CHAPTER 8

PHONOLOGY

GUNNAR ÓLAFUR HANSSON

8.1 Introduction

The interface between the subsystems of phonology and morphology (inflectional and derivational alike) remains a productive area of research and debate within theoretical phonology. The advent of constraint-based frameworks within generative phonology, in particular Optimality Theory (Prince and Smolensky 2004), and the general shift of the theoretical mainstream from serialist transformation-based models to (largely) parallelist constraint-based models, has had a considerable impact on current thinking with respect to morphology–phonology interactions.

For example, the strong drive toward parallel, global, and monostratal computation in (standard) Optimality Theory puts into sharp focus any phenomena that have traditionally suggested the need for serial, cyclic, or multistratal derivations: 'cyclicity effects', construction-specific phonology, and the like. Furthermore, the formal notion of correspondence, and the possibility of extending correspondence constraints beyond the input–output dimension (underlying to surface representation) to the output–output dimension (relations among surface forms) has invited the explicit formalization of such traditional notions as analogical levelling and paradigm uniformity within a generative framework. Finally, the output-oriented character of Optimality Theory, in which all phonological computation is seen as inherently optimizing, has prompted a re-evaluation of many areas in which morphological processes are subject to phonological conditioning (e.g. allomorph selection, as well as other types of 'unusual' exponence; see Trommer and Zimmerman, this volume). This re-evaluation has gone hand-in-hand with a renewed focus on cross-linguistic typology, from which the field has benefited greatly.

This chapter will first briefly review current approaches to modelling the morphology–phonology interface, and the mechanisms by which information about morphological structure can affect phonological patterns (Section 8.2). We then turn
to two topics of particular interest with respect to the mutual interaction of morphology and phonology. The first of these is the interplay of paradigms and paradigm structure with phonology (Section 8.3). This includes the levelling of phonological alternations (paradigm uniformity effects), the blocking or triggering of phonological processes due to homophony avoidance (paradigm anti-homophony effects), and the occurrence of paradigm gaps which appear to be phonologically motivated. Secondly, we will examine in detail the phenomenon of phonologically conditioned suppletive allomorphy (Section 8.4), in which the direction of influence is from phonology to morphology rather than the reverse.

The morphology–phonology interface is an expansive and complex domain of inquiry, and many of the central theoretical questions cut across the divide between inflection and word formation. This chapter makes no pretence at exhaustive coverage, but rather focuses on a few theoretical issues and empirical phenomena which are of particular relevance for inflectional morphology. For other recent handbook-chapter overviews of the morphology–phonology interface in general, or of specific topics addressed within the present chapter, see Albright (2011), Bermúdez-Otero (2011, 2012), Inkelas (2011), Kaisse and McMahon (2011), Nevins (2011), Trommer (2011b), and Trommer and Zimmerman (this volume).

8.2 MODELLING THE MORPHOLOGY–PHONOLOGY INTERFACE

The phonology of a language \( L \) can be thought of as a function \( \phi_L \) from an input to an output representation: \( \phi_L(X) = Y \) thus denotes that in language \( L \) an input (lexical, underlying, phonemic) representation \( /X/ \) surfaces as \( [Y] \). (Henceforth the subscript \( L \) will be omitted in the interest of typographical clarity.) In current generative models \( \phi \) not only encodes allophonic detail but can also cover various neutralizing (feature-changing) processes, deletions, insertions, and prosodic structure-building such as syllabification, stress assignment, and so on.\(^1\) Similarly, we can take morphology to be a function \( \mu \), which can at least in some cases be seen as composed of a series of nested functions, \( \mu(\cdot) = \mu_n(\mu_{n-1}(\ldots(\mu_2(\mu_1(\cdot))))), \) where each \( \mu_i \) represents one derivational step or operation (e.g. affixation).

There are various ways in which one might conceptualize the manner by which phonology and morphology interact, such that phonological patterns and processes can be sensitive to aspects of morphological structure or vice versa. At one extreme, we might imagine that all morphology logically precedes all phonology in the sense that the input to the phonological grammar is a fully structured (and fully 'spelled-out')

\(^1\) Note that this formulation abstracts away from the question of how \( \phi_L \) is best modelled formally—for example, in terms of serially ordered transformational rules, or ranked and violable output constraints—which is of course a matter of central importance to phonological theory.
object which contains, in addition to an underlying phonological representation, all
details regarding the internal morphological make-up of the word form in question
(constituent structure, morphosyntactic feature specifications, and so forth). The
role of the phonology is then simply to compute an output (surface) phonologi­
cal representation of this object. Viewed from the function perspective, this model
can be depicted as \(\phi(\mu(\cdot))\) or more accurately as \(\phi(\mu_n(\ldots(\mu_1(\cdot))))\). In such a
model, there are in principle no limitations on the ways in which the constraints
or rules of phonology might make reference to, and thus be sensitive to, morpho­
logical factors. We thus expect extensive influence of morphology on phonology
\((M \rightarrow P)\), but there should be little or no influence of phonology on morphology
\((P \rightarrow M)\), though morphology might perhaps be sensitive to the lexical phonologi­
cal properties (that is, ones not derived by \(\phi\)) of the stems and exponents that it
manipulates.

At the other extreme, we can imagine that all phonological and morphological
structure is assembled and computed in parallel, such that each is fully capable of
influencing the other. If implemented in terms of a constraint-based architecture such
as Optimality Theory (McCarthy 2002, 2007; Prince and Smolensky 2004), this is in
essence a non-modular model, in which constraints that reference either morpholog­
ical or phonological content (or both at once) are fully interspersed and each type
of constraint can outrank the other. It is thus not only possible for consideration
of morphological factors to affect phonological patterning \((M \gg P)\) but likewise for
aspects of morphology (e.g. exponence, affix order) to be affected by constraints on
phonological well-formedness \((P \gg M)\). The full-parallellism model differs from the
all-morphology-precedes-phonology model primarily in that there is a much less clear
distinction between the functions \(\phi\) and \(\mu\), with much of what is normally attributed
to the latter being instead folded into the former (see Embick 2010 and Trommer and
Zimmerman, this volume).

In practice, existing approaches to modelling the interface and interaction between
phonology and morphology fall somewhere between these two straw-man extremes.
Each of the two is at once too permissive and unconstrained and too limited in its
abilities to adequately capture the richness of attested phenomena. In what follows,
we first (Section 8.2.1) consider a class of models that incorporate morphol­
ogy–phonology interleaving, such that phonology is applied and re-applied in a cyclic
fashion at intermediate nodes in the constituent structure. In models of this type,
construction-specific phonology is accounted for by allowing successive cycles to call
on distinct phonology functions, or co-phonologies (levels, strata). The remaining
subsections critically examine various alternative proposals for non-interleaving and
monostratal accounts of some of the same empirical phenomena. Section 8.2.2 dis­
cusses the use of constraint indexation, as well as reference to internal morphosyntactic
constituent boundaries, as a means of capturing construction-specific phonological
patterns. Section 8.2.3 considers output–output correspondence as an alternative to
interleaving in accounting for so-called cyclicity effects (see also Section 8.3.1 on the
related notion of paradigm uniformity).
One influential model of morphology-phonology interaction, which had its roots in Chomsky and Halle’s (1968) cyclic account of English stress but was more fully developed in the Lexical Phonology model (Kiparsky 1982a, 1982b; Kaisse and Shaw 1985; Mohanan 1986; Kaisse and McMahon 2011) and its constraint-based descendants, takes the phonological interpretation of complex morphological structures as involving an interleaving, or layering, of phonology and morphology. There is thus not merely a single call to the phonology function $\phi$ at the very end of the derivation of a morphologically complex form. Rather, the derivation may involve one or more internal passes through phonology as well: $\phi(\mu_n(\ldots(\phi(\ldots(\mu_1(\ldots))))))$. In other words, application of some phonological processes can precede certain aspects of morphology, creating the potential for various interactions and interference—between phonology and morphology, and between inner and outer applications of phonology—which would otherwise be inconceivable.

In a simplified view of morphology as involving only morpheme concatenation, interleaving can be illustrated schematically as in (1), where phonological interpretation takes place at each non-terminal node in a constituent-structure tree. Here \{a, b, c, d\} represent individual affix morphemes, of which \{a, b\} are cyclic, each defining a domain for phonological interpretation, whereas \{c, d\} are non-cyclic. At each node, $M$ stands for its morphological constituent structure (category labels on constituents are left out for simplicity) and $\Phi$ for its phonological content; $\sim$ is the string-concatenation operation.

\[
\Phi = \phi_k(\phi_j(b \sim \phi_i(X \sim a)) \sim c \sim d)
\]

\[
M = [ [ b [ X a ] ] c d ]
\]

The innermost constituent, $[X a]$, has its phonological properties computed by applying phonology ($\phi_i$) to the concatenation of the underlying phonological representations of the root ($X$) and the innermost suffix ($a$): $\phi_i(X \sim a)$. This derived
representation in turn acts as one of the two concatenated terms at the next level up, where the prefix (b) is attached and phonology is once again applied. The output of the phonological computation at this node in the tree is therefore \( \phi_i(b \sim \phi_i(X \sim a)) \). Finally, this representation is concatenated with those of the two outermost suffixes (c, d) and phonology is applied yet again. The output phonological representation of the word form as a whole is therefore defined as \( \phi_k(\phi_j(b \sim \phi_i(X \sim a))) \), with three separate, and nested, calls to phonology.

In (1), the successive applications of phonology have been labelled \( \phi_i, \phi_j, \phi_k \). If all of these are one and the same function, \( \phi_i = \phi_j = \phi_k = \phi_L \), then what we are seeing is pure cyclicity: the cyclic application of a single phonology function (constraint ranking, or set of rules) as dictated by the morphological structure. In practice, however, interleaving models generally presuppose the co-existence of more than one phonological sub-grammar. Models differ with respect to whether any limit is placed on how many such co-existing phonologies, 'cophonologies', the grammar can contain or on how these may differ from each other, and whether the cophonologies are assumed to be strictly ordered relative to one another.

Stratal Optimality Theory (Kiparsky 2000, 2008a; Bermúdez-Otero 2003, 2012), the constraint-based implementation of Lexical Phonology (Kiparsky 1982a, b; Mohanan 1986), posits exactly three distinct levels, or strata, of phonology: Stem, Word, and Phrase. These are strictly ordered, and any phonological derivation must involve at least one call to each of the three: \( \phi_L(\cdot) = \phi_{\text{Phrase}}(\ldots(\phi_{\text{Word}}(\ldots(\phi_{\text{Stem}}(\ldots(\cdot)))))) \). For example, Kiparsky (2000) draws upon the Stem vs. Word distinction to account for the famously opaque interaction of syncope, stress, epenthesis, and morphological structure in Levantine Arabic (Brame 1974; Kenstowicz 1983). The basic pattern of stress assignment is that a penultimate syllable will attract stress if it is heavy (e.g. CVC or CV:;) but otherwise stress falls on the antepenultimate syllable. Syncope is sensitive to stress placement, targeting short /i/ in unstressed open (and hence light) syllables. Finally, complex syllable onsets or codas are not permitted, and epenthesis of [i] breaks up otherwise unsyllabifiable consonant clusters of three or more consonants.

The puzzle, which is illustrated by the minimal triplet in (2), is twofold. First, assignment of stress to the penultimate syllable triggers syncope of unstressed /i/ before a subject marker like 1.PL /-na/ (2a), but mysteriously fails to do so before the corresponding object or possessive markers (2b). Second, epenthesis into a morphologically derived triconsonantal cluster, such as when 1.PL /-na/ is attached to the noun /fihm/ (2c), fails to trigger stress shift to the penultimate epenthetic vowel, despite the fact that it is in a heavy syllable.

