot be denied precisely because novel strong forms do arise from time to time in diachronic change (e.g. *snuck*: see §2.3.2 below) and, with varying frequencies, in nonce-probe experiments (Bybee and Moder 1983, Prasada and Pinker 1993, Albright and Hayes 2003).

(10) a. /-d/:

/-d/:	[-ɪd] after /t, d/	[ˈpæt-ɪd, ˈæd-ɪd]
	[-t] after voiceless segments other than /t/	[tæp-t, pæk-t, kəˈlæps-t, pætʃ-t]
	[-d] elsewhere	[dæb-d, sæg-d, bæn-d]

b. dɹaɪv∼dɹəʊv, ɹaɪt∼ɹəʊt, ɹaɪz∼ɹəʊz
dɹɪŋk∼dɹæŋk, swɪm∼swæm, sɪt∼sæt
bləʊ∼bluː, gɹəʊ∼gɹuː
etc.

One is therefore led to ask: do such apparently disparate patterns call for essentially different accounts, each striking its own specific balance between the roles of the lexicon and of the grammar?

Chomsky and Halle (1968) answered this question largely in the negative: they asserted that any pattern of alternation exhibiting any degree of psychological reality must be expressed as a standard rule, i.e. as a rule that applies whenever its input satisfies its structural description. As we shall see in section 2.3.2, this aspect of *SPE*'s doctrine reflected Chomsky and Halle's adherence to an important strand of structuralist thought: Bloomfield's (1933: 274) view of the lexicon as a cognitively inert list (Chomsky 1965: ch. 2, endnote 16). This conception of the lexicon left grammatical rules, and in particular the type of nonstochastic rule available at the time, as the only device capable of expressing cognitively real patterns of alternation. In consequence, *SPE* and all the single-route theories of exponence inspired by it handle Modern English strong verbs by means of mandatory ablaut rules (Chomsky and Halle 1968: 11, Halle and Mohanan 1985: 107–14, Halle and Marantz 1993: 127–9); insofar as such rules operate upon phonological features but must refer to arbitrary diacritics or lexical lists, they are incompatible with a modular approach to the interfaces between syntax, morphology, and phonology, the subject of section 2.4 of this chapter (see §2.4.3 and §2.4.4.). In section 2.3.2 I further recall how *SPE*'s assumptions favored the opportunistic use of underlying specifications and of extrinsic rule ordering for the purposes of subsuming lexically idiosyncratic patterns of alternation under independently motivated phonological rules. This, in turn, forced Chomsky and Halle to renounce the goal of devising unsupervised learning algorithms for phonological rule systems.

In contrast with *SPE*, this chapter pursues the hypothesis that alternations of different types reflect different modes of interaction between the lexicon and the grammar. The specific proposal outlined below draws upon ideas from a variety of sources: most immediately,

(i) Jackendoff's (1975) theory of lexical redundancy rules,
(ii) the phonological architectures developed in Lexical Phonology (Kiparsky 1982a, b) and Stratal OT (Bermúdez-Otero 1999, Kiparsky 2000), and
(iii) the dual-route approach to exponence (Prasada and Pinker 1993, Clahsen 1999, Pinker 1999, Ullman 2001, Pinker and Ullman 2002).

Of course, each of these sources has a rich and complex history of its own: both (ii) and (iii), for example, may be regarded as pursuing insights first expressed in the taxonomies of alternation of the Kazan School (Kruszewski 1881[1995], Baudouin de Courtenay 1895[1972]). This section provides an overview of my proposal; more detailed arguments are laid out in sections 2.3.3, 2.3.4, and 2.3.5, where I have made a sustained effort to emphasize the intellectual lineage of each idea.

In line with dual-route frameworks, then, I reject single-mechanism connectionist and exemplar-based models of exponence (e.g. Rumelhart et al. 1986, Chandler 2010); I assert instead that **some** morphophonological patterns **are** explicitly represented in the grammar by means of rules in the broad sense, i.e. by means of symbolic generalizations containing one or more typed variables (Marcus 1998, 2001: ch. 3).[7] The criteria for regularity, in this specific technical sense, have been most fully worked out for inflection (Pinker 1999: 217–18, 222–4, etc, Pinker and Ullman 2002: 458–62). By these criteria, the grammar of present-day English is found to contain symbolic generalizations driving the realization of past tense by means of the suffix /-d/, with its concomitant alternation between [-d], [-t], and [-ɪd]: see (10a). In contrast, other patterns of exponence such as strong verb ablaut (10b) satisfy our operational definition of psychological reality but fail to pass the test of regularity: following the dual-route approach, I assume that such patterns are encoded implicitly in the connections between nodes in a distributed associative memory (although in §2.3.4 I note the important alternative offered by Albright and Hayes 2003).

Within the class of morphophonological patterns represented by symbolic generalizations, however, one can still observe wide differences in productivity and susceptibility to lexical exceptions. This fact suggests that dual-route theories of exponence require further elaboration: we need a 'refined' dual-route model (Clahsen, Sonnenstuhl, and Blevins 2003: 127, 149). As emphasized in the tradition of Lexical Phonology and Stratal OT, for example, present-day English exhibits a sharp contrast between 'class-two' or word-level constructions, such as /-d/ suffixation in the past tense, and 'class-one' or stem-level constructions, like adjectival derivation with the suffix /-ɪk/.[8] The alternation affecting the suffix /-d/ (10-a) is strictly exceptionless: notably, in the wug test conducted by Albright and Hayes (2003: 151) participants failed to produce the expected alternant [-d] in only one out of 937 responses to probes ending in voiced segments other than /d/ (see also note 24 below). In contrast, the rule whereby /-ɪk/ attracts primary stress to the preceding syllable (11a) does sustain lexical exceptions (11b): see Fournier (2010: 28) for an exhaustive list.

(11) a. acrobát-ic, genét-ic, harmón-ic, idýll-ic, Miltón-ic, titán-ic, etc.
 b. Cáthol-ic, Árab-ic

Ordinary dual-route models do not provide us with the means to account for this disparity between stem-level and word-level constructions. On the one hand, the generalizations governing the assignment of stress to English stem-level forms drive diachronic processes of regularization and apply productively to novel forms, including phonotactically deviant items: accordingly, they cannot involve mere pattern association, but must rather be explicitly represented in the grammar by means of symbolic generalizations (Hayes 1982: 236–7, and see §2.3.3.1 for further details). On

[7] This chapter's concerns are mainly architectural; in general, the implementation of symbolic generalizations by means of rewrite rules, on-or-off parameters, inviolable constraints, or violable constraints will not be at the heart of the discussion, though optimality-theoretic constraint interaction will play an important role in the execution of some analyses (e.g. §2.3.3.2, §2.4.2.2) and in the formalization of some proposals (§2.4.3). The term 'rule' is therefore to be understood in the broad sense of 'symbolic generalization', unless a narrower meaning is made clear by the context.

[8] Empirically, my distinction between stem-level and word-level constructions coincides roughly (but only roughly) with Siegel's (1974: 111ff.) distinction between class-one and class-two affixation, and with Booij and Rubach's (1987) distinction between cyclic and postcyclic affixation. Conceptually, it is rather different: see §2.4.2.3 below, specially (62) and the immediately preceding paragraph. Booij and Rubach's (1987) postlexical stratum corresponds to my phrase level.

the other hand, the absence of exceptions to the alternation shown in (10a) cannot mean that stored past-tense forms are unable to block regular /-d/ suffixation, for such blocking is essential to the survival of irregular strong forms.

The solution lies in acknowledging two types of lexical listing. English stem-level derivatives, I shall argue, are mandatorily entered into the lexicon in the form assigned to them by the stem-level morphology and phonology: i.e. they are listed not as strings of underliers, but as fully prosodified stem-level output structures. This type of listing, which following Kaye (1995: 302ff.) I call nonanalytic, corresponds approximately to Jackendoff's (1975: 643ff.) notion of 'whole-form storage'. In contrast, word-level constructs, including regular past-tense forms, may be listed or unlisted; but, if they happen to be listed, they are crucially entered into the lexicon analytically, i.e. as concatenations of word-level input pieces. Analytic listing thus resembles the concept of a 'combinatorial lexical entry', which Clahsen and Neubauer (2010: 2634) postulate on the basis of independent psycholinguistic findings (for further psycholinguistic support, see Stemberger and MacWhinney 1986, 1988). Because stem-level forms are listed nonanalytically, i.e. with their stem-level phonological properties fully specified, their lexical entries are able to block the on-line application not only of stem-level affixation rules, but also of the phonological processes (such as stress assignment) that target stem-level domains; the latter may consequently sustain lexical exceptions, as we see in (11b). In contrast, the fact that word-level past-tense forms are either unlisted or listed analytically entails that the surface realization of the word-level suffix /-d/ is always computed on line and so strictly abides by the pattern shown in (10a)—excepting only cases of processing error, as shown by Albright and Hayes (2003: 151).

In the refined dual-route architecture proposed in this section, therefore, the distinction between explicit symbolic generalization and implicit pattern association, which has long been at the heart of dual-route models, is combined with a further distinction between analytic and nonanalytic listing. The result is a threefold taxonomy of exponence mechanisms, where the idiosyncratic properties of stem-level morphophonology emerge from the interaction of nonanalytic listing with symbolic generalization. In this account, stem-level processes operate in a way that closely resembles Jackendoff's (1975) lexical redundancy rules, and indeed the exposition in section 2.3.3.1 will highlight the Jackendovian pedigree of the idea.

(12)

nonanalytic listing {	distributed associative memory	
	lexical redundancy rules	} explicit symbolic generalization
	standard rules	

In section 2.3.3.2, however, I will demonstrate that Stratal OT can model the phonological effects of Jackendovian lexical redundancy rules at the stem level without any addition to its existing phonological technology. For Stratal OT to handle a form like *Árabic*, three ingredients must come together:

(13) a. nonanalytic lexical listing,
b. morphosyntactic blocking,
c. high-ranking faithfulness in the phonology.

First, it is necessary that stem-level constructs like *-ic* adjectives should be entered into the lexicon with their stem-level phonological properties (including prosody) fully specified. Second, the presence of an entry for *Árabic* in the lexicon must be able to prevent the morphosyntax from building the adjective on line through the addition of the suffix *-ic* to the noun stem *Arab*, in just the same way that the lexical entry of *drove* blocks **drive-d*. Finally, metrical faithfulness must rank high in the stem-level phonological constraint hierarchy, so that the proparoxytonic contour specified in the lexical entry of *Árabic* can withstand neutralization to the default pattern exemplified by *idýll-ic, Miltón-ic, titán-ic*, etc.

This way of capturing the phonological effects of lexical redundancy rules makes two correct predictions (§2.3.3.2). First, the principle of free ranking requires us to countenance the possibility that, in the stem-level hierarchy, some faithfulness constraints may be superordinate whilst others may be crucially dominated by markedness. But, in the model outlined in (13), superordinate faithfulness plays a crucial role in protecting exceptional phonological properties stored in nonanalytic lexical entries. Stratal OT therefore predicts that the stem-level phonology will simultaneously enact two types of generalization:

(i) default rules (like penultimate stress in *-ic* adjectives), driven by markedness constraints subordinated to exception-protecting faithfulness; and
(ii) exceptionless well-formedness generalizations, enforced by top-ranked markedness constraints.

This is a good result: exceptionless well-formedness generalizations at the stem level are needed, *inter alia*, to express inviolable phonemic inventory restrictions (Bermúdez-Otero 2007a).

Second, this account makes strikingly accurate predictions about cyclic reapplication within stem-level domains, which I shall again illustrate with evidence from present-day English. Consider the familiar triads in (14), which involve two rounds of stem-level derivation.

(14)	a. órigin	b. orígin-al	c. orìgin-ál-ity
	órigin	orígin-ate	orìgin-át-ion

The complex forms in (14c) violate a stem-level phonological generalization known as the Abracadabra Rule (after Selkirk 1984: 117), which states that a sequence of three pretonic light syllables will bear secondary stress on the initial syllable: cf. *àbracadábra, dèlicatéssen, Mèditerránean, càtamarán*. This violation of the Abracadabra Rule is a cyclic effect: stress assignment reapplies after each round of affixation, and the foot-head assigned to the second syllable of *original* and *originate* in the second cycle blocks the enforcement of the Abracadabra Rule in the third cycle.

(15)		⟦⟦⟦origin⟧al⟧ity⟧
	first cycle	órigin
	second cycle	oríginal
	third cycle	orìginálity

Several aspects of this phenomenon raise problems for Lexical Phonology in its classical rule-based versions (e.g. Booij and Rubach 1987). First, the internal cyclicity of the stem level is stipulated rather than explained: Lexical Phonology simply states as an axiom that stem-level phonological rules reapply cyclically, whereas word-level and phrase-level rules apply once across the board within their respective domains. Second, Chung (1983: 63) observes that stem-level processes exhibiting cyclic reapplication in complex forms also sustain outright lexical exceptions in monomorphemic items: the Abracadabra Rule fails in forms like *Epàminóndas* and *apòtheósis*, bearing out Chung's Generalization (22)—but Lexical Phonology offers no account for this correlation. Third, Lexical Phonology makes no provision for the fact that cyclic reapplication within stem-level domains is itself subject to exceptions: in (16a), for example, we see the Abracadabra Rule unexpectedly displaying normal application (Collie 2007: 147); (16b) shows an instance of cyclic stress preservation failing in a different type of pretonic environment (Collie 2007: 289); and (16c) is an exception to the cyclic transmission of trochaic vowel shortening (Collie 2007: 289–90, and see note 10 below).

(16)	a.	illégible	ìllegibílity	(*alongside expected* illègibílity)
	b. (tríangle)	triángulate	trìangulátion	(*alongside expected* triàngulátion)
	c. c[aɪ]cle	c[ɪ]clic	c[aɪ]clicity	(*alongside expected* c[ɪ]clicity)

Stratal OT solves these three problems of Lexical Phonology at a single stroke by adopting the recipe for lexical redundancy rules given in (13). In this view, the mandatory nonanalytic listing of stem-level forms entails that, when a speaker of English first produces or perceives the adjective *oríginal*, she immediately assigns it a lexical entry specifying a foot-head on the second syllable: /*oríginal*/. The existence of this lexical entry blocks the suffixation of *-al* to *órigin* on line, thereby preventing the derivation of **òrigin-ál-ity* from the remote base *órigin* in one fell swoop. And, in turn, faithfulness to the input foot-head in /*oríginal-ity*/ takes precedence over the alignment constraints driving the Abracadabra Rule. Thus, the sequence shown in (15) does not consist of steps in a single on-line derivation, but rather holds between historical events of stem formation and storage under a régime of nonanalytic listing, blocking, and high-ranking faithfulness (13). In this model, then, the word level does not show cyclic reapplication internally because word-level forms are either unlisted or listed analytically: properties assigned in the output of the word-level phonology are therefore not stored, and so cannot be fed again as input to the word-level phonology from the lexicon. Chung's Generalization holds because the same high-ranking faithfulness constraints that protect the input foot-head on the second syllable of complex *orìginálity* guarantee its survival in monomorphemic *Epàminóndas*. And, finally, cyclic reapplication effects can fail because they crucially depend on blocking by lexical entries, but blocking is itself variable and depends on factors such as token frequency in ways explained by parallel race models of processing (Schreuder and Baayen 1995, Baayen, Dijkstra, and Schreuder 1997, Hay 2003): this point will be developed at length in §2.3.3.3, where I examine the famous case of *còmp*[ə]*nsátion* vs. *cònd*[ɛ̀]*nsátion*.

In sum, the remainder of section 2.3 will show that we can make progress on the question of the division of labor between storage and computation of exponents by implementing the refined dual-route framework outlined in (12) with the technology of Stratal OT.

2.3.2 Bloomfield's lexicon and SPE*'s evaluation measure*

Let us return to the division of labor between storage and computation in *SPE*. Chomsky and Halle's theory imposed very strong biases on the analysis of alternations: it discouraged storing allomorphs in the lexicon, and it favored deriving alternations by means of phonological rules (cf. also Bonet and Harbour this volume: §6.2.1). These biases were implemented through two important postulates of the theory:

(i) First, Chomsky and Halle did not recognize morphology as a separate module of the grammar: the surface structures generated by the syntax provided the input to the phonology (1968: ch. 1, §4). Nonetheless, *SPE* did in fact envisage a number of processes mediating between syntactic surface structures and underlying phonological representations (ch. 1, §5; ch. 8, §6). These processes included "readjustment rules" that, among other things, handled certain cases of allomorphy: e.g. a readjustment rule mapped $[\![_{V} [\![_{V} \mathit{sing}]\!]\ \textsc{past}]\!]$ onto $[\![_{V} \mathit{s}^{*}\mathit{ng}]\!]$, where /∗/ stood for an /i/ annotated with a diacritic feature that triggered a later phonological rule of ablaut converting /i/ into /æ/ (p. 11). Significantly, the /∗/→/æ/ transformation itself was claimed to take place in the phonology. The homonymous readjustment rules of Distributed Morphology operate in a very similar way: see Embick and Halle (2005: 41) for a discussion of *sang*. Embick and Halle (2005: 42) insist that the readjustment rules of Distributed Morphology **are** phonological rules, just as the /∗/ →/æ/ transformation of *SPE* was supposed to be (cf. §2.4.3 below).

(ii) Following Halle (1959), moreover, *SPE* posited an evaluation measure that selected the system of lexical entries and phonological rules containing the smallest number of symbols: this favored the adoption of any rule that dispensed with more symbols in lexical entries than it took to state the rule itself (pp. 334, 381, 389).

The prevalent, though by no means universally shared, view today is that *SPE*'s analytic biases were wrong. First, they favored descriptively inadequate grammars, i.e. grammars which misrepresented adult speaker competence. The main problem was the derivational remoteness of underlying representations, which could be vastly different from surface forms and related to them through extremely long and opaque derivations.[9] A notorious example was Chomsky and Halle's (1968: 233–4) postulation of velar fricatives in the underlying representations of *nightingale* /nixtVngǣl/

[9] In the controversy that ensued (e.g. Kiparsky 1968, Hyman 1970, Crothers 1971, etc.), this problem was often characterized as one of excessive 'abstractness.' It must be noted, however, that underlying representations may be 'abstract' in very different ways, not all of which incur the problems that afflicted *SPE*. As we shall see in §2.4.2.2, for example, the underlying representation of a morpheme may consist of a bare mora, which may be provided with segmental content by a variety of processes: e.g. by lengthening a vowel or by reduplicating a syllable-sized string; see (53). Such an underlying representation is undeniably

→ [ˈnaɪtn̩ˌgeɪl] and *dinghy* /dinxi/ → [ˈdɪŋi]: the former "explained" an exception to trisyllabic shortening (p. 52 and *passim*); the latter "accounted for" the unexpected occurrence of [ŋ] before a tautomorphemic vowel (cf. Bermúdez-Otero 2008). It is important to understand that these—for us—obviously misguided proposals were more than a fanciful indulgence, but were highly favored by *SPE*'s analytic biases. For example, given a relatively pervasive alternation like *divīne* ~ *divĭnity*, *serēne* ~ *serĕnity*, *sāne* ~ *sănity*, etc., one can effect a net reduction in the overall symbol count of one's description by positing a single underlier for each root and setting up a rule of trisyllabic shortening targeting long vowels in antepenultimate syllables preceding a stressless penult (*SPE*, p. 52).[10] Having done that, however, one needs to exempt *nīghtingale* from this rule. If one regards rule diacritics of the format [±rule *n*] (Chomsky and Halle 1968: ch. 8, §7) as devices of last resort,[11] then *SPE*'s evaluation measure forces one to look for the alternative that spends the fewest additional symbols: this alternative may well involve adding a feature or segment to an underlier (in this case, a /x/ to /nixtVngǣl/) for the sole purpose of opportunistically triggering a rule that happens to be already lying about (in this case, a battery of rules dealing with the appearance of surface [tʃ] in *question*, *bastion*, *righteous*, etc.).

Second, the grammars favored by the analytic biases of *SPE* were unlearnable except under extremely idealized and unrealistic conditions. Chomsky (1957: ch. 6) had earlier disclaimed the ambition to provide a "discovery procedure for grammars", i.e. an unsupervised learning algorithm inferring a grammar from a corpus; he had settled instead for the lower goal of devising an "evaluation procedure", i.e. a criterion for arranging a given set of grammars in an order of preference, given a corpus (see further Chomsky 1965: ch. 1, §6–7). In this idealized scenario, an unspecified function of Universal Grammar provides the learner with a set of grammars compatible with the primary linguistic data, and the evaluation measure chooses the best. In the 1950s this was an expedient compromise: some structuralist linguists had arguably taken a wrong turn by subjecting grammars to conditions which were designed to make a certain sort of discovery procedure viable, but which in effect made it impossible to provide adequate characterizations of fundamental aspects of linguistic structure, including morphology–phonology interactions (e.g. Hockett 1942: 20–1, Moulton 1947: note 14; cf. Pike 1947, 1953). Whatever the merits of Chomsky's strategy in the 1950s, however, it **must** remain a supreme goal of linguistic theory to devise unsupervised learning algorithms that infer descriptively adequate grammars from the sort of input available to children. Yet it seems clear that there is no feasible discovery procedure

'abstract' (in the sense that it omits a great deal of concrete surface detail), but it need not require the extremely opaque derivations that characterized *SPE*: indeed, the relevant mappings may be transparent and monotonically structure-building.

[10] Prince (1990: 368–70) demonstrates that, in fact, trisyllabic shortening is trochaic shortening (Hayes 1995: 142–9) under final syllable extrametricality. This explains why the suffix *-ic*, which normally induces mere consonant extrametricality (18a), induces trochaic shortening of stressed penults, as in *c*[aɪ]*cle* ~ c[ɪ]*cl-ic* (16c).

[11] As Spencer (1991: 101) puts it, "Classically, generative phonologists have tried to use rule features as sparingly as possible."

for the grammars most highly valued by *SPE*'s evaluation measure:[12] automated rule learners do of course exist (e.g. Albright and Hayes 2003), but they do not return *SPE*-style grammars. OT provides a good term of comparison: the framework has given rise to explicit learning algorithms tested in computer simulations of nontrivial acquisition tasks (e.g. Tesar and Smolensky 2000, Boersma and Hayes 2001).

These two problems are of course intimately related: if the grammars favored by the analytic biases of *SPE* are unlearnable under realistic conditions (second problem), then they cannot be descriptively adequate (first problem). But what went wrong? The received wisdom, with which I agree, is that *SPE* struck the wrong balance between storage and computation: see e.g. the papers in Nooteboom, Weerman, and Wijnen (2002), specially Pinker and Ullman (2002), Jackendoff (2002a), and Booij (2002). However, the usual assertion that *SPE* favored the minimization of storage and the maximization of computation does no more than trivially restate the effects of *SPE*'s evaluation measure; it fails to explain **why** Chomsky and Halle chose that measure over others. Significantly, *SPE* did **not** overtly base its choice of evaluation measure on a claim that the properties of the human brain put a premium on storage in long-term memory—though this claim is made, and half-heartedly defended, by Bromberger and Halle (1989: 56–7, especially footnote 7). Rather, the root of the problem lay, as Jackendoff (2010: 37) correctly observes, in Chomsky and Halle's espousal of Bloomfield's (1933: 274) conception of the lexicon as an unstructured, cognitively inert list: "The lexicon is really an appendix of the grammar, a list of basic irregularities" (echoed in Chomsky 1965: ch. 2, endnote 16). If the grammar comprises only rules and lexical entries, and everything contained in lexical entries is an accident, then it follows that every nonaccidental pattern, i.e. every pattern exhibiting any symptom of psychological reality (§2.3.1), must be expressed as a rule.

Now consider two such patterns:

(i) The ablaut alternation present in English verbs like *string* ~ *strung*, *stick* ~ *stuck*, and *dig* ~ *dug* since the sixteenth century was generalized to the originally weak verb *sneak* in American dialects possibly as late as the nineteenth century: this yielded the innovative strong past-tense form *snuck* (Hogg 1988; see further Wełna 1997). This analogical extension shows that the alternation between a high front vowel in the unmarked form and [ʌ] in the past tense was in some sense psychologically real for speakers; otherwise it could not have been extended. Therefore, if rules provide the **only** way to represent psychologically real patterns, and if phonological rules are the preferred type of rule in this case, then one is forced to countenance phonological processes of ablaut for strong verbs in present-day English: see Halle and Mohanan (1985: 107–14) and Halle and Marantz (1993: 127–9).

(ii) The English word *righteous* appears to be semantically and phonologically related to the word *right*: someone is righteous if they do right, and the string

[12] In fact, even the evaluation measure itself was never applied as defined, since no one ever attempted to compare the global symbol counts of two complete packages of lexicon plus rules (Prince 2007: §2.1.2).

