

2. The Organization of the Grammar

K. P. MOHANAN

Subject	Theoretical Linguistics » Phonology
Key-Topics	grammar
DOI:	10.1111/b.9780631201267.1996.00004.x

0 Introduction

The central issue I wish to explore in this article is the nature of the relation between subsystems of grammar, in this instance the subsystems of phonology and morphology.¹ This issue has close links with two other issues: (1) the classification and formal expression of regularities in representations in terms of systems of rules/ constraints, and (2) the levels and content of representations.

The first question on the issue of the relation between subsystems is: do phonology and morphology interact? One can identify two answers to this question in the literature. The first is that they do not interact: regularities of sound patterning sensitive to morphology are part of morphophonology, not phonology. This was the position held, for instance, in neo-Bloomfieldian classical phonemics. A variant of this claim is that phonological regularities must be classified into two types on the basis of interaction with morphology: the phonological rules/ constraints that interact with morphology from a separate subsystem independent of the phonological rules/ constraints that do not interact with morphology. An example is the distinction between rules and processes in natural phonology (Stampe 1972), which correspond in classical phonemics to morphophonological and phonological rules respectively. The second, answer, in contrast, is given by classical generative phonology (Chomsky and Halle 1968) and its descendents lexical phonology (Pesetsky 1979; Kiparsky 1982, 1985; K. P. Mohanan 1982, 1986; Pulleyblank 1983, 1986a; Booij and Rubach 1984), nonlexical prosodic phonology (Selkirk 1980; Aronoff and Sridhar 1983; Sproat 1986), prosodic lexical phonology (Inkelas 1989), harmonic phonology (Goldsmith 1989, 1993a; Wiltshire 1992), and so on; they reject the classification of rules/constraints into morphophonological and phonological subsystems, postulating a single phonological subsystem of rules/constraints, and claiming that phonology does interact with morphology. I will show that there is sufficient evidence to hold the latter position.

If phonological patterns can be sensitive to morphological information, how is this information made available to phonology? That is, how do the two subsystems interact? There are two broad classes of answers, those that appeal to entities of representations, and those that appeal to sequential modularity. Classical generative phonology appeals to tree structure representations with labels such as N and NP. In addition, morphological information about affixation versus compounding, different kinds of affixation, and the like, are expressed representationally in terms of juncture symbols such as +, #, and ##. This information is represented in prosodic phonology by labels such as phonological root, stem, and word. In these theories, the domains of phonological rules/ constraints are specified in terms of representational units. In contrast, lexical phonology factors out his information into sequentially ordered modules, allowing the domains of phonological rules/constraints to be specified in terms of the modules. Prosodic lexical phonology employs both representations as well as sequential modularity. As it turns out, there exist configurations of facts that are analyzable in terms of sequential modularity but not representations, and others that are analyzable in terms of

sequential modularity but not representations, and others that are analyzable in terms of representations but not sequential modularity. How this dilemma can be resolved is an open question.

Another central preoccupation in phonological theory relevant to the question of the phonology–morphology interaction involves the need for a level of phonological representation that is most readily accessible for speech production, recognition, acquisition, and speaker judgements. This was perhaps the intuition underlying the levels of phonemic representation in classical phonemics. The theory of generative phonology proposed by Chomsky and Halle in *The Sound Pattern of English* (1968, henceforth SPE) made a radical departure from classical phonemics by not including this level of representation. This led to a serious dissatisfaction with SPE phonology and a series of rebellions against it. The late seventies and early eighties witnessed a return to a preoccupation with levels of representation, particularly in the form of lexical phonology, which offered a new conception of the phonology–morphology interaction (Siegel 1974; Pesetsky 1979; K. P. Mohanan 1982; Kiparsky 1982; Pulleyblank 1986a). The main contribution of lexical phonology to the question of levels of representation was the idea that even though phonological rules/constraints themselves do not fall into types, the interaction between phonology and morphology leads to a level of phonological representation that is distinct from underlying and phonetic representations. The last two decades have seen various versions of lexical phonology propounded, as well as many critiques and alternatives (Rubach 1984; Sproat 1985, 1986; Hargus 1985; Kaisse and Shaw 1985; Borowsky 1986; Halle and Vergnaud 1987; Fabb 1988; T. Mohanan 1989; Inkelas 1989; Goldsmith 1990; Wiltshire 1992; to mention a few). It appears to me that even though we may have to reject many of the assumptions of lexical phonology, including its sequential conception of grammar, we will need to retain the idea of an intermediate level of representation, as is done in some versions of phonological theory that reject sequentiality (Lakoff 1993; Goldsmith 1989, 1993a).