(2) Strata and cycles in Palestinian Arabic

\begin{align*}
\text{a. } ['fhim-na] & /fihim-na/ 'we understood' & \text{[ [ [ fihim ][St na] ][St ][Wd ][Phr] ]} \\
\text{b. } [fi.'him-na] & /fihim-na/ 'he understood us' & \text{[ [ [ fihim ][St ][Wd na] ][Wd ][Phr] ]} \\
\text{c. } ['fi.him-na] & /fihm-na/ 'our understanding' & \text{[ [ [ fihm ][St ][Wd na] ][Wd ][Phr] ]}
\end{align*}

(Kiparsky 2000)
Kiparsky's (2000) stratal analysis relies on the insight, supported by a variety of other sound patterns as well, that object/possessive markers in Levantine Arabic are attached at the word level whereas subject markers are stem-level (see, e.g., Kenstowicz 1983), as shown by the bracketing in the rightmost column of (2). Given certain differences in constraint ranking between the three strata, the facts fall out correctly. The failure of epenthesis to affect stress in (2c) is explained by the fact that epenthesis is a Phrase-level process (cf. /fih.m il.-'wa.lad/ → ['fih.m il.-'wa.lad] 'the boy's understanding', not *[fi.him. il.-'wa.lad]), whereas weight-sensitive stress assignment is (re)computed solely at the Stem and Word levels. That is, Phrase-level phonology simply preserves the input (Word-level) stress pattern faithfully. The failure of syncope to apply in (2b) arises from the fact that syncope, a Word-level process, is blocked from deleting vowels that are stressed in the input (i.e. ones which have been assigned stress in the Stem-level phonology), due to a high-ranked faithfulness constraint $\text{MAX-}\text{V-IO}$.

In terms of phonology, the derivational history of the three forms in (2) is then as shown in (3). Note that in each case, the affixation of /-na/ triggers a new cycle through the phonology of the relevant (Stem or Word) level. Note also that the root-and-template morphology of Arabic, which combines the triconsonantal root /fhm/ with the appropriate CVCVC or CVCC template and vocalism, is likewise Stem-level but has been omitted here for simplicity.

(3) 'we understood' 'he understood us' 'our understanding'

<table>
<thead>
<tr>
<th>Stem:</th>
<th>'fi.him</th>
<th>'fi.him</th>
<th>'fihm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stem (2nd cycle):</td>
<td>fi.'him.na</td>
<td>(n/a)</td>
<td>(n/a)</td>
</tr>
<tr>
<td>Word:</td>
<td>'fhim.na</td>
<td>'fi.him</td>
<td>'fihm.na</td>
</tr>
<tr>
<td>Word (2nd cycle):</td>
<td>(n/a)</td>
<td>fi.'him.na</td>
<td>'fihm.na</td>
</tr>
<tr>
<td>Phrase:</td>
<td></td>
<td></td>
<td>fi.'him.na</td>
</tr>
<tr>
<td>Output:</td>
<td>['fih.m.na]</td>
<td>[fi.'him.na]</td>
<td>['fi.'him.na]</td>
</tr>
</tbody>
</table>

An alternative to stratal Optimality Theory, cophonology theory (Orgun 1996, 1999; Inkelas 1998, to appear; Inkelas and Orgun 1998; Anttila 2002; Inkelas and Zoll 2005, 2007), abandons several of its restrictive assumptions and generalizes the stratum notion by extending it to cover all construction-specific phonological patterns. In this model, each morphological construction is associated with some phonological function (constraint ranking), a cophonology, and there is in principle no limit on how many cophonologies can co-exist within the grammar, nor how many individual morphological constructions 'subscribe' to any given cophonology. Importantly, cophonologies are not assumed to be subject to any crucial ordering. For some pair of cophonologies $\phi_i \neq \phi_j$, the fact that the order $\phi_j(\ldots(\phi_i(\ldots)))$ occurs does not automatically preclude the possibility of the reverse order $\phi_i(\ldots(\phi_j(\ldots)))$, as long as the morphological structure so dictates. Furthermore, because cophonologies are tied directly to morphological constructions, their application is economical (in the sense
of Inkelas and Orgun 1995): unlike in stratal Optimality Theory, there is no requirement that a given word form must be subjected to each cophonology (level, stratum) in the language.

Finally, there is no stipulated (theory-internal) limit on the extent to which cophonologies, or the levels of stratal Optimality Theory, may in principle diverge. Such divergence will in any event be constrained in the diachronic domain of language transmission and change (on this point see also Bermúdez-Otero and McMahon 2006; Pater 2010). Anttila (2002) approaches the question of shared vs. non-shared properties among cophonologies by modelling the phonological grammar as a whole as an inheritance hierarchy. Each node in the hierarchy defines a particular constraint ranking (cophonology), inheriting and combining ranking information from all higher nodes, such that the default or unmarked state of affairs is for cophonologies to be quite similar. In Anttila's account of the declension of Finnish /a/-final nominals, the node to which a given class of nominal belongs determines the choice between mutation (/a/ → [ɔ]) or deletion (/a/ → Ø) before plural /-i/-, the degree to which that choice is sensitive to syllable count (even- vs. odd-numbered stems) and/or the quality of the penultimate stem vowel, and the extent to which the choice is variable.

8.2.2 Parallellism: Constraint Indexation, Boundary Reference

A different line of scholarship eschews morphology–phonology interleaving and cophonologies in favour of a monostratal model (i.e. one employing a single constraint ranking), in which apparent interleaving effects are instead accounted for by other means. In the Optimality Theory tradition, a popular alternative to strata or cophonologies is to allow for certain constraints to be indexed to particular morphological constituents or lexical categories, or even to individual morphemes, lexemes or lexical strata (McCarthy and Prince 1995; Itô and Mester 1999; Pater 2000, 2010; Alderete 2001; Smith 2011). In the most common instantiation of this approach, only correspondence constraints can be indexed in this manner, in particular those regulating faithfulness (input–output correspondence).

As an illustrative example, consider the two different strategies for hiatus resolution that are employed in Turkish in different morphological contexts (Inkelas and Zoll 2007). The general pattern is epenthesis of the glide [j] between the stem-final and suffix-initial vowels, as illustrated in (4). However, before the progressive /-ljo:/ suffix, the stem-final vowel instead deletes, as seen in (5).²

² The archiphonemic symbols /I A/ denote a vowel that is underlyingly specified as [+high] or [−high], respectively, but whose value for backness (and in the case of /U/, rounding) is determined by harmony to the vowel in the preceding syllable.
(4) Turkish: glide epenthesis as default hiatus resolution strategy

a. [adam-a] /adam-A/ man-DAT 'to a/the man'
   [kapu-jə] /kapu-A/ door-DAT 'to a/the door'
b. [ev-i] /ev-I/ house-DEF.ACC 'the house (obj.)'
   [oda-jwu] /oda-I/ room-DEF.ACC 'the room (obj.)'
c. [koʃ-adʒak] /koʃ-AdʒAk/ run-FUT.PTCP 'he will run'
   [anla-jaŋak] /anla-AdʒAk/ understand-FUT.PTCP 'he will understand'
d. [al-tundʒa] /al-IndʒA/ take-CVB 'upon taking'
   [dene-ʃinjge] /dene-IndʒA/ try-CVB 'upon trying'

(5) Turkish: hiatus resolution by vowel deletion in progressives

a. [jap-ujor] /jap-Ijor/ make-PROG 'he is making'
   [koʃ-ujor] /koʃ-Ijor/ run-PROG 'he is running'
b. [anl-ujor] /anla-Ijor/ understand-PROG 'he is understanding'
   [oʃ-ujor] /oʃe-Ijor/ pay-PROG 'he is paying'
   [koʃ-m-ujor] /koʃ-mA-Ijor/ run-NEG-PROG 'he is not running'

As Inkelas and Zoll (2007) note, these facts could easily be captured in an indexed-constraint analysis, by assuming a constraint ranking such as that in (6). Here NoHIATUS stands for whatever markedness constraint is responsible for prohibiting hiatus in Turkish. (This may well be the syllable structure constraint ONSET, which bans onsetless syllables.) As input-output correspondence constraints, MAX-IO penalizes deletion whereas DEP-IO bans epenthesis (McCarthy and Prince 1995).

(6) \{ NoHIATUS, DEP-IO_{PROG} \} \gg MAX-IO \gg DEP-IO

The indexed constraint DEP-IO_{PROG} is relevant only for segments contained within a Stem + /-Ijor/ constituent. That constraint thus has the effect of preventing [j]-epenthesis in progressive forms specifically, such that the alternative repair strategy of vowel deletion emerges as optimal (7). In non-progressive contexts, where the indexed constraint is irrelevant (and hence vacuously satisfied), the default repair of [j]-epenthesis occurs (8).

(7) /anla-Ijor/ \rightarrow [anl-ujor] 'he is understanding'

<table>
<thead>
<tr>
<th>/anla-Ijor/</th>
<th>NoHIATUS ; DEP-IO_{PROG}</th>
<th>MAX-IO</th>
<th>DEP-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. an.la.u.jor</td>
<td>*!</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. an.laJu.jor</td>
<td>!</td>
<td>*!</td>
<td>*</td>
</tr>
<tr>
<td>c. *an.lu.jor</td>
<td>!</td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>
(8) /anla-IndʒA/ → [anla-jʊndʒa] ‘upon understanding’

<table>
<thead>
<tr>
<th>/anla-IndʒA/</th>
<th>NoHIATUS</th>
<th>DEP-IO_PROG</th>
<th>MAX-IO</th>
<th>DEP-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. an.1a.un.đa</td>
<td>* (\uparrow)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. an.1a.jun.đa</td>
<td></td>
<td></td>
<td>* (\uparrow)</td>
<td></td>
</tr>
<tr>
<td>c. an.1un.đa</td>
<td></td>
<td></td>
<td></td>
<td>* (\uparrow)</td>
</tr>
</tbody>
</table>

In this way, constraint indexation is able to mimic what in a cophonology analysis would be a matter of different ranking relations among the same set of three constraints: \{ NoHIATUS, DEP-IO \} \(\Rightarrow\) MAX-IO in the progressive-construction cophonology, and \{ NoHIATUS, MAX-IO \} \(\Rightarrow\) DEP-IO otherwise. The two approaches are not merely notational equivalents of one another, however. In particular, cophonology theory is embedded in an interleaving model of the morphology-phonology interface (cf. (1) above). Constituent structure and nesting relations are thus predicted to dictate the interaction and trade-off between the cophonologies (rankings) that are called upon in the derivation of a morphologically complex word form. Consider two morphological operations \(\mu_j\) and \(\mu_k\) (e.g. prefixation and suffixation, or zero-derivation and affixation). If the morphology of the language allows either to be nested within the other, such that the constructions \(\mu_j(\mu_k(\cdot))\) and \(\mu_k(\mu_j(\cdot))\) both occur, resulting in the same linear string of morphemes, cophonology theory predicts the phonological outputs for the two to be different. This is due to the different nesting relations between the relevant cophonologies: in the former case, \(\phi_j\) applies to the output of \(\phi_k\), and vice versa in the latter. In either case, the ‘outer’ cophonology can be sensitive to properties induced by the ‘inner’ cophonology, or potentially overwrite them.

One example of this sort of interaction is found in stress patterns associated with derived place names in Turkish (Inkelas and Orgun 1998, 2003; Inkelas 1999). In Turkish, the default word stress falls on the word-final syllable. However, a productive zero-derivation process, by which any word can be converted into a place name, is associated with a special cophonology in which the word (if not already stressed) receives stress on the antepenultimate syllable if it is heavy (CVC) and the penultimate syllable is light (CV), or else on the penultimate syllable. This results in pairs of place names and the words from which they are derived which differ only in the location of stress, such as [ba.'ka.đak] Bakacak (place name) vs. [ba.ka.'đak] ← /bak-AbđAk/ (look-FUT) ‘s/he will look’, or [sir.ke.'đi] Sirkeci (place name) vs. [sir.ke.'đi] ← /sirke-'đi/ (vinegar-AGT) ‘vinegar seller’. In addition, Turkish also has a large set of suffixes that subscribe to a different, pre-stressing cophonology, which assigns stress to the syllable immediately preceding the suffix. This includes /-đʒA/ MITIGATIVE ‘kind of; in the manner of; (in the) language of’, e.g. [syt.'ly.đże] ← /syt-II-đʒA/ (milk-ASSOC-MIT) ‘kind of milky’. The crucial evidence comes from cases where a pre-stressing suffix is either attached to a derived place name, or is contained within it, as the two cophonologies may then place conflicting demands on stress assignment. For example, the place name Çamlıca is derived from a form with the same internal structure as [syt.'ly.đże] just mentioned (/tfam-II-đʒA/, pine-ASSOC-MIT); it retains that same stress
pattern, [\textipa{\texti{\textam{\textae}}\texti{\textam{\textae}}}]), rather than adopting the expected place-name stress pattern (*[\textipa{\texti{\textam{\textae}}\texti{\textam{\textae}}}]). Conversely, if the pre-stressing suffix /-\textipa{\textg{\textae}}\textipa{\textg{\textae}}/ is attached to a place name like Kavakh [ka.\textipa{\textvak.\textlu}] a zero-derivation from /kavak-li/ (poplar-ASSOC) 'with poplar', it is the place-name stress pattern that prevails, [ka.\textipa{\textvak.\textlu}] 'in the' language/dialect of Kavakh', not the pre-stressing one (*[ka.\textipa{\textvak.\textlu}] which would have been analogous to [\textipa{\texti{\textam{\textae}}}\textipa{\texti{\textam{\textae}}}]).