> [ɹaɪt] is approximately contained within [ˈɹaɪʧəs].[13] Let us assume that this relationship is psychologically real for English speakers. If such a relationship can **only** be captured by a rule, and if the **only** relevant type of rule available is a phonological process mediating between underlying and surface representations, then one is forced to derive *right* and *righteous* from a common underlier and to capture the alternation between [t] and [ʧ] by means of some phonological transformation lying conveniently to hand. For the latter, the obvious candidate is *SPE*'s rule of palatalization (p. 230), which can easily be triggered by positing an underlying high vocoid in *-eous*, as found on the surface in *bil-ious* (cf. *bile*), *censor-ious* (cf. *censor*), *labor-ious* (cf. *labor*), and *uproar-ious* (cf. *uproar*). Unfortunately, however, the expected realization of an underlying /t/ in this environment is [ʃ]: cf. *expedi*[t]*e* ~ *expedi*[ʃ]*ious*, *infec*[t] ~ *infec*[ʃ]*ious*, and, with other suffixes, *abor*[t] ~ *abor*[ʃ]*ion*, *Egyp*[t] ~ *Egyp*[ʃ]*ian*, *par*[t] ~ *par*[ʃ]*ial*, etc. It therefore looks as though the common underlier for *right* and *righteous* cannot be the expected /ɹaɪt/, or rather /rīt/,[14] at least if a morphophonemic readjustment or a rule diacritic are to be avoided. This difficulty, in combination with a multiplicity of other facts, eventually leads to *SPE*'s choice of /rixt/ and /rixt+i+ɔs/ as underlying representations, thereby setting up the alibi for /nixtVngǣl/ and /dinxi/ (pp. 233–4).

The dilemma becomes pressing: if one must posit an evaluation procedure (simply because a discovery procedure is out of one's reach), and if this evaluation procedure must assign a high value to phonological rules of ablaut for strong verbs and to phonological derivations relating *righteous* to *right* via a shared underlier (simply because this is the only way to capture psychologically real patterns in one's framework), then one is likely to end up stuck with an evaluation measure that also favors /nixtVngǣl/ and /dinxi/.[15]

[13] I say 'approximately' because it is not safe to assume without further argument that one can equate a /t/ with the closure phase of a /ʧ/ either in the phonology or in the phonetics.

[14] The expected underlier in *SPE* is /rīt/ because all long vowels take a free ride on a noncyclic word-level rule of vowel shift: see note below.

[15] The afterlife of the /nixtVngǣl/ scandal holds considerable historical interest and deserves to be recorded in a footnote. Rule-based Lexical Phonology attempted to get rid of the velar fricative of /nixtVngǣl/ without altering *SPE*'s symbol-counting measure by invoking Mascaró's (1976) phonological version of the Strict Cycle Condition, which prevented cyclic rules from applying in nonderived environments. Since trisyllabic shortening was designated as a cyclic rule (despite problematic cases like (16c)), the Strict Cycle Condition stopped it from applying to underived *nightingale*, and so *SPE*'s underlier /nixtVngǣl/ could be replaced with less remote /nītVngǣl/ (Kiparsky 1982b: 57–8, Kaisse and Shaw 1985: 15–17). Nonetheless, the new underlier /nītVngǣl/ still had to take a free ride on a noncyclic word-level rule of vowel shift to become [ˈnaɪtn̩ˌgeɪl]. No greater gains in transparency were possible: the Strict Cycle Condition still allowed absolute neutralization to take place postcyclically (Kiparsky 1985: 87–8), and the symbol-counting evaluation procedure still continued to adjudicate against the replacement of a single noncyclic word-level rule of vowel shift with two cyclic stem-level rules subject to blocking in non-derived environments (cf. McMahon 1990). "Derivational simplicity is strictly subordinated to grammatical simplicity," Kiparsky (1982b: 57) had proclaimed. McMahon (2000: e.g. 53) is rightly dissatisfied with this outcome and rightly blames it on *SPE*'s simplicity (2000: e.g. 138), which she therefore replaces with a new evaluation measure incorporating Kiparsky's (1973b: 65) Alternation Condition (McMahon 2000: 84). However, Bermúdez-Otero and Hogg (2003: 107–9) argue that McMahon's own proposal falls foul of the

There are several conceivable ways out of this dilemma. In this chapter I pursue the strategy of denying *SPE*'s premise that all psychologically real patterns must be expressed as standard rules, i.e. as symbolic generalizations that apply mandatorily whenever their conditions are met. Standard rules may instead be reserved for linguistic patterns that exhibit full productivity; in contrast, semiproductive patterns (like English stem-level derivation) and sporadically extended idiosyncrasies (like the synchronic relics of ablaut in present-day English verb inflection) can be handled with other tools. These tools come within our reach if we discard Bloomfield's conception of the lexicon as an unstructured, cognitively inert repository of irreducibly arbitrary stipulations; it then becomes possible to account for morphophonological patterns of limited productivity by means of devices that capture psychologically real relations between stored lexical items. Here I shall consider two such devices: one symbolic and explicit (Jackendovian lexical redundancy rules), the other subsymbolic and implicit (distributed associative memory).[16]

2.3.3 Lexical redundancy rules

Jackendoff (1975) took a decisive step away from the Bloomfieldian lexicon by completely overhauling *SPE*'s treatment of lexical redundancy rules (cf. Chomsky and Halle 1968: 163, 171, 380–9); variants of the same idea were proposed soon after by Aronoff (1976) and Lieber (1980).[17] Jackendoff's lexical redundancy rules are particularly well suited to account for the properties of certain morphological constructions that display limited productivity and sustain lexical exceptions. Unlike *SPE*, Jackendoff proposes that the forms instantiating such constructions are listed the lexicon in full—or, in the terminology I adopted in §2.3.1, nonanalytically. However, the grammar also contains symbolic generalizations that capture the regularities holding over such forms: these rules may occasionally be used to generate novel forms (whence their semiproductivity), and their application is blocked by contradictory specifications in lexical entries (thereby sustaining exceptions).

In this section I use Jackendoff's idea to elucidate certain aspects of stress assignment in English stem-level derivatives. The analysis rests on the premise that the output of each round of stem-level morphology and phonology receives its own lexical entry, and that certain stem-level phonological rules can be defeated by lexically prespecified structure (§2.3.3.1). Two interesting results follow. We shall see, first, that those stem-level phonological processes that are blocked by prespecified structure become prone to developing cyclic reapplication in the course of history. This explains the recurrent coupling of exceptions with cyclic effects within stem-level

Duplication Problem (Clayton 1976, Kenstowicz and Kisseberth 1977: 136ff.), a difficulty which does not arise in a constraint-based framework (McCarthy 2002: 71ff.). In addition, McMahon's system still relies on a version of the Strict Cycle Condition, which is now known to be false (Kiparsky 1993, Bermúdez-Otero forthcoming).

[16] Albright and Hayes's (2003) alternative, which I cannot evaluate in this chapter, is to retain a single-mechanism rule-based framework but to replace *SPE*'s deterministic derivations with probabilistic rule-systems that generate multiple competing outputs annotated with numerical confidence values: see §2.3.4 below.

[17] Jackendoff (2010: 35–9) provides a brief outline of the historical origins and development of his idea. Further discussion can be found in Jackendoff (1997: 123–30, 2002b: 53, 158ff.).

domains (Chung 1983: 63): both arise from the same grammatical mechanism, namely nonanalytic listing and blocking (§2.3.3.2). Second, if blocking is implemented in processing by means of a parallel race between the lexicon and the rule-system, we can explain the effect of token frequency not only on exceptionality but also on stem-level internal cyclicity. This will be illustrated with a new look at an old classic: *SPE*'s discussion of pretonic vowel reduction in *compensation* vs. *condensation* (§2.3.3.3). These results show that, at least in some key cases, the question of the interaction between morphology and phonology cannot be separated from the problem of demarcating storage from computation.

2.3.3.1 Lexical redundancy at the stem level A salient property of English stem-level derivation is its semiproductivity. It is not the case that, if a hypothetical derivative complies with all the semantic, syntactic, morphological, and phonological subcategorization requirements of a stem-level affix, it will thereby come into existence. For example, speakers of English have to learn whether or not, for each verb, there happens to be a stem-level nominalization and, if there is one, which exponent it takes and which idiosyncratic semantic restrictions it may have: see the examples in (17).

(17)		*-ion*	*-al*	*-ance*
a. *commit*		*commission*	*committal*	*committance*
	OED entry?	yes	yes	yes ("obsolete, rare")
	tokens per 10^6 words in BNC	112.04	2.65	0
b. *permit*		*permission*	*permittal*	*permittance*
	OED entry?	yes	no	yes
	tokens per 10^6 words in BNC	33.84	0	0
c. *submit*		*submission*	*submittal*	*submittance*
	OED entry?	yes	yes ("rare")	yes ("obsolete")
	tokens per 10^6 words in BNC	15.66	0	0

OED = *Oxford English Dictionary*
BNC = British National Corpus

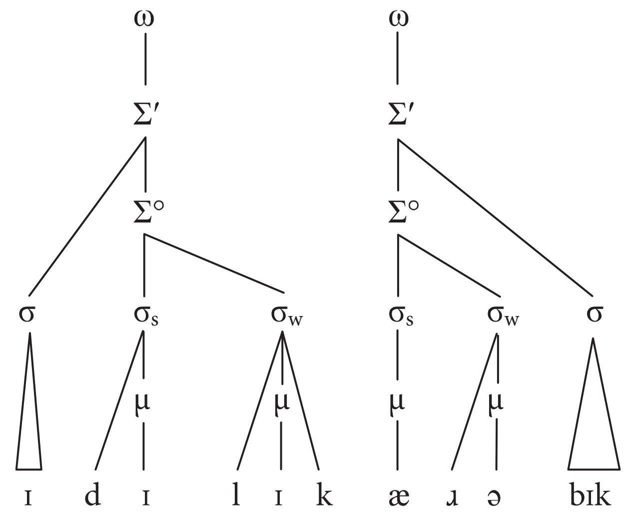
A possibility may remain unrealized indefinitely; but, should the need be felt, affixation may be used generatively to create a new derivative. For example, the noun *submittal* (17c) appeared in American English in the nineteenth century: the *OED*'s first attestation, from an American source, is dated 1888. *Submittal* has since gained currency, particularly in academic and legal discourse, with the meaning 'act of submitting a document', which is based on the transitive use of *submit*:[18] the Corpus of

[18] This appears to be true for most speakers. However, Google does return some tokens of *submittal* corresponding to the intransitive use of *submit* ('yield to a higher power'): e.g. "prayerfulness and submittal to the will of the Lord" (http://www.benpres-holdings.com/speeches.php?id=50, accessed on 27 April 2010).

Contemporary American English (COCA) contains 41 tokens, 37 of which occur in the academic section of the corpus. Since the meaning of *submittal* is thus idiosyncratically restricted, the noun must have its own lexical entry. Significantly, *submittal* does not occur in British English, where its uses are covered by *submission*: in contrast with the 3 tokens in the American National Corpus (ANC) and the 41 in COCA, the BNC contains none.

Moreover, many of the regularities that hold over stem-level derivatives sustain outright exceptions. This is particularly clear in the case of phonological patterns. We saw above, for example, that *righ*[t] ~ *righ*[t͡ʃ]*-eous* is an exception to the normal pattern of palatalization: cf. *infec*[t] ~ *infec*[ʃ]*-ious*. Similarly, *ob*[iː]*se* ~ *ob*[iː]*s-ity* is an exception to trisyllabic shortening: cf. *ser*[iː]*ne* ~ *ser*[ɛ]*n-ity*. Stress assignment is particularly revealing. Unlike word-level suffixes, stem-level suffixes affect the location of stress, but the rules governing their metrical behavior are riddled with exceptions. As we saw in (11), for example, the foot structure of *Árab* ~ *Árab-ic* departs from the usual pattern of final consonant extrametricality (Hayes 1982) associated with *-ic* suffixation, which is at work in *ídyll* ~ *idýll-ic*.[19]

(18) a. *idýlli<c>* b. *Ára<bic>*



Yet, at the same time, there is a great deal of evidence to show that the stem-level stress assignment rules of English can nonetheless be used generatively (e.g.

[19] In (18a), the extrametricality of the final consonant is reflected in its failure to project a mora. In (18b), the whole final syllable is extrametrical, and so excluded from the foot's zero projection (Σ°). The evidence for attaching stray syllables within stem-level domains to a higher foot-projection (Σ′), rather than directly to the prosodic word (ω), comes from /t/-flapping in American English: in expressions like *a* [tʰ]*obóggan* and *Mèdi*[tʰ]*erránean*, rightward attachment of the pretonic stray syllable to Σ′ places the /t/ in initial position within a foot-projection, so protecting it from flapping (Jensen 2000: 189, 209–11, Davis and Cho 2003: 613–14); direct attachment to ω could account for the absence of flapping in *a*[tʰ]*obóggan* (on the assumption that the left edge of ω is a strong position), but it cannot account for the absence of flapping in *Mèdi*[tʰ]*erránean* (*pace* Selkirk 1996: 197–8). In contrast, stray syllables affiliated to word-level suffixes do attach to ω: see Bermúdez-Otero (2011: §4) for durational evidence.

Hayes 1982: 236–7). Notably, their application is responsible for the regularization of exceptions in diachronic change: e.g. conservative British RP *ápplic-able*> advanced British RP *applíc-able* (Wells 1990: *sub voce*); see also the discussion of diatonic pairs like *tòrmént*$_V$ ~ *tórmènt*$_N$ in §2.4.2.3 below.[20] Similarly, they are used generatively in loanword adaptation: e.g. Russian *Nínotchka, bábushka* > English *Ninótchka, babúshka*. It is interesting to note that these loanwords have been nativized metrically, but not segmentally: the clusters [ʧ.k] and [ʃ.k] occur tautomorphemically only in borrowings like *A*[ʃ.k]*enazi* and *A*[ʃ.k]*elon*. As argued by Pinker (1999: 217, 222, 230, 232) and Pinker and Ullman (2002: 458), application to phonotactically deviant items is a hallmark of linguistic generalizations explicitly represented in the mind as symbolic rules (§2.3.4).

It thus looks as if the morphological and phonological patterns holding over English stem-level derivatives are explicitly represented in the grammar by means of symbolic generalizations, i.e. by means of 'rules' in the broad sense. Nonetheless, many of these rules exhibit a cluster of properties (notably, semiproductivity and exceptions) that distinguish them from standard rules, i.e. from the sort of rule that applies mandatorily whenever its structural description is met. Following Jackendoff (1975), we can solve this apparent contradiction by postulating that linguistic rules can apply not only in standard mode, but also in lexical redundancy mode. Indeed, in agreement with Jackendoff (2010: 32), I shall assume that lexical redundancy rules have exactly the same format as ordinary rules, and differ from the latter only in their mode of application.[21] In this view, then, stem-level morphological and phonological rules (but not, I assume, word-level ones) can be designated as applying in lexical redundancy mode. The lexical redundancy mode involves a special relationship between the rules so designated and the contents of the lexicon. This special relationship has two aspects: the nonanalytic listing of rule outputs, and the blocking of rules by prespecified information. Of these two properties, the first is defining: some standard rules, such as the word-level processes involved in English regular past-tense marking, undergo blocking just like lexical redundancy rules (§2.3.1); but only the output of lexical redundancy rules is listed nonanalytically.

We must first assume, then, that the **output** of each full cycle of the stem-level morphology and phonology is stored in its own lexical entry. This means that stem-level derivatives are entered into the lexicon nonanalytically (Kaye 1995: 302ff.): i.e. as full forms (Jackendoff 1975: 643ff.), rather than as concatenations of underlying morphs. For example, the adjective *idýllic*, derived from *ídyll* by means of stem-level suffixation, will have its own entry in the lexicon, and this entry will contain

20 Diachronic processes of regularization show different behavior from sporadic extensions of irregular patterns, as in the case of *snuck* in §2.3.2 above (Fertig 1999; see further Prasada and Pinker 1993).

21 Accordingly, recognizing the lexical redundancy mode of application does not tell us whether the format of exponence processes is, in Stump's (2001: 1) terms, 'lexical' or 'inferential': e.g. Anderson (1992: 186) and Bochner (1992) posit lexical redundancy rules within inferential frameworks, but Jackendoff (1997: chs 5 and 6, 2002b: 161–2, note 3) does so within a piece-based architecture. Jackendoff's rejection of 'process morphology' is grounded on considerations of modularity (see §2.4.2.1 and §2.4.2.3 below), and is independent of his assumptions about the balance between storage and computation in the Language Faculty.

a phonological representation consisting not of a string of two underliers (19b), but rather of a single fully prosodified stem-level output structure (19a), including a foot-head on the penult. However, the lexical entry of *idyllic* will **not** record the effects of word-level and phrase-level phonological rules, which do not apply in lexical redundancy mode.

(19) *Nonanalytic listing of a stem-level form:* idyllic
 a. ✓ IDYLLIC ↔ [$_{\omega}$[$_{\Sigma'}$ɪ[$_{\Sigma\circ}$ˈdɪ$^{\mu}$.lɪ$^{\mu}$k]]] = (18a)
 b. ✘ IDYLLIC ↔ ⟦$_{SL}$ɪdɪl-ɪk⟧

In contrast, word-level constructs, if entered into the lexicon at all (a possibility supported by Baayen et al. 2002), will be listed analytically: i.e. as they appear in the **input** to the word-level phonology.

(20) *Analytic listing of a word-level form*: loaded
 a. ✓ <LOAD, past> ↔ ⟦$_{WL}$[$_{\omega}$[$_{\Sigma}$lə$^{\mu}$ʊ$^{\mu}$d]]-d⟧
 b. ✘ <LOAD, past> ↔ [$_{\omega}$[$_{\Sigma}$ˈlə$^{\mu}$ʊ$^{\mu}$]dɪd]

Note that the concept of analytic listing is independently needed to account for the behavior of phrasal idioms like *pull* ⟦$_{PossP}$X⟧ *leg* 'tease X'. This construction must have its own lexical entry because its semantics is noncompositional; but, at the same time, its internal constituent structure must be visible to the syntax, since it can undergo operations like *wh*-extraction (e.g. ⟦*Whose leg*⟧$_i$ *are you trying to pull* t$_i$?). Therefore, the syntactic structure of this idiom must be stored analytically (see Jackendoff 2002b: 167–72).[22] This contrast between analytic and nonanalytic listing also finds a parallel in Clahsen and Neubauer's (2010: 2634) distinction between 'combinatorial lexical entries' and 'unanalyzed lexical entries' (see further Stemberger and MacWhinney 1986, 1988).

Second, we must assume that a rule in lexical redundancy mode can build, but cannot change, structure: in other words, lexical redundancy rules capture default patterns, but are **blocked** by prespecified information. This postulate, together with the hypothesis of stem-level nonanalytic listing, accounts for the prevalence of exceptions among stem-level morphophonological patterns. Under the assumption of nonanalytic listing, for example, the lexical entry of the adjective *Arabic* contains the phonological representation shown in (21a), rather than the decomposed alternative given in (21b). Now suppose that the stem-level phonological rules of final consonant extrametricality and foot construction are designated as applying in lexical

[22] Marantz (1997b) claims that the noncompositional reading of an idiom like *pull* ⟦$_{PossP}$ X⟧ *leg* is a contextually determined property of a single root ($\sqrt{\text{PULL}}$), but the claim is rejected by Williams (2007: 359–61) and by Horvath and Siloni (2009). More generally, Marantz's theory of domains for noncompositional meanings, as developed by Marvin (2002: 39), Arad (2003: 740), Embick and Marantz (2008: 11), and Embick (2010: 44), is beset with difficulties: see e.g. Alexiadou (2009), Anagnostopoulou and Samioti (2009a,b), Bermúdez-Otero (2013), Borer (2009), Harley (2009), and Wechsler (2008).

redundancy mode.[23] If that is the case, then the stored nonanalytic representation in (21a) will block the regular assignment of stress to the penultimate syllable.

(21) *Nonanalytic lexical entry for the adjective stem* Arabic

a. ✓ ARABIC ↔ [$_{\omega}$[$_{\Sigma'}$[$_{\Sigma\circ}$ˈæ$^{\mu}$.ɹæ$^{\mu}$]bɪk]] (C-extrametricality and penult stress blocked)

b. ✘ ARABIC ↔ ⟦$_{SL}$æɹæb-ɪk⟧ (→$_{SL}$ * [$_{\omega}$[$_{\Sigma'}$ æ[$_{\Sigma\circ}$ˈɹæ$^{\mu}$.bɪ$^{\mu}$k]]])

In psycholinguistic terms, I assume that blocking is implemented by means of a parallel race between the lexicon and the rule-system (Schreuder and Baayen 1995, Baayen et al. 1997) within a dual-route model of processing (e.g. Prasada and Pinker 1993, Clahsen 1999, Pinker 1999, Ullman 2001, Pinker and Ullman 2002). This predicts—correctly, as we shall see—that blocking, and therefore the liability of stem-level redundancy rules to bear exceptions and to show cyclic reapplication, will correlate with token frequency (§2.3.3.3).

When there is no lexical entry to block its application, however, a set of lexical redundancy rules **can** be used generatively; but, in line with our first hypothesis, the output of such instances of generative use will automatically undergo nonanalytic listing. In this view, the noun *submittal* is entered into a speaker's lexicon when first produced or encountered: from that point onwards, neither the suffixation of *-al* nor stress assignment need apply on line (unless lexical retrieval proves too slow); the rules revert to merely encoding the predictable aspects of the relationship between *submit* and *submittal*. This fact crucially distinguishes structure-building operations designated as applying in lexical redundancy mode from structure-building operations applying in standard mode; an example of the latter is regular past-tense formation by means of /d/-suffixation in present-day English (§2.3.1). Although both types of rule undergo blocking, the outputs of standard structure-building rules may remain unlisted or, if entered into the lexicon, do so in analytic form: cf. (19a) and (20a). Crucially, the assumption that word-level regular past-tense forms are either unlisted or listed analytically, as in (20a), correctly predicts the fact that the phonological alternation of the suffix /-d/ between [-d], [-t], and [-ɪd] is strictly exceptionless, as shown experimentally by Albright and Hayes (2003: 151).[24]

[23] As I anticipated in §2.3.1, not all stem-level phonological rules apply in lexical redundancy mode; some, e.g. those enforcing inviolable inventory restrictions, are exceptionless. See the discussion surrounding (24) below for another example, with an optimality-theoretic implementation.

[24] Words like *beloved* [bɪˈlʌvɪd] 'dear' and *winged* [ˈwɪŋɪd] 'possessing wings' must be treated as stem-level items and listed nonanalytically, but, since they are adjectives, they do not disprove the assertion that the past-tense forms of regular weak verbs are word-level and therefore either unlisted or listed analytically: cf. *I* [ˈwɪŋd] *it* 'I improvised,' not **I* [ˈwɪŋɪd] *it*. A more intriguing case is that of the verb *text* 'send a message to a mobile phone': its expected past-tense form is [ˈtɛkstɪd], but the variant [tɛkst] has attained wide currency. The rise of this variant may plausibly be connected with the fact that, in nonpast forms like /tɛkst-Ø/ and /tɛkst-s/, the stem-final consonant is very highly vulnerable to *t*-deletion (Guy 1994). I therefore conjecture that frequent exposure to perceived tokens of nonpast [tɛks] led some speakers to produce or parse [tɛkst] as a regular past-tense form: i.e. /tɛks-d/ ↔ [tɛkst]. Other speakers in turn treated this novel past-tense form as a listed irregular, which was liable to became entrenched in the lexicon because of its similarity with verbs like *burst*, *cost*, etc. (see §2.3.4 below). If this hypothesis is on the right track, then past-tense [tɛkst] does not provide evidence for irregularity in the phonological alternation of the suffix /-d/.

2.3.3.2 The emergence of stem-level cyclicity: órigin, oríginal, orìginálity As I anticipated in §2.3.1, a striking array of morphophonological facts finds its explanation in the hypothesis that stem-level constructs are stored nonanalytically and block stem-level phonological rules applying in lexical redundancy mode. Notable among these facts is Chung's Generalization:

(22) *Chung's Generalization* (after Chung 1983: 63)
If a stem-level phonological process can sustain lexical exceptions in monomorphemic items, then it can show cyclic reapplication in complex stem-level forms, and vice versa.

As an example of cyclic reapplication inside a stem-level domain, I shall use the case of *órigin* → *oríginal* → *orìginálity*, presented in (14) above. In this section the logic behind Chung's Generalization will reveal itself with special clarity once we have determined how lexical redundancy rules in the stem-level phonology are to be formalized in a constraint-based framework like Stratal OT (Bermúdez-Otero and McMahon 2006: 400, Kiparsky 2007, Collie 2007: 252ff., 2008, Bermúdez-Otero 2008).

As we have already seen, English has a stem-level phonological generalization known as the Abracadabra Rule (after Selkirk 1984: 117), whereby a pretonic sequence of three light syllables bears secondary stress on the first syllable, creating an initial dactyl.

(23) *The Abracadabra rule*
$[_{\omega}$ σ̆σ̆σ̆σ́… → $[_{\omega}$ σ̆̀σ̆σ̆σ́…
àbracadábra, dèlicatéssen, Mèditerránean, càtamarán

This generalization sustains outright lexical exceptions among monomorphemic items (e.g. *Epàminóndas*) and among items derived from bound roots (e.g. *apòtheósis*); these exceptions follow clear trends, but are nonetheless unpredictable (Collie 2007: 155–8). The Abracadabra Rule thus proves to be a stem-level phonological rule applying in lexical redundancy mode: it defines a default pattern, but it is blocked by prespecified information. In an optimality-theoretic framework, this can easily be expressed as an emergence-of-the-unmarked effect by means of the stem-level ranking MAX-Head(Σ) ≫ ALIGN(ω,L;Σ°,L) ≫ ALIGN(Σ°,R;ω,R).[25]

(24) a. MAX-Head(Σ)
For every input segment that is the designated terminal element of a foot, assign one violation mark if it does not have an output correspondent that is the designated terminal element of a foot.

b. ALIGN(ω,L;Σ°,L)
For every prosodic word, assign one violation mark if its left edge is not aligned with the left edge of some foot's zero-projection.