The issue of the phonology–morphology interaction is also closely involved with the formal statement of regularities of distribution and alternation, that is, the nature of the rule/constraint system. Classical generative phonology is deeply entrenched in the metaphor of grammar as a production system that gives rise to a conception of grammar in which modules as well as principles are sequential. In this conception, a principle is typically stated as a rule, that is, a procedure that takes an input and yields an output. The same input–output relation holds for levels of representations as well: a module with its set of principles takes a level of representation as the input, and yields another level of representation as the output. Between two levels of representation defined by the grammar are intermediate stages in the derivation which have no theoretical status in the overall model.

This sequential conception in phonology is currently being challenged by a growing body of work that views grammatical principles not as procedures, but as well–formedness statements analogous to the laws of physics. Unlike an instruction in a computer program, the laws of gravity and magnetism do not take an input and yield an output. They are simply statements of certain relations holding on a set of constructs. Kisseberth (1970) and Sommerstein (1974) are often cited as being influential in triggering the use of various types of well–formedness statements to build a nonsequential conception of phonology in works such as Singh (1984), Karttunen (1989), Lakoff (1989), Goldsmith (1989), K. P. Mohanan (1989, 1992), Bird (1990), Coleman (1991), Scobie (1991), Smolensky and Prince (1993), and Goldsmith, ed. (1993b). The elimination of input–output relations between levels of representation also results in the elimination of intermediate stages in a derivation that connects two levels. In such a conception, levels of representation are simultaneously present in a multidimensional space, as in Halle and Vergnaud's (1980) ringbound notebook metaphor. As a consequence, a principle of the grammar can state a relation holding between two entities internal to a given of representation, or between two levels of representation. The sequentiality of SPE and lexical phonology has no place in this emerging declarative conception of grammatical principles, modules of the grammar and levels of representation.²

In this article, I will explore the basic issues and hypotheses alluded to above, provide a critical evaluation, and attempt to outline what I consider to be the unresolved problems. The questions that I will address are:

- What is the relation between phonology and morphology?

- How is morphological information made available to phonology?
- How do controversies on rule/constraint types, levels of representation, and sequentiality affect our view of the phonology–morphology interaction?

I must warn the reader that it is not my purpose here to introduce the reader to current theories of phonology or to defend any of them. My purpose is rather to identify and explore a set of issues on the organization of grammar which have come to revolve around various debates on classical phonemics and morphophonemics, classical generative phonology, lexical phonology, and subsequent work. In doing so, my primary focus will be the concepts that are often obscured by debates on formalisms. The reader who is looking for a summary of or introduction to the latest formalisms in phonology may, I am afraid, be disappointed.³

1 Preliminaries

1.1 Three Pointers from SPE

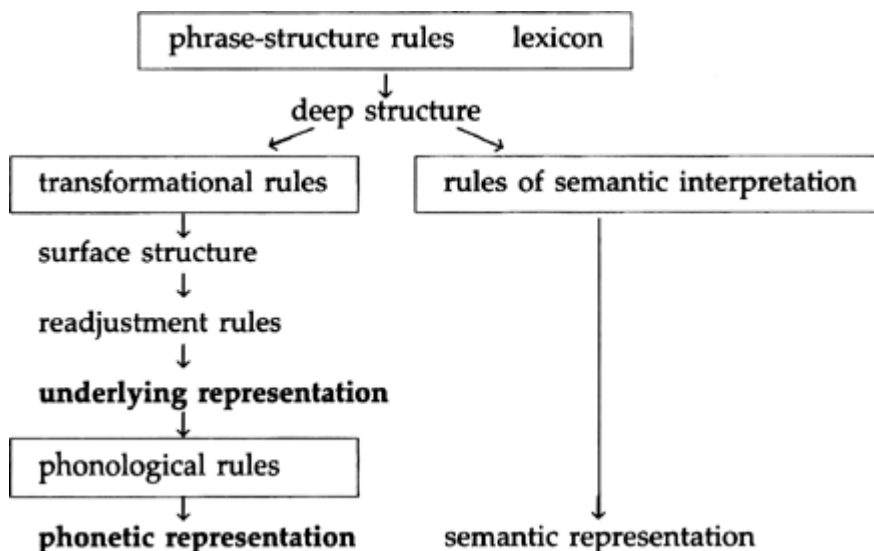
The three themes that will connect the pieces in this article are:

- 1 Dependence on morphology and syntax: There is a correlation between the structures by syntax and morphology on the one hand, and by phonology on the other.
- 2 Nonconvergence: There is a mismatch between the structures motivated by morphology and syntax, and the structures needed for phonology.
- 3 Construction types: Phonology requires additional information about different kinds of morphological “construction types” such as affixation and compounding, different kinds of affixation, and so on.

As with most issues in current phonology, insights on these three issues can be found in *SPE*. therefore, it is useful to see how *SPE* dealt with them.

The place that *SPE* assigns to phonology within the overall organization of grammar is given in (1):

(1)

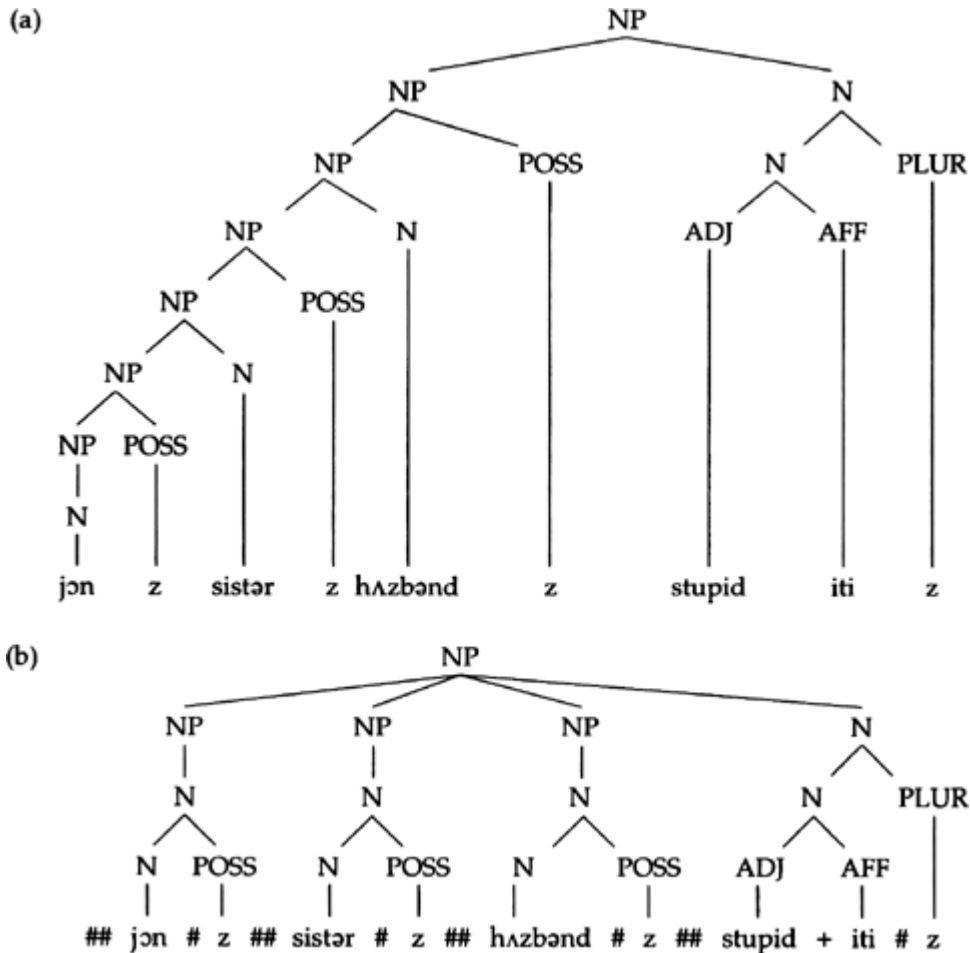


That surface structures are an (indirect) input to the phonological component shows that *SPE* recognized the correlation between morphosyntactic patterns and phonological patterns. That these surface structures are modified by readjustment rules shows that *SPE* recognized the mismatch between morphosyntax and phonology. That these readjustment rules introduced juncture symbols like + and # shows that *SPE* recognized the need to refer to information about construction types not