The observed outcome is thus in all cases consistent with the 'innermost wins' principle that appears to guide stress assignment throughout the Turkish lexicon (see Inkelas and Orgun 2003). This interaction pattern falls out straightforwardly in the cophonology model because of how that model takes phonology to be interleaved with, and therefore inevitably reflective of, constituent structure. All that is required to assume that stress assignment in Turkish is structure-building, applying only to forms that are not already stressed (e.g. due to containing a lexically stressed root or suffix; Inkelas and Orgun 2003: 141–2). As a consequence, whichever stress-assigning cophonology happens to be innermost determines stress placement (in the absence of lexical stress); any outer layers of phonology are unable to alter this in any way. As Inkelas (2011: pp. 74–75) points out, the constraint indexation model does not fare as well. Since it assumes a single constraint ranking, the constraints crucially associated with the pre-stressing pattern must either outrank or be outranked by the constraints that drive the place-name stress pattern (as well as those that define the default, word-final stress pattern). The erroneous prediction is thus that we should see either consistent pre-stressing ([\textipa{\texti{\textam{\textae}}}\textipa{\texti{\textam{\textae}}}], *[ka.\textipa{\textvak.\textlu}]) or consistent enforcement of the place-name stress pattern (*[\textipa{\texti{\textam{\textae}}}\textipa{\texti{\textam{\textae}}}], [ka.\textipa{\textvak.\textlu}]), regardless of scope relations between the zero-derivation and /-\textipa{\textg{\textae}}\textipa{\textg{\textae}}/ suffixation.

One argument that is sometimes adduced in support of the constraint indexation approach over interleaving approaches (cophonologies, strata) is that it is supposedly more restrictive, in that construction-specific phonology is predicted to exhibit what Alderete (2001) refers to as grammar dependence (see also Benua 1997; Itô and Mester 1999). Indexed correspondence constraints are interdigitated within the overall ranking of markedness and (other) correspondence constraints in the language, and this is claimed to restrict the degree of language-internal phonological variation that morphological constructions can exhibit; in particular, markedness reversals are claimed to be impossible. However, Inkelas and Zoll (2007) show how this gain in restrictiveness is largely illusory, and in any case crucially depends on the stipulation that indexation is limited to correspondence constraints. Much recent work, by contrast, has argued that indexation of markedness constraints must also be allowed for (Pater 2000, 2007, 2010; Ota 2004; Flack 2007).

A different way in which sensitivity to morphological structure can be captured in a monostratal model is by reference to constituent edges or boundaries. Explicit use

---

3 Thanks to Orhan Orgun for providing native-speaker judgements on the relevant place name forms cited in this paragraph.
of juncture or boundary symbols was commonplace in classical generative phonology (Chomsky and Halle 1968) as well as earlier structuralist analyses, and boundary reference remains a relevant mechanism for morphology–phonology interaction in current constraint-based frameworks. In particular, alignment constraints (McCarthy and Prince 1993b) can reference the edges of morphological constituents of various types—either directly or else indirectly via such ‘morpho-prosodic’ constituents as the prosodic word, stem, or root (PWord, PStem, PRoot; Nespor and Vogel 1986; Inkelas 1989, 1993; Downing 1998, 1999, 2006; Raffelsiefen 2005; cf. also Selkirk 1986).

For example, in an analysis of hiatus resolution strategies in Shona, Mudzingwa (2010) proposes that a noun class prefix combines with a following (non-derived) nominal stem to form a PStem constituent. However, derived nominal stems (e.g. deverbal agentives and infinitives) also constitute their own PStem domain, whereas non-derived stems do not, resulting in a recursive domain structure in the former case. A noun class prefix will therefore be separated from the following nominal stem by a PStem boundary only if the latter is derived (9b) but not otherwise (9a).

(9) Hiatus resolution and constituent structure in Shona (Mudzingwa 2010)

a. Non-derived nominal stems: [PWord [PStem Prefix Stem ] ]
   [mù.isé] /mù-isé/ cl3.sg-tail ‘tail’
   [vù.isé] /vù-ósé/ cl2.pl-all ‘all [the boys]’

b. Derived (deverbal) stems: [PWord [PStem Prefix [PStem Stem ] ] ]
   [mù.ji.mbi] /mù-imbi/ cl1.sg-sing-nmlz ‘singer’
   [vù.wo.ngo.ro.ri] /vù-öngörör-ri/ cl2.pl-spy-nmlz ‘spies’

This difference in morphological (and morpho-prosodic) structure correlates with a difference in how hiatus is repaired at the prefix–stem juncture. PStem-internal hiatus as in (9a) is resolved by elision, glide formation, or consonant-vowel fusion ([Cu/ → [Cw]), depending on the segmental context. Hiatus across a PStem boundary as in (9b), however, is consistently resolved by epenthesis (of [j], [w], [], or [], again depending on segmental context). Mudzingwa (2010) analyses this as being due to an alignment constraint ALIGN(PStem, L, σ, L) which demands that the left edge of a PStem coincide with the left edge of a syllable.

8.2.3 Cross-derivational Interactions

In the approaches discussed in the previous sections, the phonological component accesses information about morphological structure by means of making reference to morphological constituents as such, or to their edges, either directly or indirectly

4 The same type of recursion arises in noun-class prefix stacking (so-called pre-prefixation) as well. For example, [má.zi.wu.me] /má-zì-ú-mé/ (cl6.pl-cl21.aug-cl14-fruit) ‘big ume fruits’ has the domain structure [PWord [PStem má [PStem zì [PStem ú - mé ] ] ]].
This mode of morphology–phonology interaction can thus only refer to structural properties that are inherent in the word form in question; there is no explicit reference to other morphologically related words or word forms, such as other forms within the same inflectional paradigm.

In the context of Optimality Theory, alternative approaches have been developed which instead (or in addition) allow direct reference to another output word form that is a member of the same morphological paradigm, or even to the entire set of output forms that make up the inflectional paradigm of the word in question. The former of these proposals relies on a formal device usually referred to as transderivational, or output–output, correspondence (Benua 1995, 1997; Crosswhite 1998; Kager 1999: ch. 6: see also Burzio 1994, 1998, 2002a and the notion of base identity in Kenstowicz 1996). Constraints that demand correspondence between a given output form and a designated ‘base’ form in the same morphological paradigm can have the effect of either extending (overapplying) the effects of a phonological process that would otherwise not have been applicable, or conversely suppressing (underapplying) a process that ought to have affected the word form in question.

This is illustrated in (10), which depicts the Levantine Arabic case discussed in (2)–(3) above. The vertical arrows denote the dimension of input-to-output correspondence (faithfulness), the horizontal arrow the added dimension of base-to-affixed-form (output-to-output) correspondence. The difference between the forms ['fhim.na] ‘we understood’, in which syncope is enforced, and [fi.'him.na] ‘he understood us’, where it fails to apply, is that the second form is also subject to correspondence to the independently-occurring unsuffixed form ['fi.him] ‘he understood’. The forms ‘he understood’ and ‘we understood’, by contrast, do not constitute a base vs. affixed form pair in this sense, due to the fact that the morphosyntactic feature content of the former is not a proper subset of the latter (Kager 1999: ch. 6; though see Kiparsky 2000 on the problems with this criterion).

(10) ‘we understood’ ‘he understood’ ‘he understood us’
/fihim-na/ /fihim/ /fihim-na/

\[ \begin{array}{c}
\text{I-O corr.} \\
\downarrow \\
\text{['fhim.na]}
\end{array} \quad \begin{array}{c}
\text{I-O corr.} \\
\downarrow \\
\text{[fi.him]}
\end{array} \quad \begin{array}{c}
\text{B-A corr.} \\
\downarrow \\
\text{[fi.'him.na]}
\end{array} \]

Here the base ['fi.him] plays a similar role in the derivation of the affixed form [fi.'him.na] ‘he understood us’ as it did in the stratal (cyclic) account described in example (3). In the stratal account, ['fi.him] constituted an intermediate representation in the overall derivation: the output of the Stem stratum. This meant that the initial-syllable [i] was already stressed in the input to the Word stratum, and protected
against syncope in that stratum by a high-ranked input–output correspondence constraint \textsc{Max-V-IO}. The monostatal account portrayed in (10) utilizes that very same type of constraint, but defined instead over the output–output dimension: \textsc{Max-V-BA} (called \textsc{HeadMax-BA} in Kager 1999: 283).

One limitation of the transderivational correspondence approach is that it crucially requires that the source of influence (the ‘base’) occur independently as a fully fledged output form. This can create challenges in situations where the putative base is not a morphologically simplex form, as is often the case for inflectional morphology. A related problem is how the criteria for the unique identification of bases should be formulated, especially in inflectional paradigms. A related line of thinking avoids the problem of base identification by appealing to a broader and more symmetric concept of paradigm uniformity, or anti-allomorphy (Kenstowicz 1996; Steriade 2000; Burzio 2002a; McCarthy 2005). Since such proposals often explicitly invoke the notion of inflectional paradigm, they are discussed in Section 8.3.1.

### 8.3 The Phonology of Paradigms

#### 8.3.1 Levelling and Paradigm Uniformity

The most obvious way in which paradigmatic relations can have repercussions for the phonological structure of word forms is in the phenomenon of levelling, whereby an otherwise expected phonological alternation in the surface shape of some item (morpheme, stem, lexeme) is either suppressed or generalized throughout the inflectional paradigm. The Palestinian Arabic example in the preceding section can be seen as an instance of levelling, in that the syncope alternation that is evident in the subject-agreement paradigm (e.g. [fiihim] ‘he understood’, [fim-na] ‘we understood’, [fihm-u] ‘they understood’) is suppressed in the sub-paradigm defined by object-agreement suffixation ([fihim-na] ‘he understood us’, etc.). For the present purposes, the term ‘(paradigm) levelling’ will be used as a purely descriptive label, referring to any synchronic state of affairs that fits the above definition. As we will see toward the end of this section, it is far from certain that levelling exists as a distinct type of process in its own right.

Various works in the theoretical phonology literature propose that the synchronic source of at least some levelling phenomena is a demand for \textit{paradigm uniformity}: the uniform exponence of some morphological entity (e.g. the stem) throughout the paradigm (Steriade 2000; Kenstowicz 1996; Burzio 2002a; McCarthy 2005), rather than

\footnote{This is a slight oversimplification, in that it is only with respect to the vowel in the stem-initial syllable that syncope alternations are suppressed (the vowel that is stressed in the unsuffixed form [fiihim] ‘he understood’). Syncope applies as expected in the second syllable when its conditions are met: /fiihim-u/ $\rightarrow$ [f\textsubscript{\textmu}him\textsubscript{\textmu}] ‘he understood them.’}
identity to some designated base form. The optimal paradigms model of McCarthy (2005), for example, takes it as a given that inflectional paradigms do not have bases (for counterarguments, see Hall and Scott 2007; Albright 2008, 2010; Bobaljik 2008). Instead of evaluating the well-formedness of individual word forms in the paradigm against the benchmark of some designated base form within that same paradigm (as in Section 8.2.3), the overall phonological well-formedness of the entire paradigm is assessed all at once. The candidate outputs under evaluation by the phonology are thus not single word forms but whole sets of forms that together make up an inflectional paradigm. Paradigm uniformity is enforced by correspondence constraints that compare the stem portions of every pair of output forms that are contained within the (candidate) paradigm.

Consider, for example, a hypothetical language in which high vowels are obligatorily lowered before uvulars (*HiV/\_Uvu \(\Rightarrow\) IDENT[±high]-IO), and whose nominal inflection includes a plural suffix -q. In such a language, a high-ranked paradigm uniformity constraint IDENT[±high]-OP will have the effect shown in (11). The process of pre-uvular lowering overapplies, in that its effects become extended even to the unsuffixed form where no uvular is present.