[25] As we saw in note 19 above, present-day English allows foot recursion. For this reason, the constraint definitions in (24b,c) refer specifically to zero projections of the foot: i.e. Σ°. For the sake of legibility, higher projections (Σ′, Σ″, etc.) are omitted throughout section 2.3.3.2.

c. Align(Σ°,R;ω,R)
For every foot's zero-projection, assign one violation mark for every syllable intervening between its right edge and the right edge of the prosodic word.

d.

			Max-Head(Σ)	Align(ω,L;Σ°,L)	Align(Σ°,R;ω,R)
i. *default pattern:* σ̆σ̆σ̆[$_{\Sigma°}$ σ̆́]σ		[$_{\omega}$ σ̆[$_{\Sigma°}$ σ̆̀σ̆][$_{\Sigma°}$ σ̆́]σ]		1!	2+1=3
		[$_{\omega}$ [$_{\Sigma°}$ σ̆̀σ̆]σ̆[$_{\Sigma°}$ σ̆́]σ] ☞			3+1=4
ii. *exception:*	[$_{\omega}$ a[$_{\Sigma°}$ pò.the][$_{\Sigma°}$ ó]sis]	[$_{\omega}$ [$_{\Sigma°}$ à.po]the[$_{\Sigma°}$ ó]sis]	1!		3+1=4
		[$_{\omega}$ a[$_{\Sigma°}$ pò.the][$_{\Sigma°}$ ó]sis] ☞		1	2+1=3

The high-ranking faithfulness constraint Max-Head(Σ) preserves lexically specified pretonic stress contours in stored items, whether regular like *àbracadábra* or exceptional like *apòtheósis*. If pretonic footing is not specified in the input, then the ranking of the subordinate alignment constraints imposes the default initial-dactyl pattern: in particular, the requirement that the prosodic word should begin with a foot's zero projection takes precedence over general rightward alignment. Zuraw (2010) uses the same optimality-theoretic technology in her analysis of nasal substitution in Tagalog prefixal constructions: dominant faithfulness constraints protect lexical exceptions in stored items, including nonanalytically listed complex words; subordinate markedness constraints express the default pattern, which applies to novel instances of prefixation. Interestingly, Zuraw further shows that, under stochastic ranking with the Gradual Learning Algorithm (Boersma 1997, Boersma and Hayes 2001), the strength of different defaults will be reflected in the absolute ranking values of subordinate markedness constraints, acquired on the basis of lexical frequencies (see the discussion of (35) in §2.3.3.3 below; cf. Kager 2009: 412).

Incidentally, high-ranking Max-Head(Σ) is also responsible for violations of FtBin (Prince and Smolensky 1993: §4.3) like *fàscístic* [ˌfæ.ˈʃɪ.stɪk] and *mòdérnity* [ˌmɒ.ˈdɜː.nɪ.ti]. Nonetheless, there appear to be no exceptions to foot binarity of the type [$_{\omega}$[$_{\Sigma}$ σ̆̀σ̆] [$_{\Sigma}$ σ̆̀]σ́…]. If this is true, then the fact may be captured by ranking Max-Head(Σ) below a fairly specific markedness constraint m forbidding unary feet in the relevant environment. This shows how, in Stratal OT, the stem-level constraint hierarchy can do double duty, simultaneously expressing both exceptionless well-formedness restrictions (e.g. m ≫ Max-Head(Σ)) and exception-tolerating lexical redundancy rules (e.g. Max-Head(Σ) ≫ FtBin ≫ Align(ω,L;Σ°,L)). As we saw in §2.3.1, this is a welcome result, for we often need exceptionless well-formedness generalizations at the stem level to express inviolable phonemic inventory restrictions (Bermúdez-Otero 2007a). However, the correlation with cyclic reapplication inside complex stem-level forms enshrined in Chung's Generalization (22) holds only for exception-tolerating default patterns, and Stratal OT **predicts** this fact.

Now, given the constraint ranking established in (24d), the hypothesis of stem-level nonanalytic listing predicts that the Abracadabra Rule will be liable to cyclic misapplication effects. Consider, for example, a hypothetical English speaker whose lexicon contains the words *órigin* and *oríginal*, but who has not yet encountered the derived form *originality*. *Ex hypothesi*, this speaker's lexical entries for *órigin* and *oríginal* will specify the location of foot-heads, since these are created at the stem level and so are subject to nonanalytic listing:

(25) a. ORIGIN ↔ [$_{\omega}$[$_{\Sigma^\circ}$**ó**.ri]gin]

b. ORIGINAL ↔ [$_{\omega}$o[$_{\Sigma^\circ}$r**í**.gi]nal]

Now suppose that this speaker has cause to use the stem-level morphology and phonology generatively to create *originality*.[26] The on-line derivation will start with *oríginal* rather than *órigin*, because the existence of a lexical entry for *oríginal* blocks the application of *al*-suffixation to *órigin*. Accordingly, the morphology will suffix *-ity* to (25b) and will submit the result to the phonological constraint hierarchy in (24d). As shown in tableau (26), the result is *orìginálity*, with failure of the Abracadabra Rule.

(26)

[$_{\omega}$ o[$_{\Sigma^\circ}$ r**í**.gi]nal] - ity	MAX-Head(Σ)	ALIGN(ω,L; Σ°,L)
[$_{\omega}$ [$_{\Sigma^\circ}$ ò.r**i**]gi[$_{\Sigma^\circ}$ ná.li]ty]	*!	
[$_{\omega}$ o[$_{\Sigma^\circ}$ r**ì**.gi][$_{\Sigma^\circ}$ ná.li]ty] ☞		*

The Abracadabra Rule thus provides a clear example of Chung's Generalization; for other cases, see the discussion of *cònd*[ɛ̀]*nsátion* in §2.3.3.3, of British *addréss*$_{N}$ and American *áddrèss*$_{N}$ in §2.4.2.3, and of *swi*[ŋ]*ómeter* in Bermúdez-Otero (2008).

As we saw in §2.3.1, this theory of lexical redundancy rules improves on classical versions of Lexical Phonology (e.g. Booij and Rubach 1987), which simply stipulated the fact that internal cyclic reapplication effects exist only at the stem level, and not at the word or phrase levels: in Lexical Phonology, neither the existence of 'stratum-internal cyclicity,' to use Odden's (1993: 115) terminology, nor its confinement to the stem level followed from anything else. In contrast, the theory outlined above **explains** the existence of cyclic reapplication inside stem-level domains, and moreover correctly predicts that such reapplication affects only processes capable of sustaining lexical exceptions, as per Chung's Generalization. Nonetheless, by saying that internal cyclic reapplication effects within the stem level emerge from nonanalytic listing and blocking, we are in no way denying that such effects are genuinely cyclic. The essence of the phonological cycle is that the computation of the phonological properties of the parts precedes and feeds the computation of the phonological properties of the whole (Bermúdez-Otero 2011: §2), and this is precisely what happens in (26): the computation of the stem-level phonological representation of the adjective stem *original* precedes and feeds the computation of the stem-level phonological representation of the derived noun stem *originality*.[27] The key difference from the classical account

[26] A similar situation arises if a speaker whose lexicon already contains the words *órigin* and *oríginal* encounters the word *originality* for the first time in its standard orthographic representation, which does not indicate the location of stress: that speaker will need to rely on his morphological and phonological knowledge to compute the pronunciation of *originality*.

[27] This too is what justifies using the term *cyclicity* as a hyperonym of both *stratum-internal cyclicity* and *interstratal cyclicity* (Odden 1993: 115), even though the latter differs from the former in its properties (it

lies in two facts: (i) the output of the inner cycle is stored in the lexicon before the computation shown in (26) takes place, and (ii) the output of (26) is itself immediately entered into the lexicon, so that, thereafter, the cyclic relationship between the two stem-level representations holds off line. This account has nothing to do with output-output (OO) correspondence (e.g. Benua 1997), since the input to (26) consists of the stem-level representation of the constituents of *original-ity* and makes no reference to the surface properties of *original* in any environment.

This theory further predicts that, diachronically, cyclic reapplication effects like (26) spread by lexical diffusion. For example, suppose that, at a hypothetical historical stage in Early Modern English, the Abracadabra Rule applied normally: i.e. *órigin* ∼ *oríginal* ∼ ***ò**riginálity*. Suppose further that, at some later time, a new generation of speakers acquired exceptions to the initial dactyl effect, perhaps through some process of contact (Bermúdez-Otero 2007c: 514) or of restructuring, so that MAX-Head(Σ) was promoted above ALIGN(ω,L;Σ°,L) in the stem-level constraint hierarchy. In this scenario, speakers exposed to the conservative pronunciation *òriginálity* with an initial dactyl will simply retain it: recall that, by the hypothesis of nonanalytic listing, the initial dactyl will be stored in the noun's lexical entry. However, tokens of innovative *orìginálity* will gradually emerge by the mechanism described above: under the new stem-level ranking, suffixing *-ity* to *oríginal* on line yields *orìginál-ity*. Given enough time, cyclic reapplication will diffuse through the lexicon, in a process reminiscent of inflectional regularization (see note 20). Unfortunately, little appears to be known about the actual historical development of the Abracadabra Rule in Early Modern English. However, the prediction that cyclic reapplication inside stem-level domains spreads gradually by lexical diffusion is corroborated by the observation that, in present-day English, many stem-level derivatives retain noncyclic stress patterns, contrary to the expectations of standard Lexical Phonology. The next section (§2.3.3.3) discusses a classic example (recall also (16) above).

2.3.3.3 The role of token frequency: còmp[ə]nsátion *and* cònd [ɛ̀]nsátion In present-day English, cyclic stress preservation is notoriously irregular among stem-level derivatives containing pretonic sequences of two heavy syllables of which the second is closed by a sonorant consonant.[28] In such forms, the second syllable is consistently unstressed, and therefore surfaces with a reduced vowel, if the corresponding syllable is also unstressed in the base (Liberman and Prince 1977: 299, Pater 2000: 252): see (27a). If, in contrast, the base bears stress on the second syllable, this **can** be cyclically transferred to the derivative (27b)—but, crucially, it need not be.

is not bound by Chung's Generalization) and its origins (it does not emerge from nonanalytic listing and blocking). Bermúdez-Otero (forthcoming) connects the emergence of interstratal cyclicity in acquisition to a sequence of landmarks in the child's morphosyntactic development: see the last paragraph of §2.4.2.3 for a pointer to the basic facts and ideas.

[28] The literature on this phenomenon is truly enormous. Interesting contributions include Chomsky and Halle (1968: 38–9, 116, 161), Liberman and Prince (1977: 299–304), Kiparsky (1979: 428–9), Halle and Kenstowicz (1991: 460–1), Burzio (1994: §6.3), Pater (2000), Marvin (2002: 60–70), Hammond (2003), Collie (2007: ch. 2), and Kraska-Szlenk (2007: §8.1.2), among others.

(27) a. cómp[ə]nsàte còmp[ə]nsát-ion
cónt[ə]mplàte cònt[ə]mplát-ion
b. cond[ɛ́]mn cònd[ɛ̀]mn-átion
imp[ɔ́]rt ìmp[ɔ̀]rt-átion

If cyclic reapplication within stem-level domains emerges diachronically from non-analytic listing and blocking, as suggested in §2.3.3.2, then the cyclic effect in (27b) should comply with Chung's Generalization (22)—and so it does. Among monomorphemic words with the same configuration of syllables, the default pattern is clearly for the second syllable to be unstressed (28a), but exceptional items with peninitial pretonic stress do exist (28b): see Pater (2000: 250–1). Exceptions of this sort also include forms like (28c), which bear stress on the second syllable but are based on bound roots (Halle and Kenstowicz 1991: 460), for roots do not trigger cycles (Kiparsky 1982a: 144–5, 1982b: 32–3, Inkelas 1989: §3.5.5, Bermúdez-Otero 2007b: 283).[29] This evidence confirms that the stem-level constraint hierarchy protects input foot-heads over heavy syllables in the environment [σ̆̀__σ́…].

(28) a. Gòrg[ə]nzóla b. chìmp[ǽ]nzée
Pènns[ə]lvánia Mòz[ǽ]mbíque
c. ìnc[ǽ]nt-átion
òst[ɛ̀]nt-átion

The crucial point for our purposes lies in the fact, known since *SPE*, that many stem-based derivatives fail to undergo the cyclic transfer shown in (27b):

(29) a. cons[ɜ́]rve b. còns[ə]rv-átion
trànsp[ɔ́]rt trànsp[ə]rt-átion

The data in (29) pose a stiff challenge to theories such as *SPE*, Lexical Phonology, and Distributed Morphology, which stipulate stem-level internal cyclicity as an innate principle of Universal Grammar (§2.3.1). In such frameworks, the noncyclic stress pattern displayed by the items in (29b) can only be explained by claiming that these forms do not in fact synchronically derive from the stems in (29a), but are rather based on roots (which, as we saw above, do not define cyclic domains). For example, Chomsky and Halle (1968: 39, 112, 116) acknowledged the existence of *cònd*[ə]*nsátion* alongside expected *cònd*[ɛ̀]*nsátion* (cf. *cond*[ɛ́]*nse*), but they claimed that the two phonological variants reflected different syntactic and semantic structures: cf. (30a,b) and (30c,d), and see further Marvin (2002: 66–70).

(30) a. cònd[ɛ̀]nsátion ⟦$_N$⟦$_V$ condense⟧ ation⟧ 'act of condensing'
b. Andrew's skilful cond[ɛ̀]nsation of the argument into a few sentences helped me to see the point.
c. cònd[ə]nsátion ⟦$_N$⟦$_{\surd}$ condense⟧ ation⟧ 'condensed substance'
d. I used a cloth to wipe the cond[ə]nsation from the windscreen.

[29] If the minimal units of lexical storage are stems rather than roots, as argued by Bermúdez-Otero (2013), then the failure of roots to trigger cycles follows automatically.

Chomsky and Halle's examples in (30) clearly point towards the distinction between argument-structure nominals and referential nominals (Borer 2003: §4, after Grimshaw 1990; see further Alexiadou 2010: 500–1). If so, *SPE*'s empirical claim would be that derivatives like *condensation* display cyclic stress transfer when they are used as argument-structure nominals, as in (30b), but exhibit noncyclic stress patterns when used as referential nominals, as in (30d). This account works well for nouns like *information* (Chomsky and Halle 1968: 112, note 64), which unlike *condensation* is not ambiguous but can only be used as a referential nominal: *information* never inherits the argument structure of the verb *inf*[ɔ́]*rm*, and so it is correctly expected to be realized with noncyclic stress, i.e. *ìnf*[ə]*rmátion.*

(31) a. The butler suddenly informed the lady that her guests were in the drawing room.
 b. *The butler's sudden information of the lady that her guests were in the drawing room threw her into a panic.

Among nominals that are ambiguous between argument-structure and referential readings, however, the expected correlation between syntax and phonology breaks down in both ways, as Kiparsky (1979: 428–9) and Pater (2000: 261) observe, and as Marvin (2002: 69–70) concedes. In sentence (30d), for example, *condensation* is used as a referential nominal with the meaning 'condensed substance,' yet it is possible to pronounce it with a full vowel in the second syllable. Conversely, the noun *transportation* is clearly used as an argument-structure nominal in sentence (32), where it transparently inherits the argument structure of the verb *trànsp*[ɔ́]*rt*; yet the noncyclic stress pattern *trànsp*[ə]*rtátion*, with a schwa in the second syllable, is perfectly acceptable in this context, even in a slow, deliberate, and formal speech style.

(32) In *Noboa*, the plaintiffs argued that the airline's transp[ə]rtation of the human ashes in the valuable cargo section of the aircraft [...] was sufficient to justify a finding of wilful misconduct on the part of the airline.
(International Air Transport Association, *The Liability Reporter*, 9, February 2006)

Following Marantz's (1997b: 217–18) account of *Caesar's destruction of the city*, one might try to rescue *SPE*'s analysis and dispose of (32) by arguing that the meaning of the root $\sqrt{\text{Port}}$ entails an event with an external agent, that the identification of the possessive NP with this external agent is a matter of pragmatics, and that therefore *trànsp*[ə]*rtátion* does not synchronically derive from the verb *trànsp*[ɔ́]*rt*. Yet, by parity of reasoning, this argument would cancel *SPE*'s premise that, in (27b), *ìmp*[ɔ̀]*rtátion* derives from a verb stem. In fact, Marantz's (1997b) account has been argued to be false (see note 22 above and Borer 2003: §7). Thus, even though some instances of noncyclic stress do probably reflect the absence of an embedded stem in the synchronic grammar (31b), we still need an explanation for the failure of cyclic stress transfer in semantically transparent stem-based forms like the argument-structure nominal in (32).

Relative token frequency appears to be an important factor in these cases: noncyclic stress is reported to be common when the derivative is significantly more frequent than the base (Kraska-Szlenk 2007: §8.1.2).[30]

(33)

			$\times$ per 10^6 words in spoken section of COCA		
			base		derivative
a.	*cyclic stress*				
	cond[ɛ́]mn	cònd[ɛ̀]mn-átion	7.09	>	2.57
	imp[ɔ́]rt	ìmp[ɔ̀]rt-átion	5.15	>	0.62
b.	*variable stress*				
	cond[ɛ́]nse	cònd[ɛ̀∼ə]ns-átion	0.28	≈	0.22
c.	*noncyclic stress*				
	cons[ɜ́]rve	còns[ə]rv-átion	1.65	<	9.11
	trànsp[ɔ́]rt	trànsp[ə]rt-átion	7.23	<	23.54

In a careful statistical study, Collie (2007, 2008) finds a similar effect of relative token frequency on cyclic stress transfer among stem-level derivatives with trisyllabic pretonic sequences of the types σ̄σ̆σ̆σ́… and σ̄σ̄σ̆σ́…: e.g. *antìcipát-ion* ∼ *ànticipát-ion* (cf. *antícipate*).

The hypothesis that stem-level internal cyclicity emerges diachronically from nonanalytic listing and blocking can easily make sense of the role of relative token frequency. Let us suppose that, at some stage in history, *trànsp*[ɔ̀]*rt-átion* bore stress on its second syllable as a result of being derived on line from *trànsp*[ɔ́]*rt* under a stem-level constraint hierarchy that preserved input foot-heads in the environment [σ̆̀__σ́…].[31] Nonetheless, we may plausibly assume that a stressed syllable in this position has a relatively high chance of being misperceived as unstressed owing to a conspiracy of bottom-up and top-down factors. As shown in (34), the second syllable of *trànsp*[ɔ̀]*rtátion* stands surrounded by metrically stronger syllables; its vowel is therefore likely to be relatively shorter in duration, which will render it perceptually closer to schwa (see e.g. Steriade 2009: 174 and references therein).

(34)

```
              *
  *           *
  *.....*.....*
  *     *     *  *
transp[ɔ]rtation
```

[30] Hammond (2003) asserts that the probability of noncyclic stress grows in direct proportion to both derivative frequency and base frequency, but see Collie (2007: 177–86) for a refutation of this claim.

[31] The first attestation of *transportation* in the *OED* dates back to the last decade of Henry VIII's reign (in an Act of Parliament of 1540). I do not know of reliable evidence to determine the stress pattern of the noun at that time. If the pattern was originally noncyclic, the explanation given in this paragraph still goes through as an account of why *transportation* withstood the spread of the cyclic pattern better than *importation*. The logic remains the same.

In addition, the two-sided clash shown in (34) is a highly marked configuration, violating structural constraints that are ranked relatively highly in the grammar of English: indeed, as we saw in (28a), the default option for a syllable in the environment [σ̆̀__σ́...] is to be unstressed. Acoustic stimuli in danger of being heard as *trànsp*[ə]*rtátion* will accordingly not undergo top-down correction at the prelexical perception stage; if anything, the bias induced by markedness constraints on surface forms will run in the opposite direction (Boersma 2009b: 65–72). Thus, phonetic realizations of *trànsp*[ɔ̀]*rtátion* have a sizeable chance of being misperceived as *trànsp*[ə]*rtátion.* In turn, by nonanalytic listing, listeners' exposure to perceived tokens of *trànsp*[ə]*rtátion* will give rise to lexical entries lacking a prespecified foot-head on the second syllable: i.e. *trànsp/ə/rtátion.*[32] The outcome, then, is lexical variation between conservative *trànsp/ɔ̀/rtátion* and innovative *trànsp/ə/rtátion*. Moreover, this lexical variation is controlled by the balance of two opposing forces: whilst phonetic pressures reinforce *trànsp/ə/rtátion*, on-line derivation through the suffixing of *-ation* to *trànsp/ɔ́/rt* boosts *trànsp/ɔ̀/rtátion.*

It is at this point that the **relative** token frequencies of base and derivative become critical. Recall that, in a dual-route race model, the existence of a lexical entry for *trànsp/ə/rtátion* can block the on-line derivation of *trànsp*[ɔ̀]*rt-átion* from *trànsp/ɔ́/rt*; and, crucially, the probability that a stored lexical entry will block the on-line generation of a competing form depends upon the ease with which the entry itself can be retrieved, which in turn partially depends on its resting activation, which in turn partially depends on its token frequency. In consequence, high-frequency *trànsp/ə/rtátion* has a fair chance of being successfully retrieved before the processor can access its lower-frequency base *trànsp/ɔ́/rt* and run the rules to construct *trànsp*[ɔ̀]*rt-átion* (Hay 2003: 10–12, ch. 4). In the case of ˀ*ìmp/ə/rtátion*, the situation is precisely the opposite: its frequency, and so its resting activation, is much lower than that of its base *ìmp*[ɔ́]*rt*: see (33a). The overall result is that *trànsp/ə/rtátion* is good at blocking the generation of *trànsp*[ɔ̀]*rt-átion*, whereas ˀ*ìmp/ə/rtátion* is bad at blocking the generation of *ìmp*[ɔ̀]*rt-átion*. The trend affecting *transportation* reaches its limiting case in *information*: the fact that *information* can no longer inherit the argument structure of *inform*, as shown in (31), suggests that the noun can no longer be derived by suffixing *-ation* to *inf/ɔ́/rm* on line, and so, in the absence of tokens of **ìnf*[ɔ̀]*rm-átion* generated by rule, the bottom-up and top-down pressures favoring *ìnf/ə/rmátion* decide the outcome.

The overwhelming majority of derivatives with three pretonic light syllables do display cyclic stress transfer, as in *oríginal* ~ *orìginál-ity* and *imágine* ~ *imàgin-átion*; but a noncyclic stress pattern is also sporadically attested, *pace* Pater (2000: 261). As noted in Collie (2007: 147), for example, Wells (2000) reports several instances of variation like the following:

32 High-frequency derivatives like *transportation* may be expected to lead in this development, whilst low-frequency derivatives like *importation* are likely to lag, because lenition tends to be greater in high-frequency items (see Dinkin 2008 and references therein). This is important, but of course the story of cyclic vs. noncyclic stress cannot end there: although there is a statistically significant effect of the absolute frequency of the derivative, there are also significant effects of (i) the absolute frequency of the base and (ii) the frequency relation between base and derivative (Collie 2008: 513–17). See the examples in (33) again.

(35)

	a. *base*	b. *derivative: cyclic variant*	c. *derivative: noncyclic variant*
	dissímilate	dissìmilátion	dìssimilátion
	horrípilate	horrìpilátion	hòrripilátion
	illégible	illègibílity	ìllegibílity
	vatícinate	vatìcinátion	vàticinátion

Our approach to stem-level internal cyclicity offers a plausible explanation for the greater resilience of the cyclic pattern in derivatives of the *originality*-type than in those of the *transportation*-type. The second syllable of *orìginálity*, unlike that of *trànsp*[ɔ̀]*rtátion*, is surrounded by metrically weaker syllables, so its relative prominence is not perceptually compromised. As regards vowel duration, in particular, ω-initial syllables do tend to display some lengthening, but this is localized on initial consonants (Fougeron and Keating 1997, Turk and Shattuck-Hufnagel 2000). In addition, the stress contour of *orìginálity*, unlike that of *trànspòrtátion*, is perfectly alternating, and hence it is exempt from a top-down bias against clash. All these considerations suggest that *orìginálity* will have a relatively low chance of being misperceived as *òriginálity* compared with the likelihood of *trànsp*[ɔ̀]*rtátion* being heard as *trànsp*[ə]*rtátion*. Thus, metrical factors and their associated perceptual effects cause cyclic transfer to fail at different rates in derivatives with pretonic syllable sequences of different types.[33] Although these differences emerge diachronically in the way I have outlined, we may assume, following Zuraw (2010), that present-day English learners discover their existence by tracking lexical frequencies and encode the resulting knowledge in the synchronic grammar in terms of the ranking values of subordinate markedness constraints in the stem-level hierarchy (see the discussion of (24d) above).[34] More generally, the key to the prevalence of cyclic vs. noncyclic patterns of pretonic secondary stress among English stem-level derivatives lies, as in all cases of lexical diffusion, in the balance of bottom-up and top-down pressures on lexical entries (Bermúdez-Otero 2007c: 508–12); *pace* Pater (2000: 262), there is no need for lexically indexed constraints.

In sum, the theory of stem-level morphological and phonological rules applying in lexical redundancy mode explains a remarkable array of facts about cyclic effects inside stem-level domains: namely, (i) that such cyclic effects exist, (ii) that they go hand in hand with lexical exceptions as per Chung's Generalization, and (iii) that they diffuse through the lexicon at rates affected both by phonetic factors and by relative token frequency. As we noted above, this represents a considerable improvement on the position of Lexical Phonology, where stratum-internal cyclic reapplication was restricted to the stem level purely by stipulation. Moreover, Lexical Phonology had the same difficulties as *SPE* and Distributed Morphology in cases like (32), where syntax and phonology become partially decorrelated.

[33] The disparity between derivatives like *originality* and *transportation* is obvious. The difference between the types represented by *originality* and *anticipation* is less conspicuous, but becomes apparent under statistical analysis (Collie 2007: 148–9).

[34] It would be interesting to conjecture whether this in turn causes diachronic feedback effects, possibly through the interaction between markedness constraints and cue constraints in prelexical perception (Boersma 2009b).

The account presented here is also superior to one relying on constraint indexation (Pater 2000). First, it is not immediately apparent that one can deduce the link between relative token frequency and stem-level cyclicity from the axioms of indexation theory (cf. the cursory remarks on frequency in Pater 2000: 261); additional stipulations seem to be needed. Second, constraint indexation lends itself readily to applications that miss Chung's Generalization by severing the link between lexical exceptions and cyclic reapplication of stem-level processes: this happens, for example, if one uses indexed input-output (IO) faithfulness constraints for the former and indexed OO-identity constraints for the latter (Pater 2000: 254), for the two types of constraint may be ranked independently. We return to Pater's theory of constraint indexation below (§2.4.3).

2.3.4 Distributed associative memory

The theory of lexical redundancy rules surveyed in §2.3.3 breaks away from *SPE* only in part: nonanalytic listing entails a non-Bloomfieldian lexicon whose entries contain a great deal of redundant information, and lexical redundancy rules do not apply whenever they can, but act mostly as static representations of predictable relationships between items listed in full form; nonetheless, it remains the case that the grammar is held to encode these relationships explicitly by means of symbolic generalizations, in line with *SPE*'s assumption that all psychologically real patterns are expressed as rules in the broad sense (§2.3.2). As we saw in §2.3.3.1, moreover, Jackendoff (2010: 32) describes lexical redundancy rules as sharing the format of standard rules, crucially including typed variables (on which see below): indeed, the optimality-theoretic implementation sketched in §2.3.3.2 used a single constraint hierarchy to represent exceptionless well-formedness restrictions and exception-tolerating lexical redundancy rules simultaneously.

Since the late 1980s, however, there have been scholars, notably including Pinker (e.g. 1999), who have argued for a more drastic departure from *SPE* in the analysis of phenomena such as the sporadic extension of ablaut patterns among English verbs (recall the discussion of *snuck* in §2.3.2). Pinker proposes that, whereas regular past-tense inflection by means of /-d/ suffixation involves a genuine symbolic generalization, the morphological patterns that gave rise to *snuck*, though psychologically real, are nevertheless not mentally represented as rules, but are rather implicit in the connections between nodes in a distributed associative memory, as suggested by connectionist models of cognition (Rumelhart et al. 1986, especially Rumelhart and McClelland 1986).

The case for this synthesis of insights from connectionism (Rumelhart et al. 1986) and classical symbolic computation (e.g. Pylyshyn 1984) rests on two considerations. On the one hand, Fodor and Pylyshyn (1988: §3) make a strong argument that, without combinatorial symbol systems, one cannot account for key properties of higher cognition such as productivity, systematicity, compositionality, and inferential coherence. Pinker and Prince (1988, 1994) support this view with a detailed empirical critique of Rumelhart's and McClelland (1986) connectionist model of the English past tense. This line of research suggests that, at a minimum, the combinatorial nature of grammar requires typed variables, as well as a distinction between permanent lexical storage and short-term working memory (Marcus 1998, 2001). On the other hand, if properties like productivity and systematicity provide one's main argument for the existence of rules, one is *ipso facto* compelled to consider the possibility that rules may not be the right tool to describe patterns lacking those properties. Pinker therefore refuses to take it for granted that, for every psychologically real linguistic pattern, there must be a rule. Rather, he has sought to develop a set of empirical tests of regularity, particularly in the realm of inflection: only if an inflectional pattern meets these criteria, Pinker argues, can it be safely assumed to be mentally represented by a genuine symbolic generalization containing one or more typed variables. The checklists provided by Pinker (1999: 217–18, 222–4, and *passim*) and by Pinker and Ullman (2002: 458–62) include diagnostics such as application to very rare or phonotactically deviant items, to loanwords, to rootless items (e.g. names, onomatopoieas), to exocentric constructs, etc. By these criteria, the ablaut pattern extended to *sneak* ~ *snuck* is not regular, and this makes it a plausible case of pattern association. The case for a dual-route framework is further underpinned by Ullman's (2001) neurocognitive model of declarative vs. procedural memory (Pinker and Ullman 2002: 457).

We should note, however, that Albright and Hayes (2003) dispute the claim that pattern association provides an adequate account of the sporadic extension of irregular ablaut patterns like *sneak* ~ *snuck*. Albright and Hayes argue that analogical theories incorrectly predict that a novel irregular form may be created on the basis of a set of models displaying 'variegated similarity': i.e. the novel form resembles different members of the model set in different ways, but no single generalization holds across the entire set (though cf. Chandler 2010). Instead, Albright and Hayes retain *SPE*'s axiom that every psychologically real pattern is represented by a rule (§2.3.2), but reject *SPE*'s assumption that rules apply deterministically. According to their theory, learners construct probabilistic rule-systems that generate multiple competing outputs annotated with numerical confidence values; rule-building proceeds by minimal generalization, enabling the formulation of relatively narrow statements that exploit localized 'islands of reliability' (Albright 2002b). Important though these arguments and proposals are, however, I cannot engage with them in the confines of this chapter.

2.3.5 Refined dual-route models

Let us recapitulate. We have found it necessary to impose analytic biases on the interactions between phonology, morphology, and the lexicon in order to help the linguist solve the problem of analytic underdetermination, in order to explain how the child overcomes the logical problem of language acquisition, and in order that the theory of grammar should have empirical content (§2.2). However, *SPE*'s extreme bias against allomorph storage and in favor of derivation by phonological rule proved wrong and resulted in excessively remote underlying representations (§2.3.2). This suggests that it may be unwise to attempt to capture **all** psychologically real morphophonological patterns by means of a single mechanism, viz. classical (mandatory, deterministic) rules. Instead, the arguments developed in sections 2.3.3 and 2.3.4 may be taken to support a threefold taxonomy of patterns, each encoded in the grammar in a different way:

(36) *A refined dual-route model of morphophonology*

	Pattern type (with English examples)	*Grammatical encoding* (with properties)
(a)	Family resemblance between irregulars: e.g. strong-verb inflection (*string* ~ *strung*, *stick* ~ *stuck*, *sneak* ~ *snuck*)	Distributed associative memory • subsymbolic, implicit • nonanalytic listing • sporadic extension
(b)	Semiproductive pattern: e.g. stem-level derivational morphology (*divine* ~ *divinity*, *impress* ~ *impression*)	Lexical redundancy rules • symbolic, explicit • nonanalytic listing • structure-building only • leave gaps; when used generatively, the new outputs undergo nonanalytic listing
(c)	Productive pattern: e.g. regular weak-verb inflection (*play* ~ *played*, *talk* ~ *talked*, *load* ~ *loaded*)	Standard rules • symbolic, explicit • outputs can be unlisted or listed analytically • specifiable as structure-building or structure-changing • no gaps; fulfill Pinker's criteria for regularity

It is important to understand that these three types of morphophonological pattern arise from the interaction of just **two** basic mechanisms: nonanalytic listing and symbolic rules. **Three** types of pattern result, rather than just two, simply because lexical redundancy rules involve both explicit symbolic generalization and nonanalytic listing: recall (12) above.

The architecture outlined in (36) predicts that a subtle gradation of morphophonological processing effects will emerge in psycholinguistic experiments. Assume, for example, that the existence of appropriate symbolic rules in the grammar makes it possible for listeners to decompose complex forms.[35] Table (36) posits explicit symbolic generalizations both for productive constructions and for semiproductive patterns: *loaded* is related to *load* by means of word-level rules applying in standard mode, and *divinity* is related to *divine* by means of stem-level rules applying in lexical redundancy mode. Consequently, effects of decomposition (e.g. priming of the base) should be observed in both cases: robustly with productive patterns, and more weakly with semiproductive ones. In contrast, let us suppose that having a separate lexical entry enables a form to display surface frequency effects: i.e.

[35] This is a particularly natural assumption in a bidirectional framework like Boersma's (2009a).

effects caused by the form's own token frequency, rather than by the frequency of its base (Baayen et al. 2002: 62–3; though cf. Clahsen and Neubauer 2010: §6.1). If so, then items of types (36a) and (36b) should display strong and systematic surface frequency effects (e.g. on response latencies) because both are subject to nonanalytic listing. However, weaker surface frequency effects are also predicted for fully regular forms that happen to be listed analytically (Stemberger and MacWhinney 1986, 1988); indeed, such effects should be absent only in the case of unlisted regular forms. There is some suggestive empirical evidence in favor of such a gradation of processing effects. Notably, Clahsen, Sonnenstuhl, and Blevins (2003) report that native speakers of German process derivatives containing the suffixes *-ung*, *-lein*, and *-chen* in ways which match neither irregular inflection nor regular (default) inflection: on the one hand, these derivatives pattern like the outputs of default inflection by fully priming their bases; but, on the other hand, they behave like irregularly inflected wordforms in unprimed lexical decision tasks, where they display similar surface frequency effects. These findings suggest that it may be necessary to posit a more elaborate architecture than Jackendoff (2002b: 158ff.), who conflates (36a) and (36b); the tripartite taxonomy in (36) bears a greater likeness to 'refined' dual-route models, as advocated in Clahsen et al. (2003: 127, 149).

More generally, considerations of elegance and simplicity may be thought to favor single-mechanism frameworks like *SPE*'s, but one may argue that such arguments do not carry much conviction against the weight of the empirical evidence. Jackendoff (2002b: 160–2) speaks persuasively for "the necessity of a heterogeneous theory". Pinker (2002: xii) draws an interesting analogy with biological systems, which often combine two mechanisms with different costs and benefits: e.g. slow-twitch muscle fibers (which contract slowly but are very resistant to fatigue) and fast-twitch muscle fibers (which contract fast but fatigue easily). Another term of comparison may be sought in Labov and Kiparsky's resolution of the neogrammarian controversy: this required the empirical recognition that both neogrammarian change and lexical diffusion exist, and the theoretical realization that each reflects innovation in a different component of the grammar (Bermúdez-Otero 2007c: 501ff.). Indeed, the refined dual-mechanism framework outlined in (36) amounts to much more than an *ad hoc* strategy to save the phenomena; it has a rich deductive structure of its own. For example, the hypothesis that stem-level forms are stored nonanalytically (§2.3.3.1) correctly predicts that there will be cyclic reapplication effects within stem-level domains and that these cyclic effects will obey Chung's Generalization (§2.3.3.2); and, when embedded in a parallel race model of processing, the hypothesis further predicts that the incidence of stem-level cyclic reapplication will vary with token frequency (§2.3.3.3).

In the next section we move on to address the division of labor between morphology and phonology: we shall ask to what extent morphology should be allowed to manipulate phonological material and how much morphosyntactic information should be available to phonology, and we shall propose restrictive answers to these questions based on cognitive principles of modularity and locality. Throughout this discussion, however, the reader should bear in mind the implications of these proposals for lexical storage: do they favor derivationally proximate or derivationally remote underlying representations?;[36] do they converge with the results of refined dual-route models like (36)? At least in some cases, the fit will prove satisfyingly snug: in §2.4.3, for example, I shall argue that considerations of modularity speak against the analysis of present-day English strong-verb ablaut by means of morphologically triggered phonological processes ('readjustment rules'), converging with the conclusions of sections 2.3.2 and 2.3.4.

2.4 Morphology vs. phonology

The notions of modularity and locality are widely regarded as forming the conceptual bedrock on which all broadly generative approaches to linguistic interfaces should be built (see e.g. Scheer 2010). In section 2.4.1 I introduce these ideas and outline the ways in which they can help us to tackle the difficult challenges posed by analytic underdetermination (§2.2). Subsequent sections implement a modular and local program for the morphology-phonology interface by means of four specific hypotheses (37); I shall refer to this as 'the Four-Hypothesis Program'.

(37) *The Four-Hypothesis Program*

a. According to the **Morph Integrity Hypothesis** (41), the representational currency of morphology is the morph: morphology is not allowed to operate directly upon elements of phonological representation such as features, segments, nodes, or association lines (§2.4.2).

b. In section 2.4.3 I adapt Inkelas's (1989[1990: 10ff]) strong formulation of the **Indirect Reference Hypothesis** to an optimality-theoretic framework (71): in this version, Indirect Reference prevents phonological constraints other than those on prosodic alignment from referring to morphosyntactic information.

c. The **Phonetic Interpretability Hypothesis** (76) asserts that derived phonological representations must be phonetically interpretable. This forbids the presence of diacritics of morphosyntactic affiliation in phonological output representations (§2.4.4).

d. Finally, in line with the **Cycle Hypothesis** I assume that certain morphosyntactic constituents define domains over which the phonology applies iteratively, starting with the most deeply embedded domains and moving progressively outwards (see e.g. Bermúdez-Otero 2011). Alone and in combination with Phonetic Interpretability, this assumption imposes locality restrictions on the way in which phonology can refer to morphosyntactic structure during a cycle, both outwardly and inwardly (§2.4.4).

This is an austere prescription: it decrees that all morphology must be concatenative (§2.4.2.2), it bans the use of indexed constraints and of readjustment rules (§2.4.3), and it undermines the rationale for construction-specific cophonologies (§2.4.2.3).

[36] Recall that, in the usage I have adopted in this chapter, 'derivationally remote' does not mean the same as 'abstract': see again note 9.

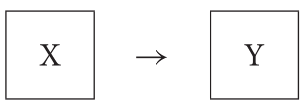
Of course, only a great deal of time and effort will allow us to find out whether these restrictions can survive sustained empirical scrutiny. Here I can do no more than sketch the empirical issues: in particular, I will show how phenomena such as reduplication, widely thought to require less parsimonious accounts, can in fact be analyzed within the limits of this program along the lines insightfully drawn by Saba Kirchner (2010) (§2.4.2.2); and I will illustrate the program's heuristic power by showing how its strictures force us to search more deeply when confronted with superficial counterevidence (§2.4.2.3).

Because this volume focuses on themes that cut across the boundary between morphology and phonology, I shall devote the lion's share of my exposition to the Morph Integrity Hypothesis, which bears on the classic debate between 'item-and-process' and 'item-and-arrangement' morphology (Hockett 1954) and which has momentous consequences for the way in which morphology and phonology share the burden of description in instances of apparently nonconcatenative exponence. Indirect Reference and Phonetic Interpretability have far less impact outside of phonology and phonetics, and so the sections dealing with these principles will consist of extremely succinct summaries of the issues with a few pointers to the literature. Cyclicity is, of course, a vast subject: for example, the question whether syntax, morphology, and phonology do run on the same cycles as each other is in very urgent need of addressing, but far exceeds the scope of this piece (nor is it settled by Bye and Svenonius in this volume); for specifically phonological arguments for the cycle, see Bermúdez-Otero (2011). Yet it would be impossible to discuss Morph Integrity without some consideration of Indirect Reference, Phonetic Interpretability, and the cycle. Indeed, throughout the following discussion I shall from time to time remark upon the close solidarity between the components of the Four-Hypothesis Program: we shall see, for example, that the cycle can crucially underpin analyses that uphold Morph Integrity (§2.4.2.2), that Indirect Reference cannot keep morphosyntactic information out of phonology without the help of Morph Integrity (§2.4.2.3), and that inward locality restrictions cannot be enforced by the cycle alone but require Phonetic Interpretability (§2.4.4).

2.4.1 Modularity and locality at the morphology–phonology interface

Most linguists working in the broad tradition of classical symbolic computation subscribe to the assertion that the architecture of grammar is modular. This statement expresses a relatively simple intuition: it is the idea that the grammar consists of a number of separate components, that each component works with its own set of representations encoding a particular aspect of linguistic structure, and that the flow of information from one component to another is restricted by relatively narrow channels of communication called 'interfaces.' This idea is often depicted by means of box-diagrams like the following:

(38) *A modular serial interface*



Here, the boxes evoke the modular character of the interaction; the single-headed arrow, its serial (i.e. functionally asymmetric) character.[37] A diagram like (38) is usually meant to convey a number of specific claims (cf. Jackendoff 1997: 24 for a somewhat different statement):

(39) a. Each of the two modules *X* and *Y* possesses its own representational vocabulary.

b. The computations performed in *Y* have no effect upon those carried out in *X*.

c. The computations performed in *X* have an effect upon those carried out in *Y*.

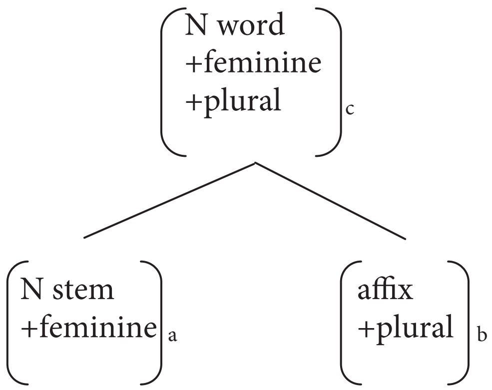
d. *X* affects *Y* in the following manner. The output of *X* is displayed at the interface with *Y*. A specific subset of the elements present in the output of *X* is put in correlation with elements in the input to *Y* by means of lexical look-up, realization statements, or mapping rules. The computations performed in *Y* can refer to the correlations so established. However, *Y* does not have access to the internal operations of *X*, or to elements in the output of *X* that do not enter into correlation with elements in the input of *Y*.

The mapping of syntactic structure onto prosodic categories in mainstream versions of OT incorporating alignment constraints (McCarthy and Prince 1993a) and the correspondence theory of faithfulness (McCarthy and Prince 1995a) can easily be understood as an instance of (39). In this case, module *X* is the syntax;[38] module *Y*, the phonology. Let us use the terms 'underlying representation' and 'surface representation' to refer to the input and the output of the phonology, respectively, and ignore cyclicity and stratification for the time being. The underlying representation consists of an assembly of phonological pieces, each of which stands in a relationship of exponence with some node in the output of the syntax; in section 2.4.2.1 below I address the way in which such exponence relationships are created. Now suppose that a certain piece in the underlying representation—say, the segment string /tike/—is the exponent of a stem node in the syntactic output. If so, the alignment constraint ALIGN(stem,R;ω,R) will in effect require that the rightmost surface element standing in correspondence with some component of the underlying sequence /tike/ should be final in some prosodic word. This is illustrated in (40), where I use subscript Roman letters to notate exponence, and subscript integers to notate the correspondence between underlying and surface representations inside the phonology.

[37] Of course, interfaces may be parallel (i.e. functionally symmetric) rather than serial. In that case the arrow would be double-headed.

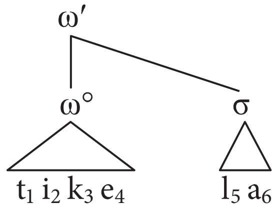
[38] Note that, throughout §2.4, I use the term 'syntax' to include both 'word syntax' and 'phrase syntax.' Of course, postulating a 'word syntax' is perfectly compatible with upholding the atomicity of grammatical words in the phrasal syntax along with all the other clauses of the Lexicalist Hypothesis (Williams 2007: 356): for theories of word structure combining both properties, see e.g. Selkirk (1982) or Ackema and Neeleman (2004, 2007). I assume a grammatical framework of this general type, although the specific details will not be relevant here. The assumption that words have internal constituent structure need not be interpreted as an endorsement of Hale and Keyser's (1992) syntactic theory of lexical semantics (cf. e.g. Kiparsky 1997).

(40) a. *syntactic output*



b. *underlying representation* $[\![_c [\![_a\ t_1 i_2 k_3 e_4]\!] [\![_b\ l_5 a_6]\!]]\!]$

c. *surface representation*



This architecture enables us to impose nontrivial modularity restrictions on phonology. Let us consider four obvious ones.

First, the subscript Roman letters in (40b) link phonological pieces in the underlying representation with nodes in the syntactic **output**, but encode no information about the computational background of the latter. In consequence, the phonology receives no information about more remote levels of representation: e.g. lexical semantics.

Second, phonological constraints cannot alter the relationships of exponence specified in the underlying representation; these are established extraphonologically by the morphology, ultimately on the basis of lexical information (see §2.4.2.1 for details). In the optimality-theoretic literature, this principle has come to be known as 'Consistency of Exponence' (McCarthy and Prince 1993b: §2.3). Van Oostendorp (2007: 129–35) notes and refutes a couple of proposals to turn Consistency of Exponence into a violable requirement. In the architecture illustrated by (40), the phonology cannot possibly infringe Consistency of Exponence, for exponence relationships are defined to hold between the syntactic output and the **underlying** representation. By the same token, output candidates engage in correspondence with the underlying form, but themselves bear no information about morphosyntactic affiliation: there are no subscript Roman letters in (40c); see further section 2.4.2.1 and 2.4.4. It is therefore simply impossible for Gen to create, or for Eval to select, a surface representation that violates Consistency of Exponence.

Third, the theory of phonological derivations provides further means to limit the use to which the phonology puts the information about exponence relationships contained in the underlying representation. In an optimality-theoretic framework, for example, the power to refer to this information may naturally be confined to a particular family of constraints: namely, constraints like ALIGN(stem,R;ω,R) above, which align the edges of prosodic categories with those of the exponents of syntactic nodes. This limitation is, in effect, nothing other than the optimality-theoretic version of the well-known Indirect Reference Hypothesis (see §2.4.3 below). A suitably stringent version of the hypothesis (71) will go on to impose restrictions on prosodic alignment constraints themselves, for example by allowing them to mention the category membership of a syntactic node (e.g. 'stem', 'word', 'N') but not its feature content (e.g. [+feminine], [+plural]). Furthermore, I assume that lexical indexation is absolutely banned for **all** constraints: thus, alignment constraints may refer to the category membership of the node exposed by a particular piece, but not to the lexical identity of the piece itself (cf. McCarthy and Prince 1993a; see Bye and Svenonius this volume).

Fourth, the surface form is the representation displayed at the interface of phonology with the phonetic implementation module; phonetics cannot peer through the surface representation into the inner workings of the phonology. Therefore, since the surface form bears no record of exponence relationships, it automatically follows that phonetics may be sensitive to prosody, but not to morphosyntax (Bermúdez-Otero 2010; cf. Kawahara 2011: §2.3.3).

The advantages of such modularity restrictions should be obvious in the light of the discussion in section 2.2 above. When confronted with a morphologically conditioned phonological alternation, modular frameworks rule out large sets of logically conceivable analyses for the simple reason that those analyses require modules to communicate in ways which are not permitted. Under (40), for example, one must reject descriptions in which the phonology directly refers to semantic or deep syntactic information, in which the phonology alters a morpholexically determined relationship of exponence, in which constraints other than those on prosodic alignment refer to the morphosyntactic affiliation of underlying phonological pieces, or in which the phonetics is directly sensitive to morphosyntax. This provides for a contentful theory of grammar: one is not allowed to rescue a cherished conjecture by reanalyzing a troublesome counterexample in a forbidden way. It also equips the learner with a useful set of strong priors (in the Bayesian sense). Moreover, modular frameworks are themselves amenable to empirical evaluation: a modular framework fosters a progressive research program if most apparent counterexamples are eventually resolved, and particularly if they are resolved in ways which reveal previously unnoticed structure in the data; conversely, a modular framework leads to a degenerating research program if it accumulates unresolved problems or solves them only by weakening its empirical content (Lakatos 1970).

In this sense, the debate on the type of modularity appropriate to natural language in general, and to morphology–phonology interactions in particular, can ultimately be settled on empirical grounds. My approach has so far been rather more Jackendovian (e.g. Jackendoff 1997: §2.6, 2002b: §7.5) than Fodorian (e.g. Fodor 1983). I am persuaded by two of Jackendoff's arguments against the applicability of Fodor's concept of 'module' in linguistics. First, and most importantly, mappings across interfaces within the grammar itself seem far too complex to be handled by computationally trivial mechanisms akin to "transducers" (Fodor 1983: 41). Second, the properties that Fodor attributes to modules do not appear to cluster consistently together: e.g. overlearned abilities like driving can be highly automatic but are obviously not innate. Indeed, I am in principle sympathetic to the idea of emergent modularity (see §2.4.5, epecially note 60). However, see Scheer (2010) for a more stringently Fodorian take on the morphosyntax–phonology interface.