(11) Levelling (by overapplication) due to paradigm uniformity

<table>
<thead>
<tr>
<th>/sani/ + {∅, q, ...}</th>
<th>IDENT[±high]-OP, *HiV/_Uvu</th>
<th>IDENT[±high]-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.  (sani, saniq, ...)</td>
<td>!</td>
<td>*</td>
</tr>
<tr>
<td>b.  (sani, saneq, ...)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>c.  (sane, saneq, ...)</td>
<td></td>
<td>**</td>
</tr>
</tbody>
</table>

As McCarthy (2005) points out, the inherently asymmetric nature of paradigm uniformity makes specific predictions regarding patterns of levelling in inflectional (as opposed to derivational) paradigms, predictions which would not be consistent with either a base-identity approach (Section 8.2.3) or a cyclic/stratal approach (Section 8.2.1). For example, the forms that act as attractors should be ones that are phonologically less marked. Furthermore, ‘majority rule’ effects are expected to be possible, in which the sheer number of forms exhibiting each alternant (each stem allomorph, say) will determine the direction of levelling. Finally, inflectional paradigms should never exhibit levelling that is manifested through under- rather than overapplication of a phonological process. It should be noted, however, that these predictions of the paradigm uniformity hypothesis pertain to levelling in the sense of systematic, regular patterns that can reasonably be ascribed to the synchronic grammar of the language in question. The question of how (or whether) these considerations apply to levelling as a diachronic process will be taken up at the end of this section.

The paradigm uniformity notion has potential applications that may at first glance seem counterintuitive. McCarthy (1998, 2005) uses it as a means of accounting for templatic or canonical shape restrictions on stems or roots. For example, McCarthy (2005) calls attention to the fact that Classical Arabic verb stems can only end in /...CVC/, whereas noun stems are less restricted and can also end in /...CV:C/ or /...CVCC/.
By appealing to paradigm uniformity constraints, he explains this as a mere by-product of the (accidental) fact that the inflectional suffixes of nouns in Classical Arabic all happen to be vowel-initial, whereas the inflectional morphology of verbs includes consonant-initial suffixes as well. Crucially, the phonology of Classical Arabic permits neither tri-consonantal clusters nor superheavy syllables, and stem-suffix sequences like */CVCC-CV/ or */CV:C-CV/ would thus require repair by vowel epenthesis and vowel shortening, respectively. A high-ranked demand for paradigm uniformity would have the effect of extending (overapplying) this repair throughout the inflectional paradigm. As a result, even if an underlying verb stem like */fa'ala:/ were to exist, its surface manifestation would be indistinguishable from that of its */CV:C/CV:CV/ counterpart */fa'ala:/, as illustrated in (12). The input */fa'ala:/ would thus not be recoverable from the point of view of the language learner (an instance of what Prince and Smolensky 2004 dub 'Stampean occultation'). The complete absence of consonant-initial suffixes in the nominal paradigm, by contrast, means that the threat of vowel-length alternations does not materialize there, and a stem of the underlying shape */CV:CCV:CV/ is thus free to surface intact, and is clearly distinct from a */CV:CVC/CV:CV/ stem, as seen in (13).

(12) Inflectional paradigm of hypothetical verb stem */fa'ala:/

<table>
<thead>
<tr>
<th>/fa'ala:/ + {a, tu, ...}</th>
<th>IDENT-μ-OP</th>
<th>*σμμμ</th>
<th>IDENT-μ-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ⟨fa'ala:a, fa'ala:ltu, ...⟩</td>
<td><img src="image1" alt="image" /></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ⟨fa'ala:a, fa'ala:ltu, ...⟩</td>
<td><img src="image2" alt="image" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ⟨fa'ala:a, fa'ala:ltu, ...⟩</td>
<td><img src="image3" alt="image" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(13) Inflectional paradigm of noun stem /kitab:/ 'book'

<table>
<thead>
<tr>
<th>/kitab:/ + {u, i, ...}</th>
<th>IDENT-μ-OP</th>
<th>*σμμμ</th>
<th>IDENT-μ-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ⟨kitabu, kitabi, ...⟩</td>
<td><img src="image4" alt="image" /></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>b. ⟨kitabu, kitabi, ...⟩</td>
<td><img src="image5" alt="image" /></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ⟨kitabu, kitabi, ...⟩</td>
<td><img src="image6" alt="image" /></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

However, as Bobaljik (2008) points out, the absolute neutralization of */CV:CVC/ and (hypothetical) */CV:CVC:CV/ verb stems at the level of the inflectional paradigm, seen in (12), is insufficient as an explanation for the categorical absence of */CV:CVC/ verb stems. This is because the underlying vowel length contrast should still manifest itself overtly in other, non-inflectional morphology, such as nominalizations. Such evidence would provide the necessary clues for learners in order to internalize the appropriate */CV:CVC/ or */CV:CVC:CV/ underlying representation for each verb stem. To cite a parallel from English, derivational formations like [dæm.n-er.ʃ] provide evidence for the verb damn having an underlying representation /dæm:n/, with final /n/, despite the complete neutralization toward /n/-less [dæm] throughout the inflectional paradigm ([dæm], [dæm-d], [dæm-ɪ], etc.)—which is similarly due to levelling by overapplication, in this case of coda cluster simplification.
What unites all accounts of levelling phenomena that attribute them either to output-output correspondence (base identity) or some notion of paradigm uniformity is one fundamental assumption: that there exists a universal desideratum or preference for identity of exponence, and thus for the absence of morphophonological alternations that disrupt such identity. As was noted earlier, proposals for paradigm uniformity constraints are intended to account for cases in which there appears to be evidence for levelling as a systematic, synchronic pattern that can thus be attributed to the grammar of the language in question (as a model of speakers' implicit knowledge). What the relationship is between this notion and that of levelling as a diachronic process (often sporadic) is a question that is largely left unaddressed in the phonological literature advocating these approaches. While the idea of uniform exponence certainly has deep roots in the historical linguistics tradition (see e.g. Hock 1991: 168)—and is indeed implicit in the traditional term 'paradigm levelling'—the notion of levelling as a distinct type of language change has increasingly come under challenge in that field in recent years. Evidence now seems to suggest, rather, that all putative cases of paradigm levelling as a diachronic process may instead be instances of analogical extension, in that levelling appears to crucially depend on the prior existence of a pattern of non-alternation that can serve as an analogical model (Albright 2005; Garrett 2008; see also Hill 2007). Albright (2005) argues against paradigm uniformity explanations of the historical levelling of Latin rhotacism alternations ([s] ∼ [r]), instead advocating an account that is based on a formally explicit model of morphological learning. In effect, this attributes paradigm levelling to imperfect learning, which has long been the standard account for analogical change in general (Kiparsky 1965, 1968).

8.3.2 Paradigm Anti-homophony Effects

As noted in the previous section, levelling and paradigm uniformity have the effect of reducing phonological differences among the set of word forms that make up an inflectional paradigm, by eliminating alternations within the stem portion that these forms share. Such anti-allomorphy effects can in principle involve either the suppression of the operation of a phonological process or an extension of its application into contexts which do not fit its conditioning environment. Conversely, departures from the normal application of a phonological process sometimes appear to serve the goal of maintaining or creating a difference in phonological form between morphosyntactically distinct forms within the paradigm. Such phenomena are usually referred to as anti-homophony effects, or intra-paradigmatic homophony avoidance (Crosswhite 1999; Blevins 2005; Gessner and Hansson 2005; Ichimura 2006; Blevins and Wedel 2009; Mondon 2009), though perhaps 'syncretism avoidance' would be a more apt term. Just as levelling has been attributed to a general principle (or constraint) of paradigm uniformity, so anti-homophony effects have been similarly explained by
invoking a principle of *paradigmatic contrast* and its maintenance (Kenstowicz 2005; Rebrus and Törkenczy 2005; see also Alderete et al. 1999; Kurisu 2001 for related notions).

As an example, consider Kenstowicz's (2005) discussion of various phonological irregularities that characterize 3sg.<p>perfective forms of verbs in many modern Arabic dialects, specifically when these are followed by an object suffix. In the Damascus dialect of Syrian Arabic, for example, such forms exhibit an aberrant stress pattern in precisely those cases where regular stress, and concomitant syncopation of unstressed schwa in open syllables, would have resulted in homophony with the corresponding 1sg and 2sg.<p>forms. (The latter two are themselves homophonous/syncretic, a point to which we will return later in this section.) This is illustrated in (14). The aberrant 3sg.<p>forms, and the 1sg/2sg.<p>forms with which they ought to be homophonous, were the phonology of stress and syncopation to operate as expected, are shown in boldface. The boundary between the stem and subject-agreement suffix is indicated by a hyphen, that between the subject and object suffixes by #. The regular stress pattern of Syrian Arabic is that stress is limited to a three-syllable window at the right edge of the word; within this window, stress falls on the rightmost heavy syllable (a word-final consonant counts as extrametrical and does not render the final syllable heavy), otherwise on the antepenultimate syllable.

(14) Damascus Arabic: anti-homophony effects in perfective paradigms

(Kenstowicz 2005)

<table>
<thead>
<tr>
<th>Object</th>
<th>‘he taught’</th>
<th>‘she taught’</th>
<th>‘I taught / you-sg.&lt;p&gt;taught’</th>
</tr>
</thead>
<tbody>
<tr>
<td>(none)</td>
<td>‘al.lam</td>
<td>‘al.lam-m-ɔt</td>
<td>‘al.’lam-t</td>
</tr>
<tr>
<td>1SG</td>
<td>‘al.’lam-#ni</td>
<td>‘al.lam-m-ɔt,#ni</td>
<td>‘al.’lam-t,#ni</td>
</tr>
<tr>
<td>2SG.M</td>
<td>‘al.lam-m-ɔk</td>
<td>‘al.lam-m-ɔt-#ak</td>
<td>‘al.’lam-#ak</td>
</tr>
<tr>
<td>2SG.F</td>
<td>‘al.lam-m-ɔk</td>
<td>‘al.lam-m-ɔt-#ok</td>
<td>‘al.’lam-#ok</td>
</tr>
<tr>
<td>3SG.M</td>
<td>‘al.lam-#o</td>
<td>‘al.lam-m-ɔt-#o</td>
<td>‘al.’lam-#o</td>
</tr>
<tr>
<td>3SG.F</td>
<td>‘al.lam-#(h)a</td>
<td>‘al.lam-m-ɔt-#(h)a</td>
<td>‘al.’lam-#(h)a</td>
</tr>
<tr>
<td>1PL</td>
<td>‘al.’lam-#na</td>
<td>‘al.lam-m-ɔt-#na</td>
<td>‘al.’lam-#na</td>
</tr>
<tr>
<td>2PL</td>
<td>‘al.’lam-#kom</td>
<td>‘al.lam-m-ɔt-#kom</td>
<td>‘al.’lam-#kom</td>
</tr>
<tr>
<td>3PL</td>
<td>‘al.lam-#(h)om</td>
<td>‘al.lam-m-ɔt-#(h)om</td>
<td>‘al.’lam-#(h)om</td>
</tr>
</tbody>
</table>

Kenstowicz (2005) accounts for the unexpected realization of forms like /‘al.lam-m-ɔt-#o/ → [‘al.lam-m-ɔt-#o] ‘she taught him’ (as opposed to the expected *[‘al.’lam-#o], which would have been homophonous with ‘I taught him / you-sg.<p>taught him’) by invoking a high-ranked constraint *PARADIGM CONTRAST*, as illustrated in (15). Here the label *SYNCOPE* is a stand-in for the constraint or constraints that effect syncope of schwa (for our purposes, a ban against [a] in unstressed open syllables), and *NONFINALITY* requires that no foot be word-final, thus forcing antepenultimate stress in words ending in a sequence of three light syllables.
(15) Anti-homophony enforced: ['1alla'məto] 'she taught him' ≠ ['fal'lamto] 'I taught him'

<table>
<thead>
<tr>
<th>/'1allam-ət#o/</th>
<th>PARADIGM CONTRAST</th>
<th>SYNCOPE</th>
<th>NONFINALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. fal.'la.mə).to</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. fal.'lam).to</td>
<td></td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>c. /fal.la.'mə.to</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The strongest evidence in support of a homophony avoidance account of these facts is the observation that the same aberrant stress pattern is not seen in verbs which happen to have a different stem shape in their 3SG.F perfective forms than in the corresponding 1SG/2SG.M form. Thus, for example, so-called 'hollow' verbs like /fa:f/ 'see' display a vowel quality and length alternation within their subject-inflection paradigm: ['fa:f-ət] 'she saw' but ['uf-t] 'I saw/you-sg.m saw'. For this reason, there is no risk of homophony between object-inflected forms like 'she saw him' and 'I saw him/you-sg.m saw him', and the former is free to surface with the regular stress pattern (and concomitant syncope), as shown in (16).