Let us now briefly turn to the concept of locality. To posit an analytic bias towards locality is to favor linguistic generalizations holding over relatively small domains or between elements standing relatively close to one another within a representation. The concept of locality plays a familiar role in accounts of sentence comprehension: for example, processing difficulty grows in proportion with the distance between an antecedent and a gap (partly, though not solely, because of the size of the structure that must be held in working memory): see e.g. Gibson (1998) and Lewis, Vasishth, and van Dyke (2006). This effect has been argued to provide the grounds for important restrictions on syntactic dependencies, such as subjacency: e.g. Weinberg (1988); see further Sprouse, Wagers, and Phillips (2012). In the current context, however, the benefits of locality accrue in a different way. If we say that morphological and phonological generalizations are constrained to hold over domains of certain sizes, we *ipso facto* rule out entire classes of hypotheses about morphology–phonology interactions: we are in effect saying that no information outside the appropriate domain can be relevant to the statement of an alternation. In this sense, locality principles create an *a priori* distinction between those factors that may be relevant to an alternation and those that may not. Locality restrictions thus alleviate the problem of analytical underdetermination (for the linguist) and the logical problem of language acquisition (for the child), and are a source of empirical predictions in linguistic theory.[39]

In the generative tradition, locality restrictions have typically been enforced through the cycle—although the causes and nature of the phonological cycle can be understood in radically different ways, as we saw in section 2.3.3.2 and 2.3.3.3 (see also the last two paragraphs in 2.4.2.3). I have recently laid out the empirical arguments for a cyclic approach to morphosyntactically conditioned phonology elsewhere (Bermúdez-Otero 2011), and so I shall have little to say about this topic here, except for two points. In section 2.4.2.2, I revise Saba Kirchner's (2010) mora-affixation analysis of stem expansion before -M̓U:T in Kwak'wala, assigning a more prominent role to cyclicity; this will show how the cycle can play an important role in upholding modular restrictions on the morphology-phonology interface. In section 2.4.4, conversely, I show how the Phonetic Interpretability Hypothesis (76), which is primarily a modularity constraint on the flow of information between morphosyntax, phonology, and phonetics, has the concomitant effect of tightening up the locality restrictions associated with the cycle, yielding a form of 'Bracket Erasure' (Orgun and Inkelas 2002). These connections highlight the coherence of the Four-Hypothesis Program as a unified conception of the morphosyntax–phonology interface.

[39] For related discussion of the broad concept of locality in cognitive science in general, see Fodor (e.g. 1983, 2008: ch. 4), who links locality to the notorious 'frame problem' in Artificial Intelligence (1983: 112ff., 2008: 116ff.).

2.4.2 Morph Integrity and the problem of apparently nonconcatenative morphology

2.4.2.1 Morphs as insects in amber Implementing modularity restrictions at an interface like (38) is futile unless one reins in the power of both modules simultaneously. To be strict about module *X* and lax about module *Y* amounts in practice to turning *Y* into a waste bin for all the problems encountered by one's theory of *X*, and vice versa: in either case, empirical content evaporates (see the discussion of readjustment rules in §2.4.3 below for an example). This means that modularity restrictions must bind morphology just as tightly as they do phonology.

Yet, initially at least, applying the notion of representational modularity to morphology looks like a tall order, for, as a first approximation, the business of morphology seems to be to deploy the resources of the lexicon for the purposes of linking up a syntactic and a phonological expression within an exponence relationship. In this view, the lexicon stores pairings of semantic–syntactic and phonological property bundles, some of them irreducibly arbitrary (i.e. Saussurean), some of them (at least in a non-Bloomfieldian lexicon) partly or wholly compositionally predictable (§2.3.2, §2.3.3). Drawing upon such lexical pairings, morphology connects the output representations generated by the syntax with the underlying forms in the input to the phonology: see (40a) and (40b) above. Morphological operations thus seem to range over the proprietary representational vocabulary of other modules. How, then, can morphology itself be modular?

There is no getting away from the fact that morphology does read syntactic and phonological representations simultaneously; in this sense, it may be regarded as a prime example of what Jackendoff (1997: ch. 2, 2002b: ch. 7) calls an 'interface processor' or a 'bi-domain-specific module'. But this acknowledgement need not turn morphology into the place where all empirical content leaks out of a modular conception of grammar. An obvious response to the facts is to assert that the domain of morphology is the exponence relationship itself as encapsulated in the **morph**, understood as an integral piece of phonological material specified with instructions for its use as an exponent of syntactic properties. From this perspective, then, the morph constitutes the representational currency of morphology: morphological operations will display the maximum degree of modularity compatible with their function in a comprehensive architecture of grammar if they treat morphs as inalterable units, and never change their syntactic specifications or their phonological content; in this overall conception of grammar, only syntax manipulates syntactic features, and only phonology manipulates phonological features. This, I believe, is the conceptual rationale for strictly concatenative approaches to morphology (e.g. Stonham 1994); this is also, I believe, the reason why Jackendoff adopts a piece-based morphology (see note 21 above).

I therefore propose the following hypothesis:

(41) *Morph Integrity Hypothesis*
Morphological operations do not alter the syntactic specifications or phonological content of morphs.

Although literary devices are always dangerous, the implications of (41) for morphology–phonology interactions may perhaps be best introduced by means of a

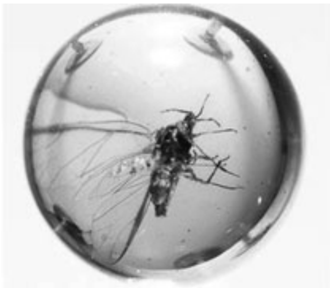


FIGURE 2.1 An insect in amber

Source: (http://www.rothamsted.ac.uk/Photos/AmberAphid.jpg)

little allegory. Under the Morph Integrity Hypothesis, then, a morph is like a translucent droplet of amber encasing a fossilized insect, the phonological content of the morph being like the body of the insect itself, and the morphology like the laboratory of an entomologist working with a collection of such specimens (see Figure 2.1).

In the course of his enquiries, the entomologist may browse through the compartment labels of his storage cabinet, pick up a particular amber globule, and place it under a magnifying glass to study the insect's body through the diaphanous resin. Upon inspection, the entomologist may decide to place the specimen in a particular spot within a working array, perhaps attaching a temporary note about work to be done. The fossilized organism is thus no less visible to the entomologist than were the labels on his storage cabinet, but the insect's resinous casing shields it from more invasive manipulation: the entomologist cannot detach an antenna, or attach a replica of a missing leg, or treat a wing with a fluorescent dye, without damaging the amber and so spoiling the precious object.[40]

Under (41), the morphology deploys analogous powers. Consider it, for example, in the task of assigning an exponent to a node in a syntactic output representation like (40a). For the sake of simplicity, let us focus on a terminal syntactic node in an initial phonological cycle. First, the morphology will search long-term memory (i.e. the lexicon) for morphs whose syntactic specifications match those of the exponendum; in a noninitial cycle triggered by a nonterminal node, the search would also include short-term memory, which would contain exponents of lower nodes generated by earlier on-line cycles (cf. §2.3.3.2 for off-line cyclicity). If a potential exponent happens to bear a syntactic subcategorization requirement, the morphology may need to scan the feature content of neighboring syntactic nodes; if a potential exponent happens to bear a phonological subcategorization requirement, it will have to scan the feature content of neighboring phonological pieces in the **underlying** representation;[41] in either case, the amount of structure available for inspection will be limited to an

[40] Until recently, attempts to extract insect inclusions from amber frequently resulted in the destruction of the fossil, but Azar (1997) reports success with a technique involving immersion in a chloroform bath (Green 2001: 238–9). In our allegory we will say that only the phonology, and not the morphology, has a supply of chloroform.

[41] As emphasized in Paster (2006: 14 and *passim*), a key characteristic of phonological subcategorization statements is that they refer to properties of phonological **input** representations; see also Bye (2007).

appropriate local domain, presumably defined by the cycle (see §2.4.4 below, and Bye and Svenonius this volume). It may so happen that the lexicon delivers two or more morphs that fit the bill. If so, several outcomes may ensue: the morphs may have different syntactic specifications, in which case the morphology may choose the most appropriate by reference to some principle like the Elsewhere Condition (e.g. Albright and Fuß this volume); or the morphs may be in free lexical variation with each other, in which case the morphology stochastically selects one or the other with a certain probability; or the morphs may arrive with an instruction that the choice should be passed on to the phonology, in which case both are inserted at the same point in the underlying representation and the disjunction is resolved in the output.[42] Moreover, in addition to selecting a suitable morph and inserting it at the right place in the underlying representation,[43] the morphology must annotate the latter with such information about exponence as the phonology is permitted to access: see the discussion of (40) in section 2.4.1. At a minimum, this will involve coindexing each phonological piece with the syntactic node it exposes; this was the role of the subscript Roman letters in (40a) and (40b). I suggested above that such coindexation is present in the phonological input only, and is only referred to by alignment constraints. We should also observe that coindexation is a property of the morph as a whole, rather than of any phonological feature, segment, node, or association line contained within it: pursuing our little allegory, it is like a sticky note temporarily attached to the surface of an amber droplet during one of the entomologist's work sessions (see §2.4.4 for further discussion). In a stratal-cyclic theory, moreover, the morphology will also use information from syntactic constituency and, in certain cases, lexical specifications of individual morphs to choose the level of the phonology to which the entire assembly of phonological pieces in a cyclic domain should be submitted: see e.g. Bermúdez-Otero (2007b: 283) and the paragraph preceding (62) below.

The metaphor of insects in amber may also help us to understand what the morphology **cannot** do to morphs in this view of the morphology–phonology interface. First and foremost, it cannot insert, delete, or otherwise alter any phonological feature, segment, node, or association line belonging to a morph: only phonology manipulates phonological material. This vastly reduces analytic underdetermination (§2.2): compare the analyst's, and by implication the learner's, predicament in a framework like Anderson's (1992) Amorphous Morphology, where, for any transformation possibly effected by a phonological rule in a phonologically defined environment, one must also reckon with the possibility that exactly the same transformation is carried out by a word-formation rule in a morphologically defined environment. By the same token,

[42] The availability of disjunctive inputs is exploited in optimality-theoretic accounts of phonologically conditioned allomorph selection by **optimization** (e.g. Tranel 1996, 1998, Kager 1996, 2009: 420ff., Mascaró 1996, 2007, Rubach and Booij 2001). This contrasts with phonologically conditioned allomorph selection by **subcategorization**, which refers to the phonological input (note 41). Lapointe (2001) and Nevins (2011b) argue that both devices are in fact needed in different cases; I agree.

[43] The linear ordering of morphs lies beyond the remit of this chapter. Here I shall merely assume that ordering relationships are fully specified by the time phonological inputs have been assembled (Paster 2009, though cf. Kim 2010), but I shall not discuss the relative roles of syntax and morphology in bringing this about.

under (41) morphology is bound by its own version of Consistency of Exponence (see §2.4.1 for Consistency of Exponence in phonology). When the morphology retrieves a piece from long-term or short-term memory, it can coindex it as a unit with a syntactic node; it may possibly coindex a syntactic node with more than one piece (as in cases of 'multiple' and 'discontinuous' exponence), and it may coindex a piece with more than one syntactic node (as in cases of 'cumulative' exponence); but it cannot reaffiliate a phonological feature, segment, node, or association line from one piece to another, nor can it designate a phonological feature, segment, node, or association line as being outside the exponence relationship in which the whole piece participates.

Finally, if the representational currency of the morphology is the morph and morphology cannot directly manipulate either syntactic or phonological features, then interesting consequences follow for the syntax–morphology interface as well. For example, a process of 'impoverishment' enforcing a systematic syncretism (see e.g. Albright and Fuß this volume) can no longer be literally formulated as a rule deleting syntactic features from syntactic output representations, but must rather be stated as a morphological constraint forbidding the use of exponents specified for a certain syntactic feature in the context of certain other features. This is precisely the line taken in Wunderlich and Fabri's (1995) Minimalist Morphology and Trommer's (2001: e.g. 113–14) Distributed Optimality.

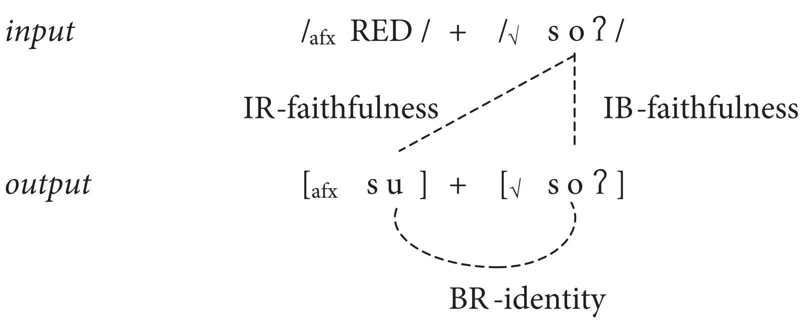
2.4.2.2 Reduplication as prosodic node affixation Without a doubt, the most striking and controversial prediction of the Morph Integrity Hypothesis (41) is that all morphology is concatenative: any instance of apparently nonconcatenative exponence must reduce to lexical allomorphy, morph concatenation, phonological derivation, or some combination thereof. The effort to live up to these strictures has generated some very interesting and encouraging results, coalescing into a line of research that seeks to reduce the role of morphology in all instances of apparently nonconcatenative exponence to the insertion of pieces of nonlinear phonological representation whose existence is independently motivated: e.g. floating features or feature-geometric treelets in the case of mutation, fully or partially bare prosodic nodes or prosodic treelets in the case of reduplication and subtraction. This approach, which on the model of Trommer and Zimmermann (2010) I shall call **Generalized Nonlinear Affixation**, was pioneered by Lieber (1992: ch. 5) and, more extensively, by Stonham (1994): see Trommer (2011) for detailed discussion. For applications to mutation, see the references in the paragraph introducing (5) above; on reduplication, see Saba Kirchner (2010); on subtraction, see Seiler (2008), Trommer and Zimmermann (2010), and Trommer (2011); and, for an even wider range of applications, see Bye and Svenonius (this volume). In this section I show that the Four-Hypothesis Program (37) favors Generalized Nonlinear Affixation above all other approaches to apparently nonconcatenative exponence, and I briefly present some striking results in the field of reduplication.

It is fairly uncontroversial at present to say that the role of morphology in reduplication is limited to concatenating pieces: this assertion is upheld by the two most influential approaches to reduplication currently being pursued. In McCarthy and Prince's (1995a, 1999) theory of phonological copying, for example, the morphology does no more than insert an underlyingly empty reduplicative morph (RED); this is then supplied with phonological content on the surface in order to satisfy constraints on base–reduplicant correspondence, which require that the output realization of RED should be identical with that of the root or stem (cf. Inkelas this volume: §10.5 for more details). Inkelas and Zoll's (2005) theory of morphological doubling assigns a completely different role to morphology, but still one that is purely concatenative: reduplicative constructions, in this view, involve the morphological combination of two tokens of an already existing piece, which may be a root, an affix, or a complex stem or word; modifications of the phonological content of these two pieces are effected by the phonology, during cycles triggered over each token and over the whole construction. These two approaches to reduplication thus comply with the Morph Integrity Hypothesis; yet, surprisingly, both prove unsuitable for our purposes because they fail to comply with the requirements of the Four-Hypothesis Program in other ways.

McCarthy and Prince's (1995a, 1999) theory violates Indirect Reference (§2.4.3) because it requires phonological constraints other than those on prosodic alignment to be indexed to particular morphological constituents, and it also offends against Phonetic Interpretability (§2.4.4) because it requires phonological output representations to be annotated with diacritics of morphological affiliation. Take, for example, McCarthy and Prince's (1995a: §4.2) account of reduplication in Akan (Niger-Congo, Kwa; Ghana), where the vowel of the reduplicant matches that of the root for the features [ATR] and [back], but is obligatorily [+high].

(42)	si-siʔ	'stand'	sʊ-sʊʔ	'carry on the head'
	si-seʔ	'say'	su-soʔ	'seize'
	sɪ-sɛʔ	'resemble'	sʊ-sɔʔ	'light'

McCarthy and Prince's so-called 'full model' posits three parallel relationships of correspondence: IB-faithfulness holds between the input and output representations of the root; BR-identity relates the surface realization of the reduplicative affix to the surface realization of the root; and, finally, IR-faithfulness holds between the input representation of the root and the output realization of the reduplicative affix.

(43)


In McCarthy and Prince's analysis, vowels undergo raising in the reduplicant but not in the root because the markedness constraint against nonhigh vowels is ranked below IB-faithfulness but above BR-identity and IR-faithfulness.

(44)

/$_{afx}$ RED/ + /$_{\surd}$ s_1 o_2 $ʔ_3$ /		IB-IDENT[high]	*[-high]	BR-IDENT[high]	IR-IDENT[high]
a.	[[$_{afx}$ s_1 u_2][$_{\surd}$ s_1 u_2 $ʔ_3$]]	*!			*
b.	[[$_{afx}$ s_1 o_2][$_{\surd}$ s_1 u_2 $ʔ_3$]]	*!	*	*	
c.	[[$_{afx}$ s_1 o_2][$_{\surd}$ s_1 o_2 $ʔ_3$]]		**!		
d.	[[$_{afx}$ s_1 u_2][$_{\surd}$ s_1 o_2 $ʔ_3$]] ☞		*	*	*

Crucially, this account only works if, as shown in (43) and (44d), the winning output candidate [susoʔ] bears morphological annotations indicating which portion realizes the reduplicative affix and which realizes the root. Merely coindexing input segments with their output correspondents is not enough: input /o/ has two output correspondents, but only one is subject to high-ranking IB-faithfulness; EVAL must be allowed to know which is which, so that it may penalize candidates like *[sosuʔ]. Thus, McCarthy and Prince's theory of reduplication entails that surface representations contain information about morphological affiliation. Yet, as we saw in the discussion of (40) above, this is problematic for at least two reasons: a theory with such labels affords the means to violate Consistency of Exponence in principle (even if it chooses not to do so in practice), and it is able to transmit morphological information to phonetic implementation.

The problem with Inkelas and Zoll's (2005) theory of morphological doubling (cf. also Inkelas this volume: §10.4.1) is different: this approach to reduplication abides by the letter of every clause in (37), but it is nonetheless antithetical to the Four-Hypothesis Program in spirit because it instantiates a framework, Cophonology Theory (Inkelas and Zoll 2007, Inkelas forthcoming), whose ultimate tendency is to abolish the division of labor in exponence between morphology and phonology. Consider Inkelas and Zoll's explanation of the fact that reduplicants typically exhibit a fixed size. In Diyari (Pama-Nyungan, South Australia, extinct), for example, the reduplicant consists of a separate prosodic word comprising exactly one syllabic trochee (Austin 1981: 30, Poser 1989: 132–3, Inkelas and Zoll 2005: 79):

(45) a. [$_{\omega'}$[$_{\omega^\circ}$[$_{\Sigma}$ wi.l̪a]] [$_{\omega^\circ}$[$_{\Sigma}$ wi.l̪a]]] 'woman'
b. [$_{\omega'}$[$_{\omega^\circ}$[$_{\Sigma}$ tʲil.pa]][$_{\omega^\circ}$[$_{\Sigma}$ tʲil.par]ku]] 'species of bird'

In Inkelas and Zoll's analysis, the role of the morphology is limited to concatenating two tokens of the base stem under a mother node and assigning each node to a particular cophonology ($\mathcal{P}_x$, $\mathcal{P}_y$, or $\mathcal{P}_z$):

(46) $\mathcal{P}_z(\mathcal{P}_x$ (/tʲilparku/),$\mathcal{P}_y$(/tʲilparku/))

$\mathcal{P}_x$(/tʲilparku/) $\mathcal{P}_y$(/tʲilparku/)

The job of truncating the lefthand-side daughter to the size of a foot devolves to the phonology: more specifically, to cophonology $\mathcal{P}_x$. In this cophonology, the constraint Lex≈ω requires that a stem should project a prosodic word, and the constraint ω≈Σ requires that a prosodic word should consist of a single foot (in other words, ω≈Σ bans unfooted syllables and dependent feet); both dominate Max, the faithfulness constraint penalizing deletion. In contrast, the righthand-side daughter belongs to the nontruncatory cophonology $\mathcal{P}_y$, in which Max dominates ω≈Σ.

(47) a. $\mathcal{P}_x$: Lex ≈ ω, ω ≈ Σ ≫ Max

b. $\mathcal{P}_y$: Lex ≈ ω, Max ≫ ω ≈ Σ

c. $\mathcal{P}_x$(/tʲilparku/) = $[_\omega[_\Sigma$ tʲil.pa]]

d. $\mathcal{P}_y$(/tʲilparku/) = $[_\omega[_\Sigma$ tʲil.par]ku]

The guiding idea in this analysis is that the parochial phonological effects contingent upon an individual morphological construction are caused by a cophonology specifically associated with that construction. Taken to its logical conclusion, then, this approach entails that there can be as many construction-specific cophonologies as there are constructions: "each individual morphological construction has its own, potentially unique, cophonology" (Inkelas forthcoming: §4.1). In this sense, Morphological Doubling Theory and, more generally, Cophonology Theory fulfill the Morph Integrity Hypothesis in a merely vacuous way: if the phonology of a language is divided into as many chambers as the language has morphological constructions, the need can rarely be felt for the morphology directly to tamper with the phonological content of morphs. Or, to put it differently, it is obvious, as we saw in section 2.4.2.1, that Anderson's (1992) word-formation rules violate (or rather deliberately ignore) Morph Integrity, but one may regard each cophonology in Cophonology Theory as an Andersonian word-formation rule, only sometimes packaged with more general phonological effects taking place in the same cycle. I say 'only sometimes' because Cophonology Theory assumes a sign-based grammatical architecture (Orgun 1996) where every constituent triggers a cycle, but this yields far more cycles than are independently required on phonological grounds: the result is that a vast number of cycles insert a morph but are otherwise phonologically vacuous (for discussion, see Bermúdez-Otero forthcoming). I return to Cophonology Theory in section 2.4.2.3 below.

In important respects, however, Morphological Doubling Theory remains very close to McCarthy and Prince's phonological copying model. In (48) I reproduce

McCarthy and Prince's (1999: 265) own analysis of Diyari, merely substituting Inkelas and Zoll's constraint ω≈Σ for McCarthy and Prince's subhierarchy Parse-σ ≫ Align(Σ,L;ω,L).

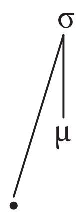
(48) Lex≈ω, IB-Max ≫ ω≈Σ ≫ BR-Max, IR-Max

As it turns out, (47a) and (47b) provide a direct translation of (48) into cophonological language: low-ranking Max in $\mathcal{P}_x$ stands for BR-Max and IR-Max; high-ranking Max in $\mathcal{P}_y$ stands for IB-Max. This reflects a fundamental assumption shared by both analyses: namely, the idea that the foot-size requirement on the reduplicant is imposed by prosodic markedness constraints, instead of reflecting a lexical property of the reduplicative prefix itself. As we shall see in section 2.4.2.3, it is this assumption that prevents the two theories from complying with the modular strictures of the Four-Hypothesis Program.

Working within the broad line of research that I have called Generalized Nonlinear Affixation, however, Saba Kirchner (2010), henceforth 'SK', has shown that many instances of reduplication—some of them of fearsome complexity—submit to much more parsimonious analyses than we have seen so far, involving neither reduplication-specific relationships of correspondence nor construction-specific cophonologies. For example, let us see how this theory deals with the fixed size of reduplicants. Take the Tangale language (Afroasiatic, West Chadic; Nigeria), which uses a CV- reduplicative prefix to mark iterativity in verbs (Kidda 1993: 29):

(49)	mál- 'beat'		mámál-	'beat repeatedly'
	ɗób- 'call'		ɗóɗób-	'call repeatedly'
	sabt- 'ruin'		sasabt-	'ruin repeatedly'

This simple pattern falls out if the underlying representation of the reduplicative prefix consists of a prosodic treelet comprising a syllable node dominating a nuclear mora and an onset position occupied by a featureless root node (•):

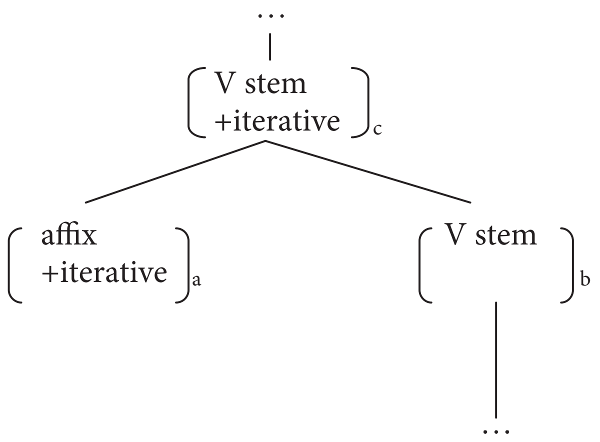
(50)


↔ [+iterative]

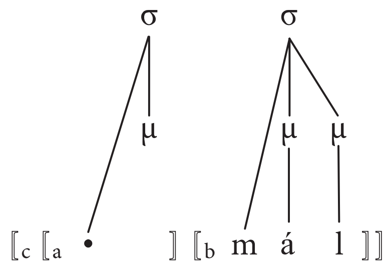
For the sake of simplicity, I shall here assume that the iterative marker is prefixed to a stem that has already undergone a phonological cycle.[44] The input to the cycle containing the reduplicative prefix will therefore look as follows:

[44] The analysis will go through in the same way, with just an added technicality, if in fact reduplication takes place in the first cycle. Empirically, the question boils down to whether the iterative marker attaches to roots or to stems.

(51) a. *syntactic output*



b. *phonological input*



Assuming OT as the theory of mappings, the computation inside the phonology proceeds as follows. First, high-ranking faithfulness constraints of the Max family require that the phonological constituents of the prefix should have output correspondents, and high-ranking Uniformity prevents this requirement from being achieved by mere coalescence with the phonological contents of the stem.[45] At the same time, high-ranking Dep-Feature prevents the prefixal prosodic treelet from being filled out with features that lack input correspondents. With Integrity ranked low, the treelet ends up being filled with duplicates of the melodic material of the stem. Moreover, Tangale tolerates onsetless syllables, but the prefixal prosodic treelet contains an onset root-node that is protected by high-ranking Max-Seg and so must be filled: in consequence, the whole initial CV sequence of the stem, and not just the vowel, must be duplicated.[46] In the optimal candidate, the linear precedence relationships

[45] Note that recourse to MaxFloat (Wolf 2007) is not necessary here (cf. SK: 6). It will nonetheless be permitted in other cases if required, because MaxFloat complies with Indirect Reference (71), unlike constraints of the RealizeMorpheme type (Akinlabi 1996, Kurisu 2001, van Oostendorp 2005a, cf. Trommer 2008c). Floating is a purely phonological attribute unrelated to morphological affiliation: in the languages of the world, floating elements occur in the underlying representations both of stems and of affixes.

[46] In this case, one could also replace (50) with a bare syllable node and rely upon an Anchor constraint to demand that a correspondent of the stem-initial segment should stand in ω-initial position: see Bye and Svenonius's analysis of Saanich in this volume. Insofar as Anchor constraints are members of the Align family, this solution complies with Indirect Reference (71) and so is compatible with the Four-Hypothesis Program (37). Nonetheless, the use of prosodic treelets, as opposed to bare prosodic nodes, will be unavoidable in those instances of reduplication in which the internal prosodic structure of the reduplicant is marked for its position and does not optimize alignment: see the discussion of (60) below.

specified in the input are disrupted, as one of the output correspondents of the stem-vowel appears on the left of one of the correspondents of the initial consonant, and so Linearity must be ranked low. Nonetheless, Linearity is still active in favoring local copying: e.g. duplicating the last consonant of the stem, instead of the first, would result in even more serious unfaithfulness to the linear precedence structure of the input. Tableau (52) summarizes the analysis: for convenience, I omit syllable nodes, I use different types of numerals to represent correspondence relationships on different tiers of representation, and I let F stand for the feature content of a segment.

(52)

μ_i μ_{ii} μ_{iii} $\bullet_1$ m_2 $á_3$ l_4 $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$	Max-μ	Max-Seg	Uniformity	Dep-Feature	Integrity	Linearity
μ_{ii} μ_{iii} m_2 $á_3$ l_4 $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$	*! (μ_i)	*! ($\bullet_1$)				
μ_i μ_{iii} m_1 $á_3$ l_4 $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$	*! (μ_{ii})	*! (m_2)				
$\mu_{i,ii}$ μ_{iii} $m_{1,2}$ $á_3$ l_4 $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$			*! ($\mu_{i,ii}$) * ($m_{1,2}$)			
μ_i μ_{ii} μ_{iii} $ʔ_1$ e m_2 $á_3$ l_4 $\mathbb{F}$ $\mathbb{F}$ $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$				*!* ($\mathbb{F}_ʔ$, $\mathbb{F}_e$)		
μ_i μ_{ii} μ_{iii} l_1 á m_2 $á_3$ l_4 $\mathbb{F}_{III}$ $\mathbb{F}_{II}$ $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$					** ($\mathbb{F}_{II}$, $\mathbb{F}_{III}$)	**!** ($\mathbb{F}_{III}>\mathbb{F}_{II}$, $\mathbb{F}_{III}>\mathbb{F}_I$, $\mathbb{F}_{III}>\mathbb{F}_{II}$, $\mathbb{F}_{II}>\mathbb{F}_I$)
μ_i μ_{ii} μ_{iii} m_1 á m_2 $á_3$ l_4 ☞ $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_I$ $\mathbb{F}_{II}$ $\mathbb{F}_{III}$					** ($\mathbb{F}_I$, $\mathbb{F}_{II}$)	* ($\mathbb{F}_{II}>\mathbb{F}_I$)

Analyses like (52) are easily falsifiable. Because of the high ranking of Dep-Feature, for example, they predict that, if phonologically driven epenthesis takes place in the same stratum as reduplication, then the epenthetic segment will be a copy of the nearest suitable neighbor, rather than an unmarked filler (i.e. /ia/→[i.ja] ≻ /ia/→[i.ʔa]);

Nonminimal prosodic treelets, e.g. (59b) or (59c) as opposed to (59a), may also be favored under an approach to acquisition in which learners discover the underlying representation of reduplicative affixes by factoring out the common properties of individual surface tokens.

an apparent counterexample to this prediction in Kwak'wala is discussed below. If the prosodic affixation approach survives this and other empirical tests, there will turn out to be no theory of reduplication as such: observe that every constraint and every element of phonological representation in (52) has independent motivation in nonreduplicative phenomena. Indeed, Generalized Nonlinear Affixation relies entirely on independently motivated theories of morphosyntactically induced opacity, such as stratification and cyclicity (Kiparsky 2000, Bermúdez-Otero 2011), to describe over- and underapplication in reduplicated forms (SK: 17–22). It thus has no means to generate so-called 'backcopying.' This appears correct: none of the cases of backcopying reported in the literature (e.g. McCarthy and Prince 1995a: §3.6) seems genuine (Inkelas and Zoll 2005: §5.7, ch. 6, Kiparsky 2010: §3). If so, the notorious 'Kager-Hamilton conundrum' (McCarthy and Prince 1999, Inkelas this volume: §10.5.2.4) is revealed as a pseudoproblem: reduplication never causes templatic backcopying simply because it never causes backcopying.

Another beautiful result of Saba Kirchner's is his demonstration that the affixation of empty prosodic nodes can insightfully account for situations in which reduplication alternates predictably with other types of apparently nonconcatenative exponence.[47] Kwak'wala (Northern Wakashan; British Columbia), for example, has a suffix -M̓UːT glossed as 'refuse, useless' (SK: 40). This suffix possesses two lexical allomorphs whose distribution is phonologically controlled (SK: 41): /-m^{ʔ}uːt/ occurs after consonants; /-muːt/ occurs after all vowels, crucially including epenthetic nonmoric [^{ə}], which is not to be confused with underlying (or copied) monomoric [ə]. The addition of -M̓UːT triggers a range of morphological and phonological effects upon the stem. On the phonological side, a regular, phonotactically driven process of epenthesis inserts a weightless [^{ə}] at the juncture between the stem and the suffix if the former ends in a laryngeally marked consonant or in a consonant cluster with a flat or rising sonority profile, neither of which can form a legal coda in Kwak'wala (SK: 37, 45). The weightlessness of this epenthetic [^{ə}] is independently confirmed by metrical evidence: notably, syllables headed by [^{ə}] are skipped in secondary stress assignment (SK: 37–40). On the morphological side, -M̓UːT requires its base to undergo one of the numerous types of 'stem expansion' found in Kwak'wala (SK: 40). The particular type of stem expansion induced by -M̓UːT is also found with the suffix /-(g)iːsaweːʔ/ 'left behind, leave behind' (SK: 40); other suffixes, like /-$^{[+\text{constr gl}]}$d^{z}ək^{w}/ 'to do before doing something else' (SK: 61), to which we shall return below, select other patterns of stem expansion. The stem-expansion pattern induced by -M̓UːT involves vowel lengthening or reduplication, in complementary distribution. This predictable alternation is controlled by syllable weight. In Kwak'wala, only plain (i.e. unglottalized) sonorant codas contribute to weight; all other codas are weightless (SK: 36). A syllable is therefore heavy (bimoric) if it contains a long vowel or a plain sonorant coda consonant, and it is light otherwise; syllables with more than two morae are prohibited (SK: 36). Under -M̓UːT suffixation, then, the stem-vowel lengthens if the bare unexpanded stem

[47] Stonham (1994: §6.6) anticipated this idea in his analysis of the formation of the actual aspect in Saanich. Bye and Svenonius (this volume) provide an optimality-theoretic implementation of Stonham's analysis.

consists of a light syllable; otherwise, there is reduplication. Or, to put it differently, the base of -M̓UːT is subject to lengthening when this is compatible with the ban on superheavy syllables, and otherwise reduplicates (SK: 46). In cases of reduplication, the location of the reduplicant is controlled by complex metrical and segmental conditions that need not concern us here (SK: 47, 52–7).

(53)

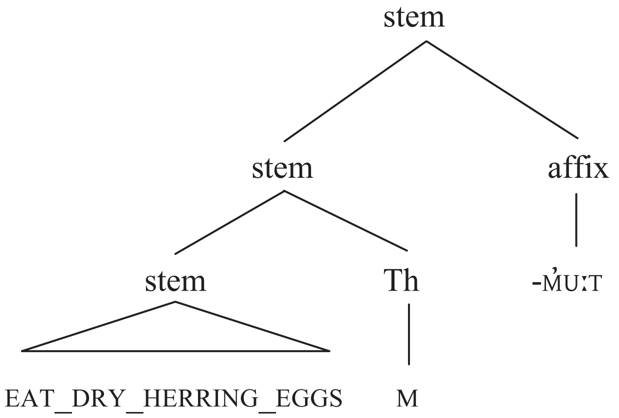
	Bare stem	*Suffixed form*	*Stem expansion*	*Bare-stem gloss*	*Source*
a.	μ təp	μ **μ** μ μ taːp.mˀuːt	lengthening	‘break’	SK: 42
b.	μ gʷəd	μ **μ** μμ gʷaː.dᵊ.muːt	lengthening (with epenthesis)	‘tie’	SK: 42
c.	μμ səl	μμ **μ** μμ səl.sə.muːt	reduplication	‘drill’	SK: 43
d.	μμ siːq’	**μ** μμ μμ sə.siː.q’ᵊ.muːt	reduplication (with epenthesis)	‘eat dry herring eggs’	SK: 44

Observe that all Kwak’wala long vowels become monomoric [ə] when shortened, and that /ə/ becomes [aː] when lengthened (SK: 39). In (53d), therefore, the bimoric stem vowel /iː/ has [ə] as its monomoric counterpart in the reduplicant.

Saba Kirchner’s momentous insight is that the stem-expansion pattern associated with -M̓UːT is caused by an underlying floating mora that is required to dock nonvacuously (SK: 48). When the bare stem is light, the floating mora can simply dock onto the vowel. When the bare stem is already heavy, however, the ban on trimoric syllables blocks vowel lengthening. The next best option is to duplicate segmental material from the stem in order to provide the floating mora with a landing site; an epenthetic vowel, i.e. [ᵊ], will not do, since schwas lacking input correspondents are required to be nonmoric. The results are shown in (53), where I have used bold typeface to indicate the location of the floating mora after docking. As Saba Kirchner notes, suffixing different empty prosodic structures will result in different patterns of stem expansion. The suffix /-$^{[+\text{constr gl}]}$dᶻəkʷ/, for example, always induces reduplication, even with bare stems consisting of a light syllable. This suggests that the stem-expansion pattern associated with /-$^{[+\text{constr gl}]}$dᶻəkʷ/ is caused not by a floating mora, but by an empty prosodic treelet like the one in (50): see SK: 61.

Saba Kirchner assumes that the floating mora that accompanies -M̓UːT originates in the underlying representation of the suffix itself (SK: 48). We have seen, however, that the suffix /-(g)iːsaweːʔ/ triggers exactly the same type of stem expansion as -M̓UːT and, more generally, that stem expansion under suffixation is a highly pervasive phenomenon in Kwak’wala (SK: 40). This strongly suggests that stem expansions are more than mere idiosyncrasies attached to particular affixes, but rather reflect a deeper structural phenomenon in the word-syntax of Kwak'wala. One possibility is that, in fact, Kwak'wala stem expansions realize a thematic position *Th* adjoined to bare stems (Oltra-Massuet 1999, Bermúdez-Otero 2007b, 2007d: 236ff.); different suffixes like -m̓uːt and /-$^{[+\text{constr gl}]}$d^{z}ək^{w}/ will then morphologically subcategorize for different exponents of this *Th* position.[48] If this is correct, then a form like [sə$^{\mu}$.si$^{\mu\mu}$.q'^{ə}.mu$^{\mu\mu}$t] in (53d) will have the following syntactic output representation:

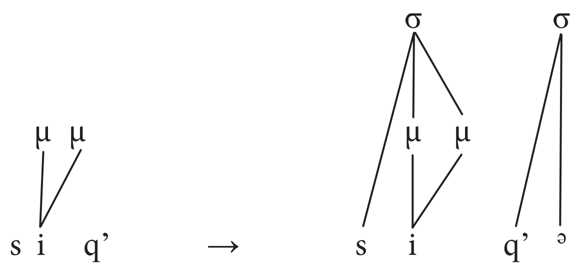
(54)



Moreover, there is evidence to suggest that -m̓uːt must be affiliated to the stem-level phonology. First, the suffix is always immediately adjacent to, or else very close to, the root (SK: 41). Second, derived forms containing -m̓uːt often display the sort of irregularities and exceptions that are characteristic of lexical redundancy rules (SK: 43–4). As we saw in section 2.3.3.2, however, stem-level lexical redundancy rules typically induce internal cyclic effects through nonanalytic listing and blocking. If so, we may expect the surface realization of (54) to show the effects of three cycles.

These proposals answer some unsolved questions hanging over Saba Kirchner's analysis. A derivation of [sə$^{\mu}$.si$^{\mu\mu}$.q'^{ə}.mu$^{\mu\mu}$t] in three cycles will proceed as follows. In the first cycle, syllabification applies to the bare stem /si$^{\mu\mu}$q'/. The stem-final ejective, being laryngeally marked, cannot be syllabified in the coda, and so a weightless [^{ə}] is epenthesized to provide an onset position for the /q'/.

(55) *First cycle*

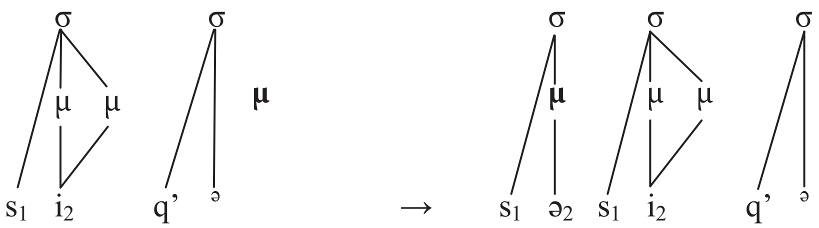


The nonmoric schwa of [sə$^{\mu}$.si$^{\mu\mu}$.q'^{ə}.mu$^{\mu\mu}$t] is thus inserted **before** the floating mora realizing the *Th* position enters the phonological derivation. This explains why, in this case, Kwak'wala fails to avail itself of the apparently most efficient solution: namely,

[48] I set aside the question whether the floating [+constr gl] feature associated with the latter suffix really belongs to it or is rather affiliated with the *Th* exponent for which the suffix subcategorizes.

to use a copy of the stem vowel to provide at a single stroke both a landing site for the floating mora and an onset position for the stem-final consonant, i.e. /$si_1{}^{\mu\mu}$q’-$^{\mu}$/ → *[$si_1{}^{\mu\mu}$.q’$ə_1{}^{\mu}$] instead of [$sə_1{}^{\mu}$.$si_1{}^{\mu\mu}$.q’^{ə}]. Output *[$si_1{}^{\mu\mu}$.q’$ə_1{}^{\mu}$] is precisely what one would expect if, as Saba Kirchner suggests, the rescue of the stem-final /q’/ and the docking of the floating mora both took place in parallel in the same cycle; metrical constraints cannot explain why *[$si_1{}^{\mu\mu}$.q’$ə_1{}^{\mu}$.mu$^{\mu\mu}$t] loses to less faithful [$sə_1{}^{\mu}$.$si_1{}^{\mu\mu}$.q’^{ə}.mu$^{\mu\mu}$t], for the former has exactly the same foot structure as optimal [$sə_1{}^{\mu}$]$^{\mu}$.$sə_1{}^{\mu}$.mu$^{\mu\mu}$t] in (53c). As it happens, the floating mora only becomes visible in the second cycle; but at this point it cannot dock onto the weightless schwa created in the first cycle, possibly because landing on this site would have a vacuous effect on length (the schwa would remain short). The floating mora must accordingly seek an anchor elsewhere. But, as we know, the stem-vowel /iː/ is not an option because trimoric vowels are prohibited. Landing on yet another epenthetic vowel is also impossible because epenthetic vowels are required to be weightless. The only way out, then, is to dock onto a copy of the stem-vowel: hence reduplication.

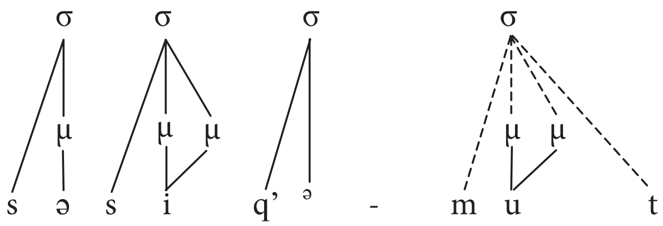
(56) *Second cycle*



The exponence of the suffix -m̓uːt itself must wait until the third cycle. At this point, the input representation of the base crucially ends with a vowel. This means that the choice between the allomorphs /-muːt/ and /-m^{ʔ}uːt/ can be settled quite simply by means of phonological subcategorization, rather than through output optimization (see notes 41 and 42 above). Subcategorization is surely the most straightforward option in this case, since a markedness motivation for the alternation appears elusive. If the whole construction were spelled out in a single cycle, as Saba Kirchner assumes, subcategorization would not be an option, for in that case the input representation of the base would end in a consonant; in fact, Saba Kirchner does not provide an analysis of these facts (SK: 41).

(57) a. -m̓uːt ↔ { /-muːt/ / V__ ; /-m^{ʔ}uːt/ }

b. *Third cycle*



These results suggest that a cyclic architecture, of the type that I shall be advocating in section 2.4.4 on locality grounds, provides the best home for Generalized Nonlinear Affixation, which I have chosen as a theory of nonconcatenative exponence for reasons of modularity. This consilience of arguments bodes well for the Four-Hypothesis Program.

It remains only to show why no problem arises over the fact that the stem-level constraint hierarchy of Kwak'wala inserts an unmarked vowel (weightless schwa) in phonotactically driven epenthesis, but copies neighboring segments in reduplication (cf. above and SK: 45, footnote 14). The solution lies precisely in the fact that Kwak'wala epenthetic vowels are generally required to be nonmoric, as shown by the facts of secondary stress. Under this weight régime, DEP-μ will favor epenthesis over copying, but a floating mora supplied in the input will be able to dock only onto a copy vowel and not onto an epenthetic one.

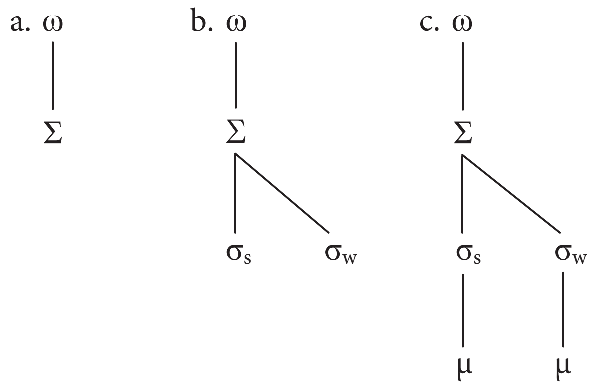
2.4.2.3 Against morph-specific phonology, against process morphology, for (emergent) levels On the surface, each token of an affix displays phonological properties originating in two different sources: some properties are idiosyncratic attributes of the affix itself, which must be listed in the lexicon and included in underlying representations; others reflect the operation of more general mechanisms in the phonology under the influence of the surrounding context. Our discussion of reduplication in section 2.4.2.2 has shown this to be also true of apparently nonconcatenative exponents. In Kwak'wala, for example, one exponent of the *Th* position, the one selected by affixes like -M̓U:T, happens to weigh precisely one mora; another, the one selected by the suffix / — $^{[+\text{constr gl}]}$d^{z}ək^{w}/, consists of one syllable. These are arbitary lexical facts, already specified in underlying representations. In the former case, however, the choice of landing site for the floating mora is not lexically idiosyncratic, but is determined by general well-formedness constraints of Kwak'wala phonology: the mora docks onto a stem-vowel unless prevented by the ban on trimoric syllables.

Similarly, the fact that the reduplicative prefix of Diyari constitutes a separate prosodic word comprising exactly one foot cannot be deduced from independent morphological or phonological generalizations: Diyari has numerous affixes that shelter under the same prosodic word as the stem, and prosodic words larger than one foot arise in a wide variety of morphological circumstances (Poser 1989: 118–19).

(58) a.	*Diyari monomorphemic words*			
	with unfooted syllables		[$_{\omega}$[$_{\Sigma}$ˈpina]du]	'old man'
	with dependent feet		[$_{\omega}$[$_{\Sigma}$ˈŋanda][$_{\Sigma}$ˌwalka]]	'to close'
b.	*Diyari polymorphemic words*			
	with unfooted syllables	ˈkan̪a-n̪i	[$_{\omega}$[$_{\Sigma}$ˈkan̪a]n̪i]	'man-LOC'
	with dependent feet	ˈkaɳa-ˌwaɽa	[$_{\omega}$[$_{\Sigma}$ˈkaɳa][$_{\Sigma}$ˌwaɽa]]	'man-PL'

What the phonological grammar of Diyari does account for is the fact that the foot containing the reduplicative prefix is a syllabic trochee (rather that a bimoric trochee or an iamb) and that its second syllable must end with a vowel (Poser 1989: 133). The analysis of Diyari reduplication in Generalized Nonlinear Affixation will reflect this division of labor transparently: (59a) is the minimal nonredundant underlying representation for the reduplicative prefix; a partially redundant underlier[49] may look like (59b) or even (59c), but it is nonetheless crucial that it should have the size of a foot.

(59)



In their respective analyses of Diyari reduplication, however, McCarthy and Prince (1999: 265) and Inkelas and Zoll (2005: 79) adopt a different strategy, which expands the role of the phonology at the expense of the lexicon. Capitalizing upon the fact that there exist markedness constraints capable of reducing a prosodic word to the size of a single foot, both frameworks shift the burden of accounting for the size of the Diyari reduplicative prefix onto the phonology: see (47) and (48) above. In this section, I show how this analytic move compromises the modular distinction between morphology and phonology. In the course of this discussion, I provide arguments against the claim that exponence can be processual and that morph-specific behavior must be accommodated in the phonology rather than in the lexicon.

The input representation of the Diyari reduplicant in McCarthy and Prince's (1999: 265) analysis contains even less information than (59a); it is purely abstract /RED/. Since BR-identity favors copying the entire base, the role of reducing the reduplicant to a single foot devolves to the markedness constraints PARSE-σ ≫ ALIGN(Σ,L;ω,L). But, as we saw in section 2.4.2.2, this analysis violates Indirect Reference (71): each IO-faithfulness constraint is split into two indexed copies, one monitoring the output realization of RED (IR-faithfulness), the other taking scope over everything else (IB-faithfulness). Concomitantly, surface representations must be supplied with labels of

[49] Section 2.3.3 provided arguments for admitting redundancy in lexical entries. See also note 46 on bare prosodic nodes vs. nonminimal prosodic treelets.

morphological affiliation, in breach of Phonetic Interpretability (76), so that EVAL can track violations of IB-faithfulness, IR-faithfulness, and BR-identity.[50]

These departures from modularity come on top of empirical difficulties arising from a putative 'Reduplicant-Default Connection', which is entailed by the use of markedness constraints to account for parochial properties of reduplicants:

(60) *Reduplicant-Default Connection* (Alderete et al. 1999: 334)
Where not copied, reduplicants are like defaults.

This alleged Reduplicant-Default Connection is contradicted by languages like Tonkawa (Coahuiltecan, Texas and Oklahoma, extinct; Gouskova 2007), where the shape of the reduplicant is not unmarked for its position. It is also challenged by languages in which a single phonological level displays different reduplicative patterns exposing different morphosyntactic features, since only one reduplicative pattern can match the prosodic and melodic defaults for that level: an example is found in the multiple stem-level reduplicative patterns of Nuu-chah-nulth (Southern Wakashan, Vancouver Island; Stonham 2007b: §3.1.3–3.1.5). Generalized Nonlinear Affixation avoids this problem because it acknowledges that the noncopied properties of a reduplicative affix have two possible origins: they may be defaults or they may be specified in the underlying representation of the reduplicative affix, as is in fact the case for **all** affixes (SK: 111ff.).

We have seen that, by asking the phonology to specify the size of the Diyari reduplicant, McCarthy and Prince (1999) give up modularity: more specifically, they breach Indirect Reference and Phonetic Interpretability. By adopting the framework of Cophonology Theory, Inkelas and Zoll (2005) make this threat to modularity even easier to see—easier to see, though not more serious than it already is in mainstream OT. In fact, mainstream OT has since its inception routinely resorted to devices whose practical effect is equivalent to the setting up of a construction-specific cophonology:

(61) *Morph-specific phonology in mainstream OT*

a. parochial alignment constraints: e.g. ALIGN($[ka]_{Af}$,L;Σ_S,R) (McCarthy and Prince 1993a: (5c));

b. templatic constraints specifying the shape of reduplicants: e.g. RED$\geq \sigma\sigma$ (McCarthy and Prince 1993b: §5.3, (40), (48));

c. BR-identity constraints indexed to individual reduplicants: e.g. $BR_{DIMINUTIVE}$-MAX-Seg vs. $BR_{DISTRIBUTIVE}$-MAX-Seg (Urbanczyk 1995);

50 McCarthy and Prince (1999: 262) further claim that a reduplicant can be designated as belonging to one of the categories *root*, *stem*, or *affix* (see Urbanczyk, 2006, for an elaboration of this idea). They suggest that the Diyari reduplicant projects a separate prosodic word because it is a 'stem'. Yet ascribing stem status to a reduplicant solely on the basis of its phonological properties, without further morphosyntactic or semantic checks, results in a viciously circular argument, and one that risks depleting the empirical content of the concepts of root, stem, and affix. In any case, no morphosyntactic categorization can explain the foot-size restriction on the Diyari reduplicant: the language allows both roots and affixes to exceed the size of a foot, and attested morphs include a tetrasyllabic suffix and a pentasyllabic root (Crowhurst and Hewitt 1997: footnote 11, citing a personal communication from the author of Austin 1981).