(16) Anti-homophony irrelevant: ['fa:f-ət#o] 'she saw him' ≠ ['uf-ət#o] 'I saw him / you-sg.m saw him'

<table>
<thead>
<tr>
<th>/'fa:f-ət#o/</th>
<th>PARADIGM CONTRAST</th>
<th>SYNCOPE</th>
<th>NONFINALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. ('fa:f.fo).to</td>
<td></td>
<td>*!</td>
<td></td>
</tr>
<tr>
<td>b. /fo: ('fa:f).to</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. /fa: ('fo.to)</td>
<td></td>
<td></td>
<td>*</td>
</tr>
</tbody>
</table>

Kenstowicz (2005) adduces a number of similar examples from other modern Arabic vernaculars, where other deviations from the regular phonology are employed to maintain the same distinctions (3SG.F vs. 1SG/2SG.M before a vowel-initial object suffix).

As the glosses above suggest, the 1SG and 2SG.M have become homophonous throughout the perfective paradigms in Syrian Arabic; both are formed with /-t/ (< Classical Arabic /-tu/ and /-ta/, respectively). This would appear to be a counter-example to the enforcement of PARADIGM CONTRAST, in particular in light of the fact that the 1SG vs. 2SG.M distinction is maintained elsewhere in the language, such as in the imperfective (e.g. ['bok.tob] 'I am writing', ['btak.tob] 'you-sg.m are writing'). The 1SG vs. 2SG.M situation is not entirely parallel, however, in that these two paradigm cells are homophonous/syncretic already in their underlying representation. What is observed in (14), and other similar cases discussed by Kenstowicz (2005), is an apparent avoidance of derived instances of homophony/syncretism: ones that would otherwise have been expected to result from the regular application of phonological processes in the language. That being said, there is no obvious reason why high-ranked PARADIGM CONTRAST could not equally well act as a trigger of (ad hoc) phonological
changes whose sole purpose would be to create a surface contrast between underly-
ingly homophonous paradigm cells. For example, one could easily imagine the 1sg vs. 2sg.m distinction being upheld on the surface by again resorting to stress shift, with /sallam-t/ → ['s'lam] (regular stress) in one case and /sallam-t/ → ['s'lam] (aberrant stress) in the other.

Whether this is a serious problem or not depends on exactly how Paradigm Contrast is defined formally. Kenstowicz (2005) does not provide an explicit definition, but refers to the one formulated by Rebrus and Törkenczy (2005), which is shown in (17).6

\[(17) \text{CON} [= \text{Paradigm Contrast}] \quad \text{(Rebrus and Törkenczy 2005: 264)}\]

The surface realizations of morphologically distinct members \(m_1 \ldots m_n\) of a paradigm \(x\) must be phonetically distinct.

The key question here is what qualifies as being ‘morphologically distinct’. If the 1sg and 2sg.m differ at the level of morphosyntactic feature specifications, but involve exponents that are accidentally homophonous (both /-t/), that should presumably still count. Alternatively, one might imagine treating /-t/ as being underspecified for the first vs. second person opposition ([±addersee]), as well as for gender (whereas 2sg.f /-t/ would be contrastively specified as [+addersee, +feminine], for example). In that case it could be argued that there is no 1sg vs. 2sg.m ‘morphological distinction’ for Paradigm Contrast to uphold in the Syrian Arabic perfective—rather, the two constitute a single syncretic paradigm cell, represented by a single output form.

Such problems of formal implementation aside, the very validity of intra-
paradigmatic homophony avoidance as a synchronic guiding principle in morphophonology is a matter of some doubt. As in the case of paradigm uniformity (see Section 8.3.1 above), a number of researchers have expressed scepticism in this regard; for a recent critique, see Mondon (2009). Several recent works attempt instead to define a principled mechanism whereby anti-homophony effects can arise through blocking (or apparent exceptions to) otherwise regular historical sound changes (Blevins 2005; Blevins and Wedel 2009; Gessner and Hansson 2005; Mondon 2009). However, this does still leave open the possibility that there is some role to play for homophony avoidance in online morphological processing and production, which could then provide a potential source for patterns that can in turn become grammaticalized. For the potential role of intra-paradigmatic homophony avoidance as a contributing factor in the emergence of paradigm gaps, see Baerman (2011, this volume) and Section 8.3.3.

6 Rebrus and Törkenczy (2005: 266) further propose that constraints on paradigm contrast and uniformity be relativized to particular feature dimensions (‘\[\text{CON}(D)\]: A form realizing some value of a morphosyntactic dimension \(D\) of paradigm \(x\) must be phonetically distinct from forms realizing other values of \(D\)’). This seems irrelevant in the Syrian Arabic case, as the same dimensions (person, gender) are involved in both the avoided 3sg.f vs. 1sg/2sg.m identity and the tolerated 1sg vs. 2sg.m identity.
8.3.3 Phonotactics, Alternations, and Paradigm Gaps

A final area where phonology intersects with paradigm structure is in the phenomenon of paradigm gaps (Hetzron 1975; Iverson 1981; Hansson 1999; Albright 2003; Raffelsiefen 2004; Pertsova 2005; Sims 2006, 2009; Rice and Blaho 2009; Baerman et al. 2010; Baerman 2011, this volume). In many cases of defective paradigms, the occurrence and distribution of gaps within the paradigm is strongly correlated with certain aspects of the phonological shape of the expected (but non-existent) word forms in question. Phonologically motivated paradigm gaps can be said to come in two flavours: ones that seem primarily driven by phonotactic well-formedness considerations, and ones that are lexically arbitrary but appear motivated by morpho-phonological alternation patterns. (However, this dichotomy may be illusory; see Albright 2009: 158–60 for discussion.)

A striking example of what appears to be a phonotactically motivated paradigm gap occurs for imperatives in Norwegian, as described by Rice (2007). Imperatives of verbs consist of the bare root, as shown in (18a). However, for verb stems that end in a rising-sonority consonant cluster—one which cannot be parsed as a complex coda—no imperative form exists (at least for certain dialects, Rice 2003), as indicated in (18b). This is a slight oversimplification, however, in that it is only in utterance-final or pre-consonantal contexts that the indicated forms are unacceptable. Should the imperative be embedded in a clausal context in which the next word happens to be vowel-initial, the imperative will surface intact, its final cluster resyllabified as a coda-onset sequence; this is illustrated in (19)–(20) with examples from Rice (2007).

(18) Norwegian: imperative gaps due to syllable structure

<table>
<thead>
<tr>
<th>INFINITIVE</th>
<th>IMPERATIVE</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. spiːs-ə 'eat'</td>
<td>spiːs 'eat!'</td>
</tr>
<tr>
<td>snakː-ə 'talk'</td>
<td>snakː 'talk!'</td>
</tr>
<tr>
<td>løft-ə 'lift'</td>
<td>løft 'lift'</td>
</tr>
<tr>
<td>b. klatr-ə 'climb'</td>
<td>Ø (*klatr) 'climb!'</td>
</tr>
<tr>
<td>sykl-ə 'bike'</td>
<td>Ø (*sykl) 'bike!'</td>
</tr>
<tr>
<td>padl-ə 'paddle'</td>
<td>Ø (*padl) 'paddle!'</td>
</tr>
<tr>
<td>oːpn-ə 'open'</td>
<td>Ø (*oːpn) 'open!'</td>
</tr>
</tbody>
</table>

(19) a. *Sykl ned bakken!
    bike.IMP down hill.DEF
    'Bike down the hill!'

b. Sykl opp bakken!
    bike.IMP up hill.DEF
    'Bike up the hill!'
Their dependence on the phrasal phonological context clearly suggests that the paradigm gaps in Norwegian imperatives are a relatively ‘late’ phenomenon, derivationally speaking. Another notable fact about these gaps is that the very same kinds of sonority sequencing violations are subject to repair, by schwa epenthesis, in the nominal morphology. Thus, for example, the unsuffixed singular form of the noun /sykI/ ‘bicycle’ (pl [sykl-ø]) is realized as disyllabic [sykIC], with epenthesis to avoid the word-final /kl/ cluster.7 For this reason, uncertainty about whether or not to apply epenthesis in the imperatives in (18b) might be a contributing factor as well.

The second type of paradigm gap in which phonological factors also seem to be implicated targets inflected forms which, had they existed, might have been expected to exhibit some phonological alternation (typically in the stem) relative to other forms in the paradigm (Hetzron 1975; Albright 2003, 2009). The well-known cases of gaps in verb paradigms of Spanish and Russian are a prime example of this. In Spanish, gaps target exactly those forms in the present indicative paradigm in which raising or diphthongization of the stressed stem vowel, or appearance of a velar stop (typically after an otherwise stem-final [θ]/[s], [l] or [n]), is observed in a large number of verbs of the corresponding inflectional classes (21). In Russian, gaps in the nonpast paradigm similarly target exactly the form in which palatalization (postalveolarization, rather) of the stem-final coronal obstruent would be expected to occur (22).

(21) Spanish: gaps in present indicative paradigms

<table>
<thead>
<tr>
<th></th>
<th>/abol-ir/</th>
<th>/sol-er/</th>
<th>/as-ir/</th>
<th>/kreθ-er/</th>
</tr>
</thead>
<tbody>
<tr>
<td>'abolish'</td>
<td>'be used to'</td>
<td>'grasp'</td>
<td>'grow'</td>
<td></td>
</tr>
<tr>
<td>1SG</td>
<td>—</td>
<td>'swel-o'</td>
<td>—</td>
<td>'kreθk-o'</td>
</tr>
<tr>
<td>2SG</td>
<td>—</td>
<td>'swel-es'</td>
<td>'as-es'</td>
<td>'kreθ-es'</td>
</tr>
<tr>
<td>3SG</td>
<td>—</td>
<td>'swel-e'</td>
<td>'as-e'</td>
<td>'kreθ-e'</td>
</tr>
<tr>
<td>1PL</td>
<td>aβo'l-imos</td>
<td>so'l-emos</td>
<td>a's-imos</td>
<td>kreθ'emo's</td>
</tr>
<tr>
<td>2PL</td>
<td>aβo'l-is</td>
<td>so'l-eis</td>
<td>a's-is</td>
<td>kreθ'emo's</td>
</tr>
<tr>
<td>3PL</td>
<td>—</td>
<td>'swel-en'</td>
<td>'as-en'</td>
<td>'kreθ-en'</td>
</tr>
</tbody>
</table>

(22) Russian: gaps in nonpast paradigms (Baerman 2008; Albright 2009)

<table>
<thead>
<tr>
<th></th>
<th>/u-bed-it/</th>
<th>/sled-it/</th>
<th>/o-bez-les-it/</th>
<th>/gas-it/</th>
</tr>
</thead>
<tbody>
<tr>
<td>1SG</td>
<td>—</td>
<td>sIe's-u</td>
<td>—</td>
<td>ga'j-u</td>
</tr>
<tr>
<td>2SG</td>
<td>ube'dl-if</td>
<td>sle'dl-if</td>
<td>obez'lesl-if</td>
<td>'gasl-if</td>
</tr>
<tr>
<td>3SG</td>
<td>ube'dl-it</td>
<td>sle'dl-it</td>
<td>obez'lesl-it</td>
<td>'gasl-it</td>
</tr>
<tr>
<td>1PL</td>
<td>ube'dl-im</td>
<td>sle'dl-im</td>
<td>obez'lesl-im</td>
<td>'gasl-im</td>
</tr>
<tr>
<td>2PL</td>
<td>ube'dl-it'e</td>
<td>sle'dl-it'e</td>
<td>obez'lesl-it'e</td>
<td>'gasl-it'e</td>
</tr>
<tr>
<td>3PL</td>
<td>ube'dl-at</td>
<td>sle'dl-at</td>
<td>obez'lesl-at</td>
<td>'gasl-at</td>
</tr>
</tbody>
</table>

Several of the examples of defective paradigms adduced by Baerman (this volume) fit this description, such as where the Finnish verb stem /erka(n)/- 'spread out' ([erkan-e:] 's/he spreads out') is missing all forms that would normally have called for a weak-grade stem alternant [era(n)-], including the infinitive *[era-ta] (cf. [vaiken-e:] 's/he is quiet', [vaie-ta] 'to be quiet'). Similarly, as Baerman points out, the GEN.PL gap in the paradigm of Russian /meft-a/ 'dream' (expected GEN.PL form: *['meft]) is at least partly due to the fact that this would be the only form in the paradigm to exhibit stem stress, as all other forms carry suffix stress: NOM.SG [meft-a], GEN.SG and NOM.PL [meft-t-i], and so forth (cf. the phonotactically equivalent and non-gapped /maft-a/ 'mast', for which the GEN.PL ['maft] receives support from stem-stressed forms like NOM.PL ['maft-i]).