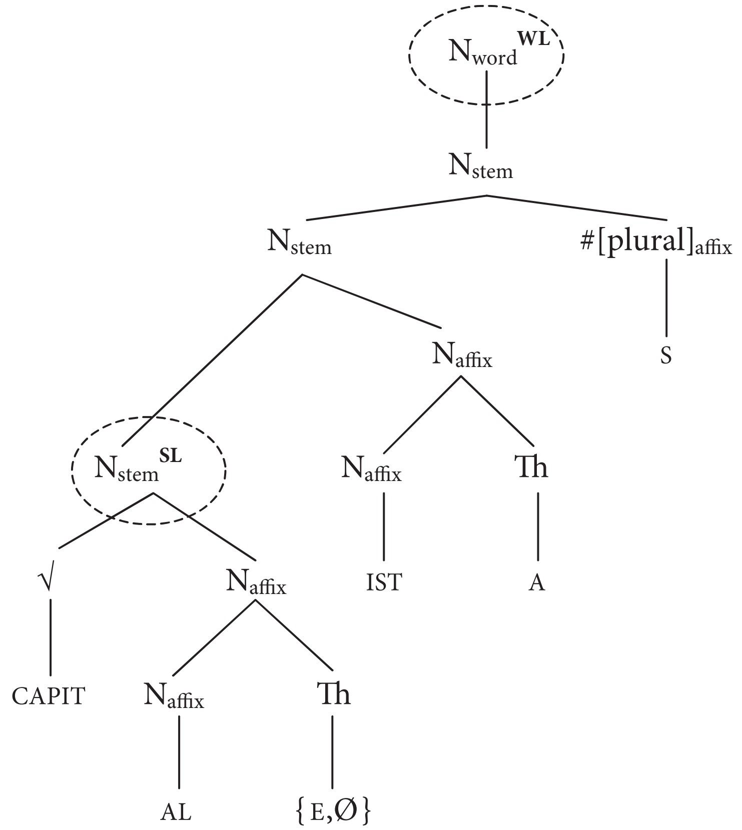
d. OO-identity constraints that arbitrarily stipulate both the base of the correspondence relationship and the affix that triggers it (Benua 1997: 25, 30, 109–10, 154–5, etc.);
e. lexically indexed IO-faithfulness constraints (Fukazawa 1997);
f. lexically indexed markedness constraints (Pater 2010: §6).

These devices are no less afflicted than cophonologies by the problems that I shall be diagnosing in this section, and the same corrective arguments apply to them. They lack, however, the two great virtues of Cophonology Theory: conceptual transparency, and the local and scopal effects of a cyclic architecture (on this point, see further §2.4.3). Benua (1997: 154) openly acknowledges that her theory constitutes a noncyclic counterpart of Cophonology Theory: "In the limit, T[ransderivational] C[orrespondence] T[heory] allows morpheme-specific phonological behavior." She envisages just one restriction: "The only phonological limitation on the variety of patterns produced in the same language is that they all obey the same relative markedness relations, because there is only one ranking of constraints" (p. 229); but even this curb is abandoned in current proposals for constraint indexation (Pater 2010: §6).

Having shown the extent to which mainstream OT has embraced morph-specific phonology, I now return to the link between Inkelas and Zoll's (2005) approach to reduplication and Cophonology Theory. In a morphological-doubling analysis, the input representation of the Diyari reduplicant is identical with that of its base; the foot-size restriction is imposed by phonological means during a cycle triggered by the reduplicant alone (46). This reduplicant cycle invokes a constraint hierarchy in which $\omega \approx \Sigma$ is top-ranked (47a,c). But this ranking does not apply beyond reduplicants: in all other situations, Diyari prosodic words are permitted to be larger than one foot (58). Base and reduplicant must therefore be assigned different constraint rankings, and Morphological Doubling Theory ends up intrinsically committed to the existence of morph-specific phonology.

Cophonology Theory (e.g. Inkelas and Zoll 2007, Inkelas forthcoming) seeks to make a virtue out of this necessity. For example, Inkelas (forthcoming: §5.7) argues that apparent phonological nonuniformity effects across constructions require the full power of construction-specific cophonologies and cannot be handled by more restrictive theories like Stratal OT. The latter countenances just three phonological subsystems within the grammar: the stem, word, and phrase levels, which are primarily associated with hierarchical levels in the syntax and only secondarily with specific affixes. The stem-level constraint hierarchy, for example, drives the phonological processes that apply in cycles triggered by root-to-stem derivation, and the word-level constraint hierarchy drives the phonological processes that apply in cycles triggered by syntactically free grammatical words (Bermúdez-Otero 2007b: 283). In a structure like (62), therefore, the affiliation of the lower circled node to the stem level and of the higher circled node to the word level is completely determined by their respective structural positions; the arbitrary properties of affixes come into play only for intermediate nodes: for example, an affix that regularly attaches to roots will usually trigger a stem-level cycle even when added to a stem.

(62) *Spanish* capit-al-Ø-ist-a-s (see Bermúdez-Otero 2007d: 236ff)
capit-al-TH-ist-TH-PL
'capitalists'



In support of Cophonology Theory, Inkelas makes two specific objections to Stratal OT. First, she states that "Stratal OT has little to say about realizational morphology or its relation to morphologically conditioned phonology" (forthcoming: §5.7). The implied premise here is that phonological constraints must bear the main burden of accounting for the phonological effects of nonconcatenative exponence, as is indeed the case in the morphological-doubling approach to reduplication (see the discussion of Diyari above). If so, the three levels of Stratal OT will certainly not be enough to specify the parochial properties of each apparently nonconcatenative construction. Stratal OT, Inkelas therefore concludes, "requires supplementation with indexed constraints or cophonologies, thus merging with the other approaches" (2009: §7.3, p. 8). But the premise of this argument, and so its conclusion, is false. Our discussion of reduplication indicates that the responsibility for describing its phonological effects should be more evenly shared between the underlying prosodic and (in cases of fixed segmentism) melodic properties of the reduplicative affix, on the one hand, and general phonological processes of the language, on the other. In the analysis of reduplication, therefore, Stratal OT asks the lexicon to bear responsibility for facts about the shape of the reduplicant that lack independent motivation in the phonology of the language, whereas it does undertake to account for the misapplication of regular phonological processes in reduplicative constructions (Kiparsky 2010). More generally, Stratal OT does have something very definite to say about nonconcatenative exponence: as a cyclic framework, it is suitable for implementing the Four-Hypothesis Program, and so can naturally be paired with that program's preferred theory of nonconcatenative exponence, namely Generalized Nonlinear Affixation.

Second, Inkelas implies that Stratal OT will be unable to cope with phonological nonuniformity effects even in cases of plain concatenation.[51] To illustrate her assertion that "each individual morphological construction has its own, potentially unique, cophonology," Inkelas (forthcoming: §4.1) adduces the behavior of the comparative suffix *-er* in present-day English. But is there anything unique about comparative *-er*, and does it raise any problems for Stratal OT? Let us begin by establishing the stratal affiliation of this affix. If for the sake of brevity we confine ourselves to phonological evidence and set other criteria such as productivity and syntactic position aside, a vital clue comes from the process of dentalization found in certain Northern Irish dialects of English, which causes the coronal noncontinuants /t, d, n, l/ to become dental when followed by /(ə)ɹ/ (Harris 1985: 58, 211ff., Bermúdez-Otero 2011: §2). Dentalization applies normally in the presence of stem-level suffixes (63a), but underapplies before word-level suffixes like agentive *-er* (63b). Comparative *-er* triggers normal application when attached to suppletive bound roots (63c), but induces underapplication in free stems (63d). We must therefore conclude that comparative *-er* is word-level, except in listed irregular root-based constructions, which, as we saw in (62), generally belong to the stem level.

(63) a.	*stem-level suffixes*	sani[t̪]-ary		
		eleme[n̪t̪]-ary		
b.	*agentive* -er	bett-er	[ˈbætəʴ]	'one who bets'
		heat-er	[ˈhitəʴ]	
c.	*comparative* -er *with suppletive bound roots*	bett-er	[ˈbæt̪əʴ]	'good.CMPR'
d.	*comparative* -er *with free stems*	fatt-er	[ˈfatəʴ]	
		la[t]-er		

Now we ask whether the other phonological properties of comparative *-er* are consistent with its affiliation to the word level. None of those cited by Inkelas (forthcoming: §4.1) proves problematic: comparative *-er* behaves like any other English word-level suffix in not affecting the location of stress and in triggering neither trisyllabic shortening nor velar softening. Inkelas also mentions the fact that comparative *-er* selects either monosyllabic stems or certain types of trochaic disyllabic stems with predominantly light final syllables; the examples in (64) come from Sproat (1998: 342) and Newell and Scheer (2007).

[51] Inkelas (1998: 134–5, forthcoming: §6) also mounts an attack on Stratal OT based on the invalidity of the Affix Ordering Generalization (Selkirk 1982: 91, after Siegel 1974: 182 and Allen 1978: 6, cf. Aronoff 1976: 85, Aronoff and Sridhar 1983, Fabb 1988). However, it is perfectly possible for Stratal OT to preserve a complete serial ordering of phonological levels without espousing the Affix Ordering Generalization (Bermúdez-Otero forthcoming).

(64) a. redder, sadder, wisher, kitscher
b. easier, happier, manlier
c. *ecstaticker, *contenter, *impoliter

This is certainly **not** a general property of English word-level suffixes, but, crucially, it is not a **derived** phonological property either. The pattern in (64) can be described by including a prosodic subcategorization frame in the lexical entry of comparative *-er:* let us say that the suffix subcategorizes for an immediately preceding prosodic word with the shape [$_{\omega}$ ˈσ(σ̆)].[52] As we saw in section 2.4.2.1 (especially note 41), the morphology will verify whether this subcategorization requirement is satisfied in the **input** to the word-level cycle in which *-er* combines with its base; the word-level phonological constraint hierarchy itself need say nothing at all about this matter.

Curiously, Inkelas omits the one phonological trait of comparative *-er* that does at first blush seem incompatible with its affiliation to the word level. The difficulty arises over the forms in (65b): as shown in (65c), English word-level suffixes regularly cause postnasal plosive deletion to overapply (Borowsky 1993: 202, Bermúdez-Otero 2011: §2), but in (65b) comparative *-er* and superlative *-est* unexpectedly fail to do so.

(65)

a.		b.				c.	
long	[lɒŋ]	long-er	[ˈ1ɒŋ.gə]	long-est	[ˈlɒŋ.gɪst]	long-ish	[ˈlɒ.ŋɪʃ]
strong	[stɹɒŋ]	strong-er	[ˈstɹɒŋ.gə]	strong-est	[ˈstɹɒŋ.gɪst]	strong-ish	[ˈstɹɒ.ŋɪʃ]
young	[jʌŋ]	young-er	[ˈjʌŋ.gə]	young-est	[ˈjʌŋ.gɪst]	young-ish	[ˈjʌ.ŋɪʃ]

The evidence of Northern Irish dentalization again holds the key to this problem: we have already seen that, in accordance with general principle, comparative *-er* behaves like a stem-level suffix in irregular root-based constructions (63c). This immediately arouses the suspicion that the comparatives and superlatives in (65b) are irregular root-based forms too. That suspicion is gratifyingly confirmed by two pieces of evidence. First, Chomsky and Halle (1968: 370, footnote 20) observe that post-nasal plosive deletion does overapply as expected in superlative forms like *cunningest* [ˈkʌnɪŋəst] and *willingest* [ˈwɪlɪŋəst], which can only be stem-based since *-est* attaches outside *-ing*. Although *SPE* describes these items as "awkward", the World Wide Web contains a nonnegligible number of tokens of <cunninger> (55,600 Google hits on 24 September 2010), and <cunningest> has made it into controlled corpora like the BNC and COCA. Even more frequent is the word *winningest* [ˈwɪnɪŋəst], used in American English to describe the sportsman or coach who has achieved the most wins (Merriam-Webster *sub voce*): on 24 September 2010 Google returned over two million pages containing <winningest>. The second piece of evidence comes from

[52] As shown by Booij and Lieber (1993), stating the prosodic selectional requirements of comparative *-er* in precisely these terms explains why the suffix is able to attach to *unhappy* (whence *unhappier* 'more unhappy'), but not to *impolite* (64c): compare the prosodifications [$_{\omega'}$ [$_{\omega}$ ˈʌn][$_{\omega}$ ˈhæ.pɪ]] and [$_{\omega}$ ɪm.pə.ˈlaɪt]. There is thus no bracketing paradox: cf. Pesetsky (1979: §2.2) and much subsequent work.

the comparative grade of *wrong*, which, as Geoffrey Pullum has observed,[53] has an extremely low token frequency: cf. the BNC token counts in (66).

(66)	a.	strong	15768	b.	wrong	15505
		more strong	12		more wrong	41
		stronger	2561		wronger	0

Since the synthetic comparative *wronger* occurs so very rarely, it is most unlikely to possess a lexicalized pronunciation; when the form is actually used, therefore, its phonological realization will have approximately the same evidential value as a response to a wug test (§2.3.1). Significantly, the editors of Merriam-Webster (*sub voce wrong*) describe the comparative *wronger* and the superlative *wrongest* as occurring "sometimes" and give the pronunciations [ˈɹɒŋə(ɹ), ˈɹɒŋəst]. While irregular pronunciations with a [g], if ever found at all, could be explained by pattern association with irregular models like *stron*[g]*er* and *stron*[g]*est* (see §2.3.4 above), the realizations actually recorded by Merriam-Webster have no explanation unless the comparative suffix *-er* and the superlative suffix *-est* belong to the word level, as required by Stratal OT, and the highly frequent forms in (65b) are in fact irregular, like (63c).

With empirical content comes heuristic power. It is only because Stratal OT offers a restrictive approach to phonological nonuniformity that it can encounter empirical challenges like the one posed by *lon*[g]*-er, stron*[g]*-er,* and *youn*[g]*-er.* The hallmark of a progressive theory is that it overcomes such obstacles not by weakening its empirical content, but by stimulating the discovery of new facts in whose light its apparent problems dissolve (Lakatos 1970). Less restrictive devices—whether cophonologies or indexed constraints of whatever sort—will be able to accommodate data like (65b) at face value, but in so doing they risk leaving us none the wiser.

Inkelas, however, maintains that we need the freedom to specify the phonological behavior of individual constructions. Yet, in effect, availing ourselves of that freedom amounts to giving up the distinction between morphology and phonology altogether. In section 2.4.2.2 I highlighted a deep affinity between Cophonology Theory and Amorphous Morphology. Exactly like an Andersonian word-formation rule, a construction-specific cophonology is a function that applies under morphosyntactically defined conditions and whose domain and range consist of sets of phonological expressions. Crucially, the morphosyntactic environment of a cophonology can be defined just as narrowly as that of an Andersonian word-formation rule, and the grammar can contain as many of the former as of the latter. A significant difference lies in the fact that, in a minority of cases, cophonologies package an Andersonian word-formation rule together with a number of more general phonological processes; this only happens in a minority of cases because, as I noted in section 2.4.2.2, only a few of the cycles postulated by the sign-based framework (Orgun 1996) of Cophonology Theory have independent phonological motivation. Yet, precisely in these cycles, we see morphology and phonology completely fused: generalizations that even Anderson (1992) would assign to different components of the grammar are implemented by

[53] Language Log, 12 February 2004, http://itre.cis.upenn.edu/~myl/languagelog/archives/000448.html

a single constraint hierarchy. Thus, modularity collapses. And, revealingly, it collapses even though Cophonology Theory upholds Indirect Reference (§2.4.3): Inkelas and Zoll (2007: 137) emphasize that "no individual constraint makes reference to morphological information". By itself, however, this adherence to Indirect Reference matters less than it may seem: the constraints themselves can be seen as merely providing an alternative formalization for the structural change of an Andersonian word-formation rule, plus additional phonological processes where relevant. Notice again the solidarity between the elements of the Four-Hypothesis Program: Indirect Reference alone, without the support of Morph Integrity, is not enough to prevent morphology and phonology from falling together.

Again, Inkelas (forthcoming: §5.6) sees virtue in the fusion of morphology and phonology: "In many cases," she asserts, "it is difficult or impossible to determine which phonological effect is the primary marker of a morphological construction (i.e. morphology), and which is the secondary phonological correlate (i.e. morphologically conditioned phonology)." Inkelas (2009: §6.6, p. 6) expressed the idea more decisively: "it is difficult or impossible or **pointless**" (emphasis mine). Let me address the argument from difficulty first. There is certainly no denying that, in many cases, we as linguists will find the distinction between primary and secondary exponence hard to make, simply because we do not know to look for the cues that children rely upon; and the fact of diachronic morphologization (e.g. Anderson 1988: 330–3) shows that even children can often get it wrong. However, the inherent empirical difficulty attendant on the distinction can be seriously compounded by linguists' theoretical decisions, as when they choose to take apparently nonconcatenative morphology at face value as processual: if a process manipulating phonological material can reside either in the morphology or in the phonology, analytic underdetermination (§2.2) will indeed rear its ugly head. In contrast, the Morph Integrity Hypothesis and Generalized Nonlinear Affixation alleviate the difficulty by withdrawing the option of having phonological transformations in the morphology: that is their very rationale.

However, the decisive issue concerns the alleged pointlessness of the distinction between primary and secondary exponence, rather than its difficulty: if a distinction is important and useful, one perseveres in the attempt to draw it regardless of difficulty; one abandons it only if it indeed proves pointless. But analyzing apparently nonconcatenative exponence into lexical allomorphy, morphological concatenation of nonlinear pieces, and phonological derivation has a perfectly clear point: to enable more constrained theories of morphology and phonology to emerge by cleansing their respective empirical domains of nongermane matter, and to avoid missing true generalizations. For example, Inkelas cites the contrast between English stress-neutral and stress-affecting suffixes (e.g. *parént-al* vs. *párent-less*) as a case of morphologically conditioned phonology (forthcoming: §2), and she gives stress shift in English verb-to-noun conversion (e.g. *tòrmént*$_{\mathrm{V}}$ $\rightarrow$ *tórmènt*$_{\mathrm{N}}$) as an instance of "process morphology" (forthcoming: §3). In the latter case, however, the forestressing of the noun can be analyzed as secondary rather than primary: i.e. as the effect of general phonological processes upon a domain created by a syntactic operation with a null exponent (see also Trommer this volume: §9.3.1 for a discussion of English conversion along these lines). Inkelas portrays this as a futile choice; but, as Kiparsky (1982b: 12) showed, it matters terribly. First, observe that, in English, it is not only suffixation but also conversion that can be either stress-neutral or stress-affecting: verb-to-noun conversion is sometimes accompanied by stress shift, though, crucially, not always, as we shall see presently; in contrast, noun-to-verb conversion preserves the stress contour of the base without exception.

(67)

			base	*derivative*
a. *Suffixation*	*stress-affecting*		párent	parént-al
	stress neutral		párent	párent-less
b. *Conversion*	*stress-affecting*	(V→N)	tòrmént$_{\mathrm{V}}$	tórmènt$_{\mathrm{N}}$
	stress neutral	(N→V)	réference$_{\mathrm{N}}$	réference$_{\mathrm{V}}$

Moreover, when converted nouns like *torment*$_{\mathrm{N}}$ shift from an oxytonic to a paroxytonic contour, they are in effect adopting the **default** location for primary stress in simple and root-based stems of their category:

(68) a. *Verbs with final -VCC rhymes:* convínce, lamént, usúrp
b. *Nouns with final -VCC rhymes:* próvince, párent, vórtèx

As we saw in the discussion of (62) above, simple and root-based stems constitute domains for the stem-level phonology. Therefore, the role of morphology in English verb-to-noun conversion may simply be reduced to submitting the base to a pass through the stem-level phonology (see again §2.4.2.1):

(69) N ↔ ⟦SL __⟧
|
V

In contrast, converted verbs like *réference*$_{\mathrm{V}}$ fail to adopt the default stress pattern for verbs; but, on the other hand, a systematic failure to induce stress shift even when this violates core metrical constraints of the language is a hallmark of English word-level constructions: cf. the stark violation of the final trisyllabic window in *mémory-less-ness*. Accordingly, the exponence of noun-to-verb conversion boils down to the following:

(70) V ↔ ⟦WL __⟧
|
N

Kiparsky's great insight was to realize that, by analyzing verb-to-noun conversion as stem-level zero-derivation and noun-to-verb conversion as word-level zero-derivation, one can explain important properties of both constructions. First, we saw earlier that the stem level is the realm of lexical redundancy rules, which are characteristically semiproductive (§2.3.3.1). This immediately makes sense of the fact that verb-to-noun conversion is far less productive than noun-to-verb conversion (Kiparsky 1982b: 12). We also saw that stem-level outputs undergo nonanalytic listing, that lexical redundancy rules are subject to blocking, and that this gives rise to internal cyclic effects that are bound by Chung's Generalization (22) and spread and recede diachronically by lexical diffusion (§2.3.3.2, §2.3.3.3). All of this predicts that verb-to-noun conversion will initially yield irregular end-stressed nouns that only become forestressed by lexically diffusing change. To understand why this must be so, observe first that the rules preventing English nouns from bearing primary stress on the final syllable sustain lexical exceptions: e.g. *cadét, Julý, hòtél, bàmbóo*. This shows that pre-specified metrical structure can block default forestressing in disyllabic nouns, with clear consequences for verb-to-noun conversion. As a stem-level output, the verb stem *tòrmént*$_V$ will be subject to nonanalytic listing, and so its stress contour will be stored in its lexical entry, even though it is the default for verbs. Therefore, when *tòrmént*$_V$ undergoes conversion to a noun, the input representation submitted to the stem-level phonology in accordance with (69) will be /ˌtɔːˈmɛnt/, whose metrical specifications will block noun extrametricality (Hayes 1982: 240). The outcome will thus be a converted noun with irregular end-stress: *tòrmént*$_N$, which, again by nonanalytic listing, will be given its own entry in the lexicon, including its oxytonic contour. However, once established as a separate lexical item, *tòrmént*$_N$ will in the course of time become subject to the regularizing effect of the default pattern (recall the parallel case of *ápplicable* > *applícable* in §2.3.3.1). As stated above, therefore, we expect to find a diachronic process of lexical diffusion whereby originally isotonic end-stressed verb-noun pairs become diatonic: e.g. *addréss*$_V$ ~ *addréss*$_N$ > *addréss*$_V$ ~ *áddrèss*$_N$. The historical evidence confirms this prediction: Sherman (1975) documents the gradual rise of diatonic verb–noun pairs from the late sixteenth century. Indeed, the change is still ongoing: e.g. British English retains conservative *addréss*$_N$, whereas innovative *áddrèss*$_N$ has become current in American dialects. Moreover, even after a verb-noun pair has become diatonic, there often remains an off-line cyclic effect: notice that, even though primary stress has moved to its default location in converted nouns like *tórmènt*$_N$ and American English *áddrèss*$_N$, the subordinate foot carried by the second syllable is a marked option; cf. simple or root-based nouns like *pár*[ə]*nt, cýpr*[ə]*s*, etc. Obviously, the subordinate foot of *tórmènt*$_N$ is a cyclic remnant of the foot structure of the base verb *tòrmént*$_V$. But this leads to yet one further prediction. Chung's Generalization requires that internal reapplication effects within stem-level domains should be mirrored by outright lexical exceptions in underived forms, and so it proves in this case: alongside default *pár*[ə]*nt* and *cýpr*[ə]*s*, one finds marked *wísènt* [ˈviːˌzɛnt][54] 'aurochs, European bison' and *ábscèss*.

Turning now to noun-to-verb conversion, we have analyzed it as word-level zero-derivation (70). If so, its outputs must be listed analytically (20). We therefore expect the absence of stress shift in converted verbs like *réference*$_V$ to be exceptionless and diachronically stable: recall our discussion of the exceptionless alternation of word-level past-tense /-d/ between [d], [t], and [ɪd] in section 2.3.3.1. Once more, the prediction proves correct (Kiparsky 1997: (27a)). Interestingly, analytic listing appears

[54] The only pronunciation recorded in the *OED* is [ˈwiːzənt], but more recently Wells (2000) gave [ˈviːˌzɛnt] as the preferred variant in both British RP and General American; Wells lists realizations ending in [zənt] as secondary variants. Merriam-Webster records [ˈviːˌzɛnt] only.

to have a similar effect on lexical semantics: the meaning of true denominal verbs is highly predictable (Kiparsky 1997: §3).

The analysis proposed in (69) and (70) thus uncovers a dense web of connections linking conversion to a wide range of synchronic and diachronic facts about English stress. It would therefore be an error to describe the stress alternation in pairs like *tòrmént*$_V$ ∼ *tórmènt*$_N$ as an instance of 'process morphology', for such an account makes none of the correct predictions we have deduced above: namely,

(i) the difference in productivity and semantic transparency between verb-to-noun and noun-to-verb conversion;
(ii) the lexically diffusing rise of diatonic pairs in cases of verb-to-noun conversion;
(iii) the exceptionlessness and stability of the forestressed isotonic pattern in cases of noun-to-verb conversion;
and
(iv) the link via Chung's Generalization between the existence of conservative derived *addréss*$_N$ and that of nonderived *cadét*, and between innovative derived *áddrèss*$_N$ and nonderived *ábscèss*.