The most convincing explanation for paradigm gaps of this type is that of Albright (2003, 2009), in which uncertainty on the part of the speaker is the driving factor: 'gaps occur when speakers know that an inflected form must stand in a certain relation to another inflected form, but the language does not provide enough data to be certain of what that relation should be' (Albright 2009: 118). For example, the Spanish stem-vowel alternations seen in (21) are not observed in all verbs belonging to the relevant inflectional classes (compare [so'I-er] 'be used to', 1SG ['swel-o], to [ko'm-er] 'eat', 1SG ['kom-o]). The lexical distribution of defective verbs, and speakers' confidence in their own production of putatively gapped inflected forms, appear to be intimately correlated with the statistics of the lexicon—namely, how prevalent alternation happens to be for particular inflectional (sub)classes, or for verb stems with particular segmental properties.

Since they typically involve alternations in the phonological shape of the stem, gaps of this kind have the effect of producing paradigms that are 'uniform' (across their non-gapped forms) in the sense of Section 8.3.1. Avoidance of paradigm-internal homophony or syncretism (Section 8.3.2) also appears to play a contributing role in some cases (Baerman 2011). For example, in the large and productive inflectional class of weak feminine nouns in Icelandic (23), many nouns have a gap in the GEN.PL. The historically inherited GEN.PL ending for nouns of this class is /-na/.\(^8\) However, some

\(^8\) Aside from weak feminine nouns, the only other inflectional class that retains /-na/ are the weak neuters, a small set of half a dozen nouns mostly denoting body parts (e.g. /lœy-al/ 'eye', /fær-al/ 'kidney').
weak feminines have adopted the */-a/ GEN.PL ending that is used in virtually all other inflectional classes, whereas others vacillate between the two or have an outright gap (with considerable inter- and intra-speaker variation as to whether a given noun can be used in the GEN.PL, and if so, with which ending).

(23) Icelandic: gaps in weak feminine noun paradigms

<table>
<thead>
<tr>
<th></th>
<th>/kʰul-a/</th>
<th>/cʰist-a/</th>
<th>/pʰer-a/</th>
<th>/hol-a/</th>
<th>/fiöl-a/</th>
</tr>
</thead>
<tbody>
<tr>
<td>SG NOM</td>
<td>kʰu:l-a</td>
<td>cʰist-a</td>
<td>pʰe:r-a</td>
<td>hol-a</td>
<td>fiöl-a</td>
</tr>
<tr>
<td>ACC</td>
<td>kʰu:l-y</td>
<td>cʰist-y</td>
<td>pʰe:r-y</td>
<td>hol-y</td>
<td>fiöl-y</td>
</tr>
<tr>
<td>DAT</td>
<td>kʰu:l-y</td>
<td>cʰist-y</td>
<td>infa:</td>
<td>hol-y</td>
<td>fiöl-y</td>
</tr>
<tr>
<td>GEN</td>
<td>kʰu:l-y</td>
<td>cʰist-y</td>
<td>pʰe:r-y</td>
<td>hol-y</td>
<td>fiöl-y</td>
</tr>
<tr>
<td>PL NOM</td>
<td>kʰu:l-yr</td>
<td>cʰist-yr</td>
<td>pʰe:r-yr</td>
<td>hol-yr</td>
<td>fiöl-yr</td>
</tr>
<tr>
<td>ACC</td>
<td>kʰu:l-yr</td>
<td>cʰist-yr</td>
<td>pʰe:r-yr</td>
<td>hol-yr</td>
<td>fiöl-yr</td>
</tr>
<tr>
<td>DAT</td>
<td>kʰu:l-yr</td>
<td>cʰist-yr</td>
<td>pʰe:r-yr</td>
<td>hol-yr</td>
<td>fiöl-yr</td>
</tr>
<tr>
<td>GEN</td>
<td>kʰu:l-na</td>
<td>cʰist-na</td>
<td>pʰe:r-na</td>
<td>hol-na</td>
<td>fiöl-na</td>
</tr>
</tbody>
</table>

As can be seen from the paradigms in (23), */-na/ is the only inflectional suffix that is consonant-initial. As such, it can trigger a number of regular phonological processes that have the effect of creating an alternation in the surface shape of the stem in the GEN.PL form vis-à-vis all others in the paradigm. These include vowel shortening ([sul-a] 'pillar', GEN.PL [sul-na]), hardening ([kʰra:v-a] 'demand', GEN.PL [kʰrap-na]), depalatalization ([cʰc-a] 'widow', GEN.PL [cʰk-na]), preaspiration ([vʰkʰ-a] 'week', GEN.PL [vʰkh-na]), cluster simplification ([tʰun-k-a] 'tongue', GEN.PL [tʰun-na]), and stop excrescence ([vʰis-a] 'quatrain', GEN.PL [vʰst-na]). Avoidance of these sorts of alternations is no doubt a primary motivation for avoiding */-na/ in favour of */-a/. For example, stems ending in a singleton palatal stop, for which */-na/ would trigger a combination of depalatalization, vowel shortening, and preaspiration, tend very strongly toward either */-a/ or a gap (e.g. *[θc:c-h-a] 'roof, cover', GEN.PL *[θc:h-θa], *[θc:c-h-a]). By contrast, stems ending in a geminate and/or preaspirated palatal, or a cluster involving a palatal, are typically much more tolerant of */-a/, since here depalatalization is the only consequence ([cʰk-a] 'church', GEN.PL [cʰk-k-na]; cf. also [cʰc-a] 'widow', just mentioned). In other cases the avoidance of */-na/ appears to be phonotactically motivated to some extent, in that it would lead to a consonant cluster that is otherwise unattested (and for which a choice of phonological repair is not immediately obvious), such as *[tʰydr-a] 'pouch', GEN.PL *[tʰydr-na].

9 Baerman (2011) uses *hola 'hole' as his illustrative example of a defective weak feminine noun. A Google search produces a considerable number of hits for each of *hol-na and *hol-a in GEN.PL contexts, suggesting that for at least some speakers this word does not have a gapped paradigm. For the other gapped noun shown in (23), *fiöl-a 'violin', both GEN.PL *fiöl-na and *fiöl-a are markedly rare. (In the definite form, GEN.PL *fiöl-a-nna is robustly attested, lending further support to a homophony avoidance account; the addition of the suffixed definite article disambiguates the GEN.PL from NOM.SG *fiöl-a-n.)
Whereas the dispreference for the /-na/ allomorph is thus arguably phonologically motivated, at least in part, what underlies the dispreference for /-a/ seems to be the fact that it renders the GEN.PL homophonous with the (much more frequent and salient) NOM.SG form (Baerman 2011). (Indeed, as Baerman points out, it is hardly an accident that the two inflectional classes of nouns that have retained GEN.PL /-na/ in the first place are precisely the ones in which the NOM.SG happens to be marked with /-a/.) The conflict resulting from these two opposing desiderata, then, is what drives speakers’ uncertainty about the GEN.PL form of many weak feminines, a paradigm gap being the most extreme manifestation of such uncertainty. Frequency is presumably a third contributing factor exacerbating the uncertainty: the relatively low token frequency of GEN.PL forms in general, and the low type frequency of /-na/ as an exponent of GEN.PL across the lexicon as a whole.

8.4 Phonologically Conditioned Allomorphy

The most obvious way in which the phonology of a language can make its presence felt in the inflectional system is by giving rise to alternations in the sound shape of some morpheme, stem, or lexeme. For example, Modern Icelandic lengthens vowels in stressed open syllables, and hardens certain voiced fricatives to (voiceless unaspirated) stops before [], and this results in a length and manner alternation in the stem of [sary-a] ‘story-NOM.SG’ vs. [sak-na] ‘story-GEN.PL’. Similarly, the regular English plural marker surfaces as [-z], [-s], or [-iz] depending on whether the preceding segment is voiced or voiceless, sibilant or non-sibilant. Whenever such an alternation can be related directly to some generalization about phonotactics, allophonic distribution, neutralization patterns, or the like—where that generalization is valid either across the board or in some well-defined set of morphological contexts—the standard assumption in generative phonology (and to some extent pre-generative frameworks as well) is that the alternation is to be factored out by positing a single, underlying (input, phonemic, lexical) representation of the morpheme in question. Thus the regular English plural is simply /-z/ as far as the morphological system is concerned, regardless of the fact that this /-z/ happens sometimes to surface as [-s] or [-iz] due to the exigencies of English segmental phonology.

Instances of alternation do however exist which, although phonologically conditioned, do not seem amenable to this kind of analysis. There can be several possible reasons (not mutually exclusive) for why this may be the case. First, the alternants may be too divergent from each other to be derivable from a single phonological representation. Second, it may be impossible to derive the differences in sound shape from the different conditioning environments in which the alternants occur. Third, the alternation may be attested too sporadically to warrant treatment as the product
of a (perhaps construction-specific) phonological process. Alternations that fail on one or more of these criteria make up the important class of morphology–phonology interface phenomena that are known collectively as phonologically conditioned allomorphy (Carstairs 1988; Nevins 2011). Since the term ‘allomorphy’ is occasionally used in a wider sense, encompassing purely phonological alternations such as English [-z]~[-s]~[-iz] as well, the less ambiguous term phonologically conditioned suppletive allomorphy is preferred by some.

A textbook example of phonologically conditioned allomorphy is the English indefinite article, which is [a] before a consonant but [an] before vowels. No other English morpheme displays this kind of [n]~Ø alternation, and there is no independently attested process of either nasal deletion or nasal epenthesis that the alternation could be attributed to. Hence we are forced to admit that the indefinite article has two phonological representations, two allomorphs, /a/ and /an/, whose relative distribution is governed by phonological factors. A more subtle case is the voicing alternation exhibited by a handful of suffixes in Basque, discussed by Mascaro (2007). Modern Basque does not have a productive process of postnasal voicing ([mendi] ‘mountain’, [kontu] ‘count’, [hon-ta$] ‘this-INSTR’). However, a small number of inflectional and derivational suffixes display an alternation [t]~[d] or [k]~[g], with the voiced alternant systematically appearing after nasal-final bases (and to some extent after /l/-final ones as well, subject to dialect variation). This is illustrated in (24) with the example of ablative [-tik]~[-dik].

(24) Basque: postnasal voicing allomorphy

\[
\begin{align*}
\text{bilbo} & \quad \text{‘Bilbao’} & \quad \text{bilbo-tik} & \quad \text{‘from Bilbao’} \\
\text{irun} & \quad \text{‘Irun’ (place name)} & \quad \text{irun-dik} & \quad \text{‘from Irun’} \\
\text{non} & \quad \text{‘where’} & \quad \text{non-dik} & \quad \text{‘from where’}
\end{align*}
\]

As a phonological process, post-nasal voicing is both phonetically natural and typologically frequent, and is commonly analysed as resulting from a high-ranked, and phonetically grounded, markedness constraint *N_Q (Hayes 1999; Pater 2001). If those affixes of Basque that show the voicing alternation in (24) formed a coherent class on independent grounds (morphological, syntactic, or semantic), then that would invite the possibility of interpreting this as a case of construction-specific phonology. For example, the affixes in question could be seen as belonging to a cophonology or stratum in which *N_Q is ranked higher than otherwise; alternatively, they could be indexed to a specific and low-ranked version of the faithfulness constraint that demands preservation of underlying [±voice] values. Mascaro (2007) argues that this is not tenable in the Basque case. The suffixes displaying the voicing alternation constitute a heterogeneous set of about a half-dozen morphemes, some of which are clearly derivational (such as [-tar]~[-dar], e.g. [bilbo-tar] ‘Bilbaean’, [irun-dar] ‘Irunese’), some inflectional (such as the ablative shown in example (24) and the future participle suffix [-ko]~[-go]), and one is even a clause-final clitic ([=ta]~[=da], an enclitic version of the conjunction /eta/ ‘and; since’).
It would appear, then, that the observed voicing alternations in (24) cannot be construed as the product of a phonological process in the usual sense. Instead, the existence of two alternative surface shapes \([-tik]\) and \([-dik]\) must to some extent be relegated to the morphology and/or the lexicon. From the point of view of the phonology, then, the Basque ablative suffix has two allomorphs, \(/-tik/\) and \(/-dik/\), which despite their similarity are, for all intents and purposes, suppletive. From an item-based perspective, the (lexical, underlying) phonological representation of the ablative suffix morpheme consists of the set \(/-tik, \text{-}dik/\). From a realizational perspective one might posit two separate suffixation rules, involving the mappings \(X \rightarrow X_{tik}\) and \(X \rightarrow X_{dik}\), respectively. The difference is not important here, but for simplicity and consistency with most of the relevant phonological literature the discussion in the following subsections is couched in item-based (morpheme-based) terms.