Admittedly, when Inkelas herself applies the term 'process morphology' to the stress shift in *tòrmént*$_V$ ∼ *tórmènt*$_N$ (forthcoming: §3), she does so without theoretical commitment; her programmatic goal is, rather, to abolish the very distinction between primary and secondary exponence. Yet, to justify discarding this distinction, one would need to show that one can reach the same standard of explanation without it. Cophonology Theory does have a way of expressing phonological commonalities across constructions: the corresponding cophonologies can be displayed as terminal nodes within an inheritance hierarchy or grammar lattice (e.g. Anttila 2002; see Inkelas forthcoming: §4.1, §8).[55] Yet, apart from the fact that it is not clear how nonanalytic listing, blocking, and Chung's Generalization fit into this picture, the main problem is that the grammar lattice offers data compression, not deductive explanation: conceptually, the cophonologies come first, and nonterminal nodes in the lattice are set up after the fact; instead, deductive explanation would require nonterminal nodes in the lattice to come first and to enable predictions about the cophonologies.[56]

The same issue of priority arises again when one thinks about the acquisition of phonological nonuniformity effects by the child. Any tolerably comprehensive view of the child's phonological development will, I think, support the inference that children learn about the shared phonological properties of morph combinations and of individual morphs earlier than they discover their differences. Indeed, we know that an infant has learnt a vast amount of phonetics and phonology before she has isolated any of the words (let alone the stems or affixes) of her target language. At one extreme, Mehler et al. (1998) report that four-day-old French infants recognize French utterances as belonging to their native language, but do not distinguish between Italian

[55] Inkelas (forthcoming: §8) also asserts that diachronic factors keep cophonology divergence at bay. On this argument, see note 57 below.

[56] 'Deductive explanation' is something of a pleonasm in strict usage: properly to explain is to deduce (Hempel and Oppenheim 1948).

and English; the effect remains even when low-pass filtering of the signal at 400 Hz removes much segmental information, suggesting that French babies are relying on their knowledge of French prosody, possibly acquired *in utero*. Later, children recognize and store multi-word phrases before they isolate individual grammatical words (Peters 1983, Arnon 2009), as shown, for example, by missegmentation errors like *Give-it the ball*; and it is uncontroversial that children recognize, store, and use words as unanalyzed wholes before they can identify individual stems and affixes. In the specific case of sublexica like those of Japanese (Itô and Mester 1995a), Ota (2004) argues on learnability-theoretic grounds that the shared properties of the superset must be acquired before the differences between the subsets.

In line with this evidence, Bermúdez-Otero (1999: 101–2, 2003: §4.1) proposes that the stratal architecture of phonology rises level by level in a sequence of developmental stages, each of which is initiated by the learner's discovery of a new, more fine-grained region of syntactic structure: the child starts out with multi-word units, then discovers words, and finally isolates stems and affixes; correspondingly, she first sets up the phrase-level phonology, and then on top of it she successively builds the word and stem levels. Such a scenario accords well with the concepts of hierarchical constructive development (Quartz 1999: 54) and sequenced bootstrap learning (Lappin and Shieber 2007: 424–5). In this framework there is only one place, the stem level, where one might expect to witness the emergence of parallel cophonologies competing for application to domains of the same hierarchical rank; this is because the stem level is the final staging post in the life cycle of phonological processes, to which phonological patterns are relegated whose opacity or irregularity prevents the learner from subsuming them under the word-level phonology (see Meir 2006 for an example of this phenomenon in Modern Hebrew).[57] Indeed, in relatively well-understood cases like that of English, only the stem level affords the sort of empirical evidence that might support a plausible argument for parallel cophonologies (Zamma 2005, Bermúdez-Otero and McMahon 2006: 404–5). Yet, even in this case, one must consider alternative scenarios. Since stem-level constructions are characterized by semiproductivity and nonanalytic listing, one should countenance the possibility that variations in lexical experience across speakers may manifest themselves as differences in the extent to which individual learners acquire morphophonological generalizations over stem-level items, including whether or not they end up building parallel cophonologies: **some** speakers may acquire **some** aspects of stem-level phonology late, or not at all. A piece of evidence supporting this scenario comes from English weak irregular verbs like *keep* ~ *kep*(-)*t*: data from rates of /t/-deletion (Guy and Boyd 1990) reveal that speakers differ in the extent to which they treat the final /t/ either as a segment

[57] On the life cycle of phonological processes in general, see Bermúdez-Otero (2007c: 503ff., 2011: §3) and specially Bermúdez-Otero and Graeme Trousdale (forthcoming: §2). Citing Bermúdez-Otero and McMahon (2006), Inkelas (forthcoming: §8) claims that diachronic factors can assume the role of limiting the divergence between different construction-specific cophonologies in Cophonology Theory. Yet this appeal to diachrony remains largely promissory, in the absence of a demonstration that Cophonology Theory can support a learning-theoretic account of the life cycle of phonological processes. Bermúdez-Otero (1999: 100–3, 2003: §4ff.) outlines a model of the acquisition mechanisms that drive the life cycle of phonological processes, but this model crucially presupposes a hierarchy of phonological levels arranged in a complete serial order, as in Stratal OT (see note 51); Cophonology Theory does not meet this requirement.

belonging to a suppletive root allomorph or as a separate morph exposing past tense and possibly causing closed syllable shortening of the stem-vowel; some speakers remain stuck with the former analysis until late in adult life. Taking all this evidence together, I suggest that parallel cophonologies, if they exist at all, may occur as part of the stem-level syndrome, along with semiproductivity, exceptions, and Chung's Generalization: see again section 2.3.3.

2.4.3 Indirect Reference: against indexed constraints and readjustment rules

Section 2.4.2 explored the far-reaching implications of Morph Integrity, which limits what morphology can do with phonological material; now I turn very briefly to the architectural principles that constrain the use of morphosyntactic information in phonology. First, the Four-Hypothesis Program assumes Indirect Reference:

(71) *Indirect Reference Hypothesis*
A phonological constraint may not refer to syntactic, morphological, or lexical information unless to require alignment between designated prosodic units and the exponents of designated syntactic (word-syntactic or phrase-syntactic) nodes.

The intuition behind Indirect Reference is as old as the prosodic hierarchy itself (e.g. Selkirk 1981: 388; see Scheer, 2010, for a historical overview), but (71) states it in particularly strong terms. First, (71) enforces Indirect Reference in all phonological cycles and not just at the phrase level (Inkelas 1989[1990: 29]).[58] Second, it denies the existence of 'morphophonemic' or 'phonosyntactic' processes, i.e. of phonological processes preceding prosodification within a cycle and directly referring to nonphonological information (cf. e.g. Selkirk 1986: 373–4, and see the discussion of readjustment rules below). Third, it exploits the fact that, in an optimality-theoretic approach to prosodification, access to syntactic structure may be reserved for constraints of a specific formal type: namely, alignment constraints.

Under (71), then, every phonological constraint that mentions a nonphonological object must conform to the following schema (see McCarthy and Prince 1993a: (4)):

(72) a. ALIGN (Cat_1, $Edge_1$; Cat_2, $Edge_2$)
b. One member of the set {'Cat_1', 'Cat_2'} is the label of a prosodic unit: e.g. 'μ', 'σ', 'Σ', 'ω', etc.
c. One member of the set {'Cat_1', 'Cat_2'} is the label of a syntactic node: e.g. 'stem', 'word', 'affix', 'X°', 'XP', etc.

In this schema, the syntactic label serves to select certain pieces in the phonological input representation: namely, those pieces exposing syntactic nodes of the designated category. The leftmost or rightmost output correspondents of these selected input pieces locate one of the edge types over which the alignment constraint ranges; the other edge type is located by reference to the initial or final elements of the designated prosodic units. The constraint itself then requires that, in output representations,

[58] On the notions of 'word syntax' and 'phrase syntax', see note 38 above.

every edge of one of the two types should coincide with an edge of the other: see again the discussion of Align(stem,R;ω,R) in section 2.4.1, specially diagram (40). Crucially, the schema provides that, if Cat_x is nonphonological, then it must **exhaustively** consist of a syntactic node label: this has the effect of forbidding parochial alignment constraints like Align([*ka*]$_{Af}$,L;Σ_s,R), for one is not allowed to identify a specific morph either through an arbitrary lexical index (e.g. '645') or by simultaneously listing both its syntactic label and its phonological content (e.g. '[*ka*]$_{Af}$'); cf. McCarthy and Prince (1993a: (5c)). I further assume that the expressive power of alignment constraints is curbed by a theory of syntactic representations distinguishing between the label of a syntactic node and its feature content (see Selkirk 1986: 385–6), but I cannot go into details here; for word-syntax, an example of such a theory would be Selkirk (1982).

Could one go further and forbid absolutely all reference to morphosyntactic information in the phonology? Two considerations speak against this possibility. First, there is ample evidence that cyclicity and stratification cannot by themselves account for all morphosyntactic conditioning of phonological processes; rather, the cycle shares this task with prosody (Scheer 2010: §6ff., Bermúdez-Otero 2011: §1, §4). Distinguishing between cyclic and prosodic effects can be difficult in practice, but analysts can rely on some robust criteria: e.g.

- a cyclic domain must be exactly coextensive with some syntactic category, whereas a prosodic unit need not be;
- morphological boundaries visible within a cyclic domain become invisible in the next cycle (§2.4.4), but prosodic structure remains accessible;
- prosodic units are visible to gradient processes of phonetic implementation, whereas cyclic domains are not.

Bermúdez-Otero and Luís (2009) provide discussion and a detailed case study. Second, there are conceptual and empirical arguments against building prosodic structure in a grammatical component preceding phonology proper (cf. Selkirk 1981: 387). In the context of the Four-Hypothesis Program, this option is ruled out by Morph Integrity, which aims to ensure that only phonology manipulates elements of phonological representation, including prosodic nodes (§2.4.2.1). Empirically, one finds that constraints on syntax–prosody alignment often lose out to purely phonological structural requirements. With just two possible exceptions, for example, Dutch word-level suffixes (identifiable by their stress-neutral behavior) consistently project a separate prosodic word **unless** they are unable to satisfy ω-minimality (Booij 1995: 111–12). If prosodification took place in a module preceding phonology proper, the Dutch pattern would require either extensive duplication of generalizations between this module and the phonology, or extensive back-tracking, with phonology having to repair the results of structurally blind alignment.

Even though it allows certain phonological constraints to refer to nonphonological information, the Indirect Reference Hypothesis as formulated in (71) outlaws all the devices whereby mainstream OT generates morph-specific phonological behavior:

see (61) above. The most extreme of these is Pater's (2000, 2010) theory of constraint indexation. Like construction-specific cophonologies, lexically specific constraints suffer from a severe lack of empirical content and heuristic power. In Pater's theory, for example, there is nothing to stop one from taking the English synthetic comparatives *lon*[g]*-er, stron*[g]*-er,* and *youn*[g]*-er* at face value, and so from erroneously annotating the comparative suffix *-er* with a different index from other word-level suffixes like *-ish*; cf. the discussion of (65) in section 2.4.2.3. Similarly, Pater (2000: 260) can easily dispatch the contrast between *ìmp*[ɔ̀]*rtátion* and *trànsp*[ə]*rtátion* by assigning different indices to these two words, but this solution fails to predict the effect of relative token frequency, for indexation, unlike blocking, has no inherent link to frequency; cf. the account of (33) in section 2.3.3.3. However, constraint indexation assumes the noncyclic architecture of mainstream OT and, in consequence, its problems far exceed those of Cophonology Theory. In mainstream OT, for example, the exceptional failure of the Abracadabra Rule in *apòtheósis* would be imputed to a high-ranking lexically specific clone of IO-Max-Head(Σ), whereas the absence of an initial dactyl in *orìginál-ity* would be attributed to OO-correspondence with *oríginal*, enforced by high-ranking OO-Max-Head(Σ); see Benua (1997: 38) and cf. section 2.3.3.2. Since IO-faithfulness and OO-identity constraints can be ranked independently, compliance with Chung's Generalization—of which we have seen so many examples in this chapter—becomes purely accidental. Moreover, Pater (2010: (22)) defines the domains of morphologically conditioned phonological processes in terms of total or partial overlap with indexed morphs: he proposes that, if the markedness constraint *[XYZ] is indexed to morph-class L, then it will penalize those tokens of [XYZ] that contain some object in correspondence with a morph of class L. Whilst this convention works for some (but not all) instances of nonderived environment blocking (cf. Inkelas 2000), there are morphologically sensitive phonological processes whose domain is clearly defined by the cycle, rather than by morph overlap. Inkelas (1998: 130–1) adduces the interesting case of tonally dominant affixes, which overwrite the tonal melody of the base: if a tonally recessive prefix occurs inside a dominant suffix, it undergoes tonal overwriting (73a); if it lies outside the scope of the dominant suffix, its tones will surface (73b).

(73) a. $[\![[\![\text{prefix}_{rec}$ - stem$]\!]\text{suffix}_{dom}]\!]$ → stem and prefix tones delete
b. $[\![\text{prefix}_{rec}[\![$stem - $\text{suffix}_{dom}]\!]]\!]$ → stem tones delete; prefix tones surface

By adopting the Indirect Reference Hypothesis, we have also deprived the readjustment rules proposed by Distributed Morphology of a home in the architecture of grammar. As an example, consider the following rule of past-tense ablaut proposed by Embick and Halle (2005: 41) for a subset of English strong irregular verbs (cf. §2.3.2, §2.3.4, and §2.3.5 above):

(74) /ɪ/ → /æ/ / X__Y[past]
X = {√Sing, √Ring, √Sink, √Begin, √Sit, . . .}

The Four-Hypothesis Program banishes this rule from the morphology: in the guise of a morphological rule, (74) would violate Morph Integrity, since it alters the phonological content of roots. Yet under the Indirect Reference Hypothesis this process cannot live in the phonology either, for it mentions specific morphosyntactic features, namely [past], and specific morphs, namely the roots in list X. As it happens, Embick and Halle (2005: 42) insist that readjustment rules like (74) apply in the phonology, but their claim is purely terminological, not substantive, because they do not mean by it to place readjustment rules under any of the restrictions that bind ordinary phonological rules: readjustment rules can effect arbitrary string transformations and, as in (74), can apply in environments defined by lists rather than structurally characterized natural classes. In fact, readjustment rules share with Andersonian word-formation rules the property of carrying out phonological transformations in morphologically defined environments, but Anderson's word-formation rules are meant to live in the morphology.

The proscription of readjustment rules is a welcome result. Embick and Halle (2005: §3) defend them by asserting that the idiosyncratic alternations described by readjustment rules ought to be kept separate from "outright suppletion of the *go/went* type"; but they fail to offer an independent **synchronic** criterion for drawing this distinction. More contentfully, Marantz (1997a), as reported in Siddiqi (2009: 42–3), claims that only functional items can display suppletion because innate principles biasing the child against synonymy (e.g. the Exclusivity Constraint of Markman and Wachtel 1988) impede the acquisition of suppletion in the open-ended lexical vocabulary. Marantz's claim is falsified by the findings of typological research (Brown et al. 2003) and ignores the possibility that children may use probabilistic clues (e.g. expected frequencies) to recognize suppletion among open-class items (see also Bonet and Harbour this volume: §6.3.2 for critical discussion).

The main argument against readjustment rules, however, is that they utterly destroy the empirical content of morphological and phonological hypotheses. For example, Embick (2010) agrees with Paster (2006) and Bye (2007) in seeking to reduce all phonologically driven allomorph selection to subcategorization, rather than optimization (cf. note 42, and Bonet and Harbour this volume). In principle, this proposal has empirical content: indeed, it predicts that outwards-sensitive allomorphy cannot refer to derived phonological properties. Embick acknowledges the counterexamples listed in Carstairs (1987: 179ff.), to which one should add the decisive Surmiran case documented by Anderson (2008, 2011; though cf. Maiden 2011). Revealingly, however, Embick dismisses this evidence as follows: "These cases do not appear to be fully suppletive; that is, it looks like the majority involve morphophonological rules, not competition for insertion" (2010: 203, note 24). Thus, Embick's theory, which initially appeared strong, turns out to have an extremely narrow empirical remit: in respect of root allomorphy, it confines itself to the relatively few cases of 'outright suppletion' admitted by Embick and Halle's (2005: §3) criteria, whatever these may be. Similarly, Embick's (2010: 87–91) discussion of French preposition-article portmanteaux eventuates in the postulation of readjustment rules like the following (notated orthographically):

(75) a. d(e)-e→ du

b. à-e → au

It is hard to imagine that a theory of grammar that allows the phonology to do such things as these could ever be falsified by any portmanteau behavior at all.

2.4.4 Phonetic Interpretability and cyclic locality

Phonological objects in output representations stand in correspondence with input pieces, but do not themselves participate in the exponence relationships linking input pieces to syntactic nodes (40). This postulate of the Four-Hypothesis Program entails that phonological output representations do not contain symbols of morphosyntactic affiliation such as labels or brackets. I take this restriction to be a special case of a more general ban on diacritics, emerging from a requirement of Phonetic Interpretability:

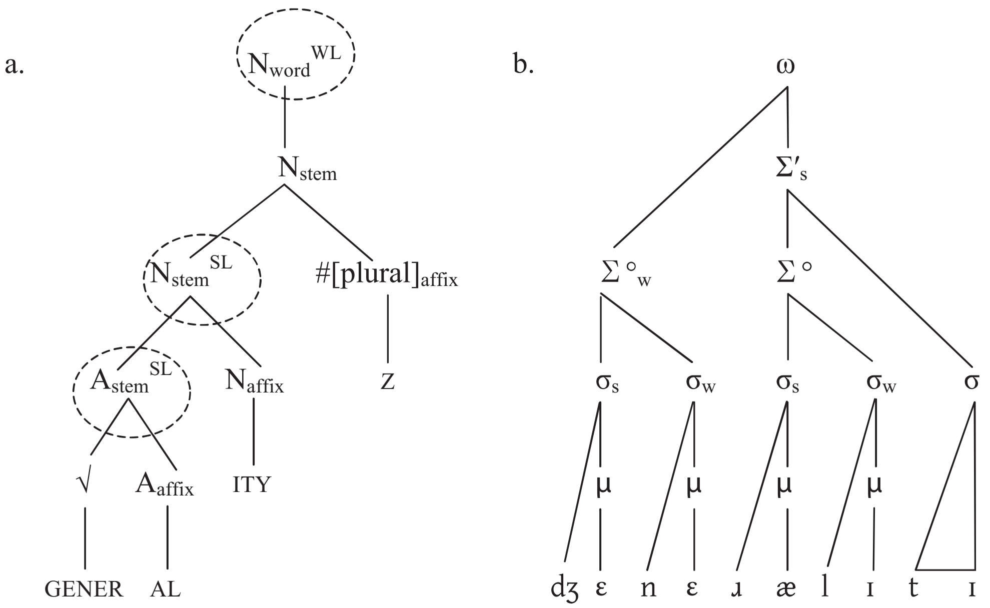
(76) *Phonetic Interpretability Hypothesis*
The contents of phonological output representations are phonetically interpretable.

This hypothesis presupposes a substantive theory of phonetics specifying what is a possible input to gradient processes of phonetic implementation. There is considerable evidence that such a theory will assign a range of phonetic interpretations to surface prosodic structure: as well as encoding categorical length and weight contrasts, for example, moric patterns have gradient timing effects (Broselow, Chen, and Huffman, 1997); the position of segments in syllable structure affects gestural magnitude and coordination (Gick 2003, Kochetov 2008); and higher prosodic categories influence F_0, duration, and amplitude (Gussenhoven and Rietveld 1992, Wightman et al. 1992, Fougeron and Keating 1997, Wagner and Watson 2010: 907–10, 925–6). In contrast, rules of phonetic implementation do not directly refer to morphosyntactic information; this statement has recently been challenged (Kawahara 2011: §2.3.3), but see Bermúdez-Otero (2010) for a defense.

I assume that the presence of indices of input–output correspondence in a phonological output representation, such as the subscript integers in (40c), does not violate Phonetic Interpretability because such indices are pure pointers, i.e. deictics without intrinsic content, which become inert as soon as the input representation to which they refer ceases to be accessible. For example, no extraphonological information will be conveyed to the phonetic module even if the surface representation (defined by the output of the phrase-level phonology) contains indices pointing towards correspondents in the phrase-level input: by the very definition of a modular serial interface given in (38) and (39), phonetics cannot see beyond the surface representation and so phonological input representations fall out of its view. McCarthy and Prince's (1995a, 1999) approach to reduplication provides the relevant term of comparison here. As we saw in (44), their theory requires that, in addition to indices of input–output correspondence, the surface representation should bear contentful morphosyntactic labels demarcating reduplicative from nonreduplicative material. In this system, it should be possible for gradient phonetic processes to refer directly to the morphological structure of reduplicative constructions.

Phonetic Interpretability has a number of beneficial effects beyond preventing morphosyntactic conditioning in phonetics. As we saw in section 2.4.1, for example, it deprives phonology of the means to violate Consistency of Exponence. Less obviously, it considerably strengthens the locality restrictions imposed by cyclic derivation. By itself, cyclicity produces outward locality effects, since information outside a cyclic domain is by definition unavailable to processes applying in that cycle: for an illustration, see the discussion of the Russian Doll Theorem in Bermúdez-Otero (2011: §3, §9). But, if the output of each phonological cycle abides by Phonetic Interpretability, then inward locality effects arise too. Consider, for example, the syntactic structure of the English word *generalities* as given in (77a), where the circled nodes trigger phonological cycles (the stem-level ones taking place off line: see §2.3.3.2). The input to the phonology in the last cycle consists of two pieces: the stem-level representation of the stem *generality* shown in (77b) (on which see note 19) and the regular plural suffix /-z/.

(77)



In this last cycle, the morphology will know that the stem GENERALITY contains the affix -ITY, since it has direct access to the syntactic representation in (77a). In the same cycle, the phonology will be able to see the phonological representation of GENERALITY in (77b) and its coindexation with a stem node. By Phonetic Interpretability, however, (77b) contains no diacritics of morphosyntactic affiliation and especially no brackets, and so at this point the phonology can no longer know about the presence of *-ity* inside *generality*, much less access the location of the juncture between *general-* and *-ity*. This result fits very well with Orgun and Inkelas's (2002) observations about so-called 'Bracket Erasure': in cases of potentiation (Williams 1981: 249–50), the morphology is able to check whether the subcategorization requirements of an affix are satisfied by the presence of another affix inside the base, even if the latter defines a cyclic domain; but the internal morphosyntactic structure of cyclic subdomains nonetheless remains invisible to the phonology. The accuracy of this subtle prediction, arising from the interaction of Phonetic Interpretability with the cycle, bears witness to the solidarity and interdependence among the components of the Four-Hypothesis Program.

2.4.5 Conclusion

The Four-Hypothesis Program (37) recommends itself to our attention for three good metascientific reasons. First, the four hypotheses, individually and in various combinations, forbid large sets of conceivable states of affairs: i.e. they make empirical predictions. This is a good thing if the theory of the morphology-phonology interface is to have strong empirical content (Popper 1959).

Second, the logical problem of language acquisition as it arises in the realm of morphology–phonology interactions becomes less daunting if the four hypotheses hold true, for they entail that the learner does not in fact contemplate many of the grammatical descriptions that seem *a priori* compatible with a given set of primary data (cf. §2.2). Yet, importantly, the four hypotheses do so without forcing us to make onerous commitments to representational innateness beyond the arrangement of morphosyntactic exponenda in hierarchical constituent structures and the existence of the morph as the unit of exponence (§2.4.2.1); this is because the four hypotheses set forth claims not about the detailed content of morphological and phonological representations or rules, but rather about the modular architecture of the grammar and the flow of information between its components.[59] Moreover, the architecture itself need not be hardwired in its entirety; many of its aspects may emerge, either from more fundamental mechanisms (as demonstrated in §2.3.3.2 and §2.3.3.3 for internal cyclicity in stem-level constructs) or from timing effects in development (as adumbrated in the two final paragraphs of §2.4.2.3 for interstratal cyclicity; see further Bermúdez-Otero forthcoming).[60]

Third, precisely because they have empirical content, the four hypotheses also have heuristic power. In face of particular phenomena, they not only instruct us to try out certain analytic moves, but, more importantly, they often force us to search more widely and deeply in order to deal with troublesome observations that could be taken at face value in less constrained frameworks. Similarly, they can perform a useful function as benchmarks for theory comparison. Of any theory that rejects one of these hypotheses we may fairly ask: Where is the evidence that the additional power is needed, and why is it impossible to describe the facts in a more constrained way? In this respect, it is essential that the Four-Hypothesis Program should bind morphology and phonology in equal measure: it is trivially easy to keep a beautiful garden on one side of a fence if we can uproot every ugly weed and throw it over on the other side. As Robert Frost reminds us, "Good fences make good neighbors."

[59] A commitment to representational nativism may be described as 'onerous' to the extent that it imposes on evolutionary linguistics the task of accounting for the phylogenesis of the proposed innate representations and that this task is hard. On the distinction between representational and architectural innateness, see Elman et al. (1996: 25–31) and Elman 1999: 3–5).

[60] See also Karmiloff-Smith (1992, 1998) on the notion of emergent modularity.