### 8.4.1 Allomorph Selection as Optimization

Even though the Basque alternation just described is not itself due to a phonological process, the distribution of the allomorphs is nevertheless conditioned by phonological factors. In Optimality Theory, the standard approach to such phonologically conditioned allomorphy has been to model allomorph selection as output optimization (Mester 1994; Kager 1996; Mascaró 1996; Rubach and Booij 2001). Because both phonological shapes are already available in the input, the phonological grammar is free to choose that output shape which is less marked (phonologically) in the environment in question, without the cost of violating input-output correspondence. In this way, low-ranked markedness constraints that are otherwise not enforced in the language are able to make their presence felt: an ‘emergence of the unmarked’ effect (McCarthy and Prince 1994).

The Basque state of affairs can be seen as reflecting a ranking \(\text{IDENT}[\pm\text{voice}]-\text{IO} \gg \text{*NC} \gg \text{*VOICED-OBR}\), where the faithfulness constraint \(\text{IDENT}[\pm\text{voice}]-\text{IO}\) penalizes any deviations from input voicing specifications, and \(\text{*VOICED-OBR}\) is a context-free markedness constraint encoding the generally marked status of voiced vis-à-vis voiceless obstruents. In cases without allomorphy, such as root-internal /nt/ in /kontu/ ‘count’ or the non-alternating instrumental /-taš/ in /hon-taš/ ‘this-INST’, the undominated constraint \(\text{IDENT}[\pm\text{voice}]-\text{IO}\) will guarantee that the stop remains voiceless into the output, despite violating lower-ranked \(\text{*NC}\). When the input representation of a morpheme contains more than one allomorph, however, as in /irun-{tik, dik}/ ‘Irun-ABL’, it is possible to satisfy \(\text{*NC}\) for free, as it were, without violating \(\text{IDENT}[\pm\text{voice}]-\text{IO}\). This is illustrated by the tableaux in (25)–(27). Following Mascaró (2007), subscript indices signal which of the input allomorphs is being referenced (for purposes of faithfulness evaluation) in the output candidate in question. Thus \([\text{irun-dik}_1]\) is an unfaithful rendering of \(/\text{irun-tik}_1/\) (with post-nasal voicing enforced, analogous to the losing candidate in (25b) \([\text{hon-dag}]\)), whereas the homophonous \([\text{irun-dik}_2]\) is a fully faithful rendering of \(/\text{irun-dik}_2/\).
In a case like the Basque one, there is an obvious and natural connection between the conditioning environment (post-nasal vs. elsewhere) and the substantive difference between the two alternants (voiced vs. voiceless stop morpheme-initially). This invites an alternative solution. Rather than represent the suffix as a set, or disjunction, of two phonological shapes /{-tik, -dik}/, we could instead posit a single phonological representation that unifies the two by factoring out their featural differences: /-Tiki/, where /TI/ stands for a coronal stop that is unspecified for [±voice]. Inkelas (1995) proposes exactly this sort of archiphonemic underspecification analysis of the lexical contrast between alternating and non-alternating stem-final obstruents in Turkish. As long as the voiced–voiceless opposition is seen as equipollent rather than privative, and as long as (some) faithfulness constraints penalize feature-value changes over and above feature-value insertion, this will allow for an analysis along similar lines as above, with Faith-IO \(\gg\) *NC \(\gg\) *VoiObs. Exactly as with the set/disjunction approach, the underspecification approach encodes the difference between alternating (allomorphic) and non-alternating (non-allomorphic) morphemes in their lexical representation, but the two deal with the alternating ones in opposite ways. The former encodes both feature values at once in the input representation ([−voice] in /-tik/, [+voice] in /-dik/), while the latter encodes neither ([0voice] in /-Tiki/).

While an archiphonemic underspecification approach could arguably handle the voicing allomorphy in Basque, it runs aground on any of the numerous cases of phonologically conditioned allomorphy where the alternating morpheme shapes cannot be
construed as anything but suppletive. A few illustrative examples are shown in (28). The first three cases involve a post-consonantal vs. post-vocalic alternation, which is often best understood in terms of syllable structure. Thus, for example, the Korean forms in (28b) both have CV.CV shape, whereas the exact opposite distribution of the two allomorphs would have resulted in the more marked structures CVC.CV and CV.V. Note that in the Udihe case in (28d), the laryngealization of the stem-final vowel (e.g. [zawe] from /zawa/ ‘take, grab’) can be interpreted as a floating [+constricted glottis] featural suffix.

(28) a. Moroccan Arabic 3RD SINGULAR MASCULINE (Harrell 1962; Mascaró 1996)

[-h] / V ___ [xt’a-h] ‘his error’

b. Korean NOMINATIVE (Carstairs 1988; Embick 2010)

[-i] / C ___ [pa.p-i] ‘cooked.rice-NOM’
[-ka] / V ___ [kʰ.o.-ka] ‘nose-NOM’

c. Djabugay GENITIVE SINGULAR (Patz 1991; Kager 1996; Rubach and Booij 2001)

[-ŋ-n] / C ___ [gaŋal.-ŋ-n] ‘goanna-GEN’
[-n] / V ___ [gu.lu.du-n] ‘dove-GEN’

d. Udihe PERFECTIVE (Nikolaeva and Tolskaya 2001; Bye 2007)

[-ge] / V[+high] ___ [dogdi-ge] ‘hear-PFV’

In some cases, synchronically suppletive allomorphy has its historical origins in non-suppletive phonological alternations. For example, both Moroccan Arabic [-h] and [-u] in (28a) are reflexes of Classical Arabic /-hu/ by historical processes of apocope and lenition/deletion.

8.4.2 The Problem of Default Allomorphs

In the Basque case in examples (24)–(27), the distribution of each of the two allomorphs can be attributed to well-known phonological factors. The context-free markedness constraint *VOICE#OBSTRUENT favours [−tik] over [−dik] and so has the effect of defining the former as the default or elsewhere allomorph, whereas in the environment N___ the higher-ranked context-sensitive markedness constraint *NC becomes relevant and favours the latter, special-case allomorph. Similarly in the Moroccan Arabic case in (28a), uncontroversial syllable-structure constraints will penalize both the use of [-u] after vowel-final bases (creating hiatus, an ONSET violation, *[xt’ā-u]) and the use of [-h] after consonant-final bases (creating a
complex coda, a *Complex violation, *[ktab-h]). The complementarity of the two allomorphs is thus correctly captured.

However, when one allomorph appears to serve as an elsewhere case, it is often unclear how its default status is to be derived, and in many cases outright stipulation seems required. In the Djabugay genitive in (28c), it is understandable why [-\(\eta\un\)] would be favoured over [-n] after consonant-final bases ([ga,\(\eta\un\)-\(\eta\un\)] ‘goanna-GEN’ avoids the complex coda of its competitor *[ga,\(\eta\un\)-n]), but it is less obvious exactly what favours [-n] over [-\(\eta\un\)] otherwise (e.g. [gu.lu du-n] ‘dove-GEN’, not *[gu.lu du-\(\eta\un\)])]. Kager (1996) encodes that preference by stipulating an obviously parochial and language-specific constraint GENITIVE=/-n/, but Rubach and Booij (2001) attribute it to the relatively marked status of dorsal place over coronal place, and hence of the velar nasal [\(\eta\)] of the [-\(\eta\un\)] alternant. In the Udihe case in (28d), the avoidance of laryngealization on a stem-final high vowel has an obvious phonological motivation in the fact that laryngealized high vowels are categorically absent from the language, but it is far from clear why suffixation of [-ge] should be disfavoured after other vowels. If the perfective suffix has the allomorphic representation /{-ge, [+constricted glottis]}/, why is *[zawa-ge] worse than [zawa] as the perfective of /zawa/ ‘take, grab’?

To address this type of problem, Mascaro (2007) borrows from morphological theory the notion of morphological defaults (cf. Brown and Hippisley 2012; Brown, this volume), and incorporates it into the way listed allomorphs are represented and manipulated in the phonology. Mascaro proposes that (lexical, underlying) phonological representations involving allomorphy be construed as partially ordered sets, where the ordering of allomorphs encodes their relative priority. The ‘first’ allomorph is the default or elsewhere case, others are special cases limited to particular contexts. Selection of a lower-priority allomorph entails violating a special faithfulness constraint called PRIORITY, and can thus only be optimal if it serves as a means of satisfying some higher-ranked constraint. Couched in Mascaro’s framework, the alternation between laryngealization of a stem-final vowel and suffixation of [-ge] in Udihe perfectives might be captured by representing the perfective suffix as the ordered set /([-constricted glottis] > -ge]/. High-ranked PRIORITY ensures that, other things being equal, laryngealization is the preferred option (despite the unfaithfulness that this causes in the stem-final vowel). However, the even higher-ranked ban against laryngealized high vowels, *[+high, +constricted glottis], forces the emergence of the lower-priority allomorph [-ge] if and only if the stem-final vowel is high. This is shown in (29)-(30).

(29) /zawa-{-[+constricted glottis] > -ge]/ \(\rightarrow\) [zawa] ‘take-PFV’

<table>
<thead>
<tr>
<th>/zawa-{-[+c.g.] &gt; -ge2]/</th>
<th>*[+high, +c.g.]</th>
<th>PRIORITY</th>
<th>IDENT[±c.g.]</th>
<th>-IO</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. (\acia)zawa1</td>
<td></td>
<td>*</td>
<td></td>
<td>*</td>
</tr>
<tr>
<td>b. zawa-ge2</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
8.4.3 Non-optimizing Allomorphy

Extrinsic ordering among allomorphs addresses cases in which the selection of one of the allomorphs cannot easily be construed as phonologically optimizing. However, there exist numerous examples of phonologically conditioned allomorphy where neither of the allomorphs appears to involve optimization. One such case is Kaititj (Koch 1980; Paster 2005), where the ergative/instrumental/locative suffix is realized as [-l] after disyllabic stems but as [-l] after stems of more than two syllables, as illustrated in (31).10

(31) Kaititj: syllable-counting [-l] /-l/ allomorphy (Paster 2005)

\[
\begin{align*}
\text{a'ki-l} & \quad \text{head-ERG/INSTR/LOC} \\
\text{il'tii-l} & \quad \text{hand-ERG/INSTR/LOC} \\
\text{ajn' pi-l} & \quad \text{pouch-ERG/INSTR/LOC} \\
\end{align*}
\]

Other cases abound in which the connection between the shape of the allomorphs and the phonological factors controlling their distribution seems equally arbitrary, or even anti-optimizing. In Korean (Lapointe 2001), the conjunctive suffix is realized as [-wa] after vowels but [-kwa] after consonants, even though the latter results in the combination of a coda and complex onset (e.g. [mom.-kwa] 'body-CONJ') that could easily have been avoided by choosing the former allomorph instead (*[mom.-wa] or *[mo.m-wa]).

A particularly striking case is that of the postposed definite article in Haitian Creole (Nikiema 1999; Klein 2003; Paster 2006; Bonet et al. 2007; Bye 2007; Embick 2010), which is either [-la] or [-a] depending on the phonological properties of the preceding noun. If the noun ends in a consonant, [-la] is used (32a). If the noun ends in a vowel, [-a] is used (32b–c), and if the final vowel is [+Advanced Tongue Root] a homorganic glide is inserted to break the hiatus between the two vowels (32c).11

10 In Kaititj /l/ → [l] when the preceding consonant is apical. The allomorph /-l/ may thus surface as [-l] in some cases (e.g. [a'yi-r-l] 'kangaroo-ERG').

11 There appears to be some variation as to whether the glide insertion pattern is confined to [+ATR] vowels or holds after [e, o] as well (see Klein 2003 for discussion); the pattern depicted in (32) is the one described by Valdman (1978). Also, phonological processes of nasalization and assimilation produce other surface alternants not shown here, e.g. [ma-ga]-l] 'mango-DEF', [madam-na] 'lady-DEF', [tigas36[w]-a] 'boy-DEF' (Nikiema 1999).
The hiatus in (32b) could easily have been avoided by selecting the [-la] allomorph instead (*pa.pa.-Ia). As Bye (2007) points out, the choice of allomorph is particularly surprising in (32c), given that hiatus is in fact being actively avoided there but by means of a phonological operation. Using the other allomorph would have been a less costly solution than epenthesis in terms of faithfulness (e.g. *pa.pje.-Ia). Finally, in (32a) the [-a] allomorph would have resulted in a less marked syllable structure by avoiding the noun-final coda (*[li.v-a]), although at the expense of misalignment between syllable and morpheme boundaries (Klein 2003; Bonet et al. 2007; see Section 8.4.4).

8.4.4 Subcategorization vs. Optimization

The pervasive occurrence of (apparently) non-optimizing cases of phonologically conditioned allomorphy like the Haitian Creole one has prompted some researchers (Paster 2005, 2006, 2009; Bye 2007; Embick 2010) to reject altogether the notion of allomorph selection as output optimization. On the basis of an extensive typological survey of phonologically conditioned allomorphy, Paster (2006, 2009) argues that the observed typology conforms much better to the predictions of an analysis in terms of subcategorization than to those of the optimization approach. In a subcategorization approach, the Kaititj -ŋ/-l allomorphy in (31) could be expressed in terms of the two subcategorization frames shown in (33), adopting the notation of Paster (2006). Since the frames in (33) stand in a subset relation, the Elsewhere Condition or Pañini’s principle (Kiparsky 1973, etc.) dictates that where relevant, (33a) will take precedence over (33b), exactly as desired.

(33) Subcategorization frames for ergative (etc.) allomorphy in Kaititj

a. \[ [ \# \sigma \# ]_{stem} /-ŋ/_{ERG\ sfx} ]_{erg\ word} \\
b. \[ [ \ ]_{stem} /-l/_{ERG\ sfx} ]_{erg\ word}
A fundamental difference between the subcategorization approach and the output optimization approach is that the former assumes that morphology (here: allomorph selection) ‘precedes’, and hence feeds, phonology, whereas in the latter phonology and morphology operate in parallel. Under output optimization, information flows in both directions (M⇒P). It is thus possible for the choice of allomorph to be influenced by the eventual outcome of whatever phonological processes might be triggered by the allomorph selection itself (an instance of what Baković 2007 dubs ‘counterfactual derivation’). Indeed, this is more than a possibility; it is what is predicted to occur, other things being equal. One of the arguments that Paster (2006, 2009) adduces for subcategorization over output optimization is that this sort of crucially output-based allomorph selection appears to be unattested, whereas clear cases of crucially input-based selection do exist.

For example, in Turkish the third person singular possessive suffix has two allomorphs (each of which is subject to vowel harmony), [-i]/[-y]/[-u]/[-u] after consonants and [-s]/[-sy]/[-sur]/[-su] after vowels. However, Turkish also has a systematic process of Velar Deletion, whereby /k/ → Ø / V вок V (Sezer 1981; Inkelas and Orgun 1995), and this creates a potential dilemma for allomorph selection in the case of /k/-final nouns. For a noun like /bebek/ ‘child’, subcategorization predicts that the input representation that is supplied to the phonology will be /bebek-1/ ‘child-3sg’ which, due to Velar Deletion, will in turn surface as [be.be.-i]. This is indeed what we find. However, with the /k/ gone, from an output perspective the base is now vowel-final, not consonant-final, and this should have made the allomorph /-I/ equally inappropriate as it is after genuinely vowel-final nouns. Output optimization would thus predict selection of /-sI/ as a means of avoiding this situation, as illustrated in (34). MAX-IO is the Faithfulness constraint penalizing deletion; we know that this constraint outranks NoCoda in Turkish because closed syllables are not generally repaired by means of consonant deletion. The sad-face symbol indicates the candidate representing the actual Turkish output form, while the pointing hand designates the incorrectly predicted output, *[be.bek.-sI].

(34) /bebek-{I, sl}]/ → [be.be.-i] ‘child-3sg’

<table>
<thead>
<tr>
<th>/bebek-{I1, sl2}/</th>
<th>*VKV</th>
<th>MAX-IO</th>
<th>NoCoda</th>
<th>Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. be.be.k-I1</td>
<td>*!</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Θ be.be.-I1</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. ≡ be.bek.-sI2</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>d. be.be.-sI2</td>
<td></td>
<td>*!</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In other words, allomorph selection in Turkish 3sg /-{I, -sI}/ is rendered opaque by the phonological process of Velar Deletion, and this is exactly as predicted by a model in which allomorph selection logically precedes (i.e. feeds) the phonological input-output mapping rather than taking place as part of that mapping itself. Paster’s (2006) point is not that the opacity as such is a problem for the optimization approach,
but rather the typological generalization that allomorph selection is always opaque (input-based) in such cases, never transparent (output-based).

Another problem with the output optimization approach is that it is rather unconstrained in practice; whether a given pattern of allomorph distribution can be construed as phonologically optimizing or not seems limited only by the ingenuity of the analyst. Considerable freedom is created if one of the allomorphs can be freely designated as default (see Section 8.4.2). Another source of descriptive power are alignment constraints that can regulate the correspondence between morphological boundaries and those of phonological (e.g. prosodic) categories. In fact, a combination of these two strategies allow Bonet et al. (2007) to formulate an output optimization analysis that captures the seemingly anti-optimizing Haitian Creole pattern in (32). First, the /-a/ allomorph is stipulated to be the default (that is, the morpheme is represented as /-a > -la/). As a result, it is effectively only the /-la/ variant whose occurrence (after consonants, (32a)) is seen as being phonologically conditioned; the choice of [pa,pa.-a] over *[pa,pa.-la], or of [pa,pje,-a] over *[pa,pje,-la], is simply a matter of using the default (highest-priority) allomorph. Second (and following Klein 2003), the selection of /-la/ after consonant-final bases is triggered by a constraint R-ALIGN-STEM-SYLL ('Align the right edge of the stem with the right edge of a syllable'), which forces the stem-final consonant to be parsed as a syllable coda. The output [liv,-la] is thus construed as ‘better’ than *[li,v-a] on the grounds of morpho-prosodic alignment (R-ALIGN), and this is what makes the former emerge as the optimal output.12

As Embick (2010) points out, however, the appeal to phonology here is entirely ad hoc and without independent support from other facts of the language. The constraint R-ALIGN otherwise plays no role in the morphophonology of Haitian Creole. In fact, consonants routinely resyllabify across morpheme boundaries, resulting in rampant R-ALIGN violations in surface forms. While Bonet et al.’s (2007) full analysis does account for this, the larger question that remains is to what extent the analysis as a whole is learnable. The challenging task that a learner has to accomplish is manifold: (i) to infer that the alternation is due to allomorphy, not an active phonological process; (ii) to infer appropriate input representations, one for each allomorph; (iii) to infer an ordering relation (if needed) among the allomorphs; and (iv) to infer a ranking of constraints (consistent with other facts of the language) that accounts for the distribution of the allomorphs. All of these different aspects of the task are mutually dependent on each other (especially (ii)–(iv)), yet all are to be carried out in parallel. Formal learning models that are designed to accomplish this have yet to be developed. (In general, little attention has been paid to learnability concerns in the literature on

12 The alternative *[li,v-a], which uses the default allomorph /-a/ while still respecting R-ALIGN, is ruled out by a constraint *C.V prohibiting CV sequences from being parsed into separate syllables. While this constraint would appear to duplicate the work of Onset, which bans onsetless syllables, the two must be separate and straddle the default-protecting constraint (*C.V ≫ Priority ≫ Onset). Otherwise /-la/ would also be selected after vowel-final bases like /pap/ in order to avoid the onsetless final syllable of [pa,pa.-a]. Bonet et al. (2007) view *C.V as part of a larger family of sonority-based markedness constraints regulating syllable contact.
phonologically conditioned allomorphy.) By contrast, in place of (ii)–(iv) a subcategorization learner has only to discover the two frames in (35) or their equivalent, a far simpler and easier task.13

(35) Subcategorization frames for definite marker allomorphy in Haitian Creole
   a. \[\[\[C #]\]_{stem} /-\text{la/DEP sfx}]_{def \text{word}}\]
   b. \[\[\[V #]\]_{stem} /-\text{a/DEP sfx}]_{def \text{word}}\]

8.5 Conclusions

This chapter has touched upon a number of ways in which the phonology of a language can interact with aspects of its morphological structure, and different approaches to the modelling of such interactions in a formal-generative framework of grammar were contrasted. We have seen how phonotactic restrictions and phonological processes can be curtailed, triggered, or extended as dictated by particular morphological constructions and constituent-structure configurations. Also examined were some instances in which phonological criteria influence the exponence of morphological properties, such as in the case of phonologically conditioned allomorph selection as well as phonologically motivated paradigm gaps. Exponence in general is, of course, a morphology–phonology interface issue by definition. Since inflectional exponence is allotted its own chapter-length treatment (Trommer and Zimmerman, this volume), I have ignored such important topics here as prosodic morphology (e.g. reduplication, infixation, truncation, templates), featural affixation, mutation, and the like. Another problem not addressed in the present chapter, which is akin to that of allomorph selection (Section 8.4), is the question of phonologically conditioned affix order (Paster 2009). Of particular relevance is so-called ‘mobile affixation’, in which an affix is either prefixed or suffixed depending on phonological characteristics of the base; the best known example is Huave (Noyer 1994; Kim 2010; Trommer and Zimmerman, this volume). The existence of variable affix order governed by phonological well-formedness conditions remains a subject of debate; Paster (2009: 36) sceptically concludes that it ‘does not really exist’, noting that ‘fifteen years after the publication of Noyer (1994), Huave remains the single most convincing example’.

Research on morphology–phonology interface issues often raises fundamental questions about the division of labour between phonology and morphology in synchronically accounting for morphophonological alternations and related phenomena. This is particularly true for the interface of phonology with inflectional morphology, given the generally productive character of the latter. Much of the subject matter that forms the

13 The subcategorization frame in (35b) could easily be expanded into separate frames for /-\text{a/}, /-\text{ja/}, and /-\text{wa/} based on the quality of the vowel. Here it is assumed that the insertion of a homorganic glide [j] or [w] as hiatus-breaker is an easily recoverable aspect of Haitian Creole phonology.
The empirical basis of theoretical phonology consists of sound patterns and alternations that, upon closer scrutiny, are limited to some extent by morphological and/or lexical factors (in ways that are often glossed over in analytical works adducing them as examples). From a phenomenological perspective, one might say that there exists a cline from regular and transparently phonological alternations at one end (which it is reasonable to attribute to the phonological grammar) to morphologically stipulated and phonologically suppletive ones at the other (in which the phonological grammar plays no obvious role). Along the middle of this cline fall the kinds of situations that have been the subject of this chapter; for example, ones that seem to involve construction-specific phonology or phonologically conditioned allomorphy. These in turn range from cases that look suspiciously 'phonological' to others for which the involvement of phonological computation and well-formedness considerations seems more remote (e.g. the allomorph selection patterns in Basque and Haitian Creole, respectively).

The existence of such a cline (as a purely observational notion) is not surprising, since morphophonological alternations most often originate in low-level, automatic (often allophonic) alternations which are themselves the product of phonetically motivated sound changes (Hansson 2011), but whose defining properties have become partly obscured by subsequent historical changes. The question for phonological and morphological theory is where, and by what criteria, to draw the line demarcating which phenomena involve calling upon the phonological grammar (some version of our function $\phi_L$ from Section 8.2 above) and which do not. So long as phonology is modelled as a function relating input (lexical, phonemic) representations to output (surface, phonetic) representations, and so long as at least some alternations in surface shape (allophonic and/or neutralizing) are treated in terms of unique underlying forms undergoing changes in the input–output mapping, such a demarcation line must exist. For the past half-century, the standard heuristic in generative phonology has been to ascribe any alternation to the phonological grammar by default, even at the expense of relatively abstract input representations or complex interactions of rules/constraints; analyses in terms of allomorphy or morphological stipulation are treated as a last resort. Many of the proposals described in this chapter, such as the invoking of constraints on paradigm uniformity or anti-homophony, or the notion of output optimization in allomorph selection, are best understood in this light. Disagreements on the particular theoretical issues are typically played out in the arena of descriptive adequacy and typological coverage (as well as formal parsimony). What has been less often addressed is the question of explanatory adequacy and how the language learner comes to reason her way, on the basis of positive data, to the assumed analysis of the morpho-phonological phenomenon in question. With the increased interest in the formal modelling of phonotactic and morpho-phonological learning in recent years (for an overview, see Albright and Hayes 2011), this is likely to change.