provided by the morphosyntax.

To illustrate, the surface structure representation of the English phrase *John's sister's husband's stupidities* is something like (2a). The readjustment rules convert (2a) into the underlying representation in (2b):

(2)



As can be seen from the underlying representation in (2b), *SPE* assumes a close interaction between phonological and nonphonological systems. First, it allows phonological rules to access information about (readjustment) syntactic constituency, grammatical categories, and grammatical features. However, the tree structures that encode this information are not identical in (2a) and (2b). The readjustment rules applying to (2a) flatten the tree structure to the one required for correct stress placement.⁴

In addition to the readjustment of constituency relations, readjustment rules also introduce juncture symbols such as +, #, and ##, which encode information about the morphological construction types crucially needed for both segmental and suprasegmental phonology. For example, the difference between class 1 derivation (e.g., *-ity*) on the hand, and class 2 derivation (*-ing*) and inflection (*-es*) on the other, is represented in terms of boundary symbols like + and # (e.g., *stupid_iti#z*). Unlike the information about constituency and grammatical categories, the information about construction types is expressed in terms of diacritic symbols which do not have any universal crosslinguistic substance.⁵

Thus, the use of readjustment rules in *SPE* indicates the recognition that the structure that is relevant for the application of phonological principles is governed by, but not identical to, that of pure morphosyntactic representation. This recognition was explicitly voiced in *SPE*:

we have two concepts of surface structure: input to the phonological component, and output of the syntactic component. It is an empirical question whether these two concepts coincide. In fact, *they do coincide to a very significant degree, but there are also certain discrepancies*. These discrepancies... indicate that the grammar must contain certain rules converting the surface structures generated by the syntactic component into a form appropriate for use by the phonological component. In particular, if a linguistic expression reaches a certain level of complexity, it will be divided into successive parts that we will call "phonological phrases," each of which is a maximal domain for phonological processes.

[Italics mine: KPM] (P.9)

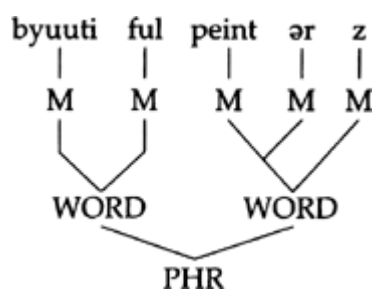
We will pursue the three themes of phonology–morphology interaction pointed out by *SPE* in the rest of this article.

1.2 Strategies for Accessing Wordhood

Accessing morphological information involves distinguishing between words and units larger than words, and identifying information internal to words. Let us begin by exploring different ways of accessing the notion WORD. This construct is relevant for phonology for at least three well-known reasons: some phonological patterns crucially hold only *within words*, some hold only at the *edge* (beginning or end) of a word, and some are governed by information about *word-internal structure* for their application. An example of the first type is homorganic nasal assimilation in Malayalam, which is obligatory within a word, and optional across words (K.P. Mohanan 1993). An example of the second type is the flapping of *t* followed by a lateral in American English: in this environment, flapping takes place if *t* is word final (e.g., *The ca[D] licked the butter*) but not if it is word internal (e.g., **ha[D]less, *bu[D]less*). An example of the third type is the gemination of obstruents in Malayalam, which applies at the junction between two stems in a subcompound where the second stem is Dravidian (K. P. Mohanan 1986).

Strategies for capturing the notion word can be divided into two broad classes, one in terms of representation labels within a single level of representation, and the other in terms of different levels of representation. The structure of the phrase *beautiful painters* in terms of representational labels is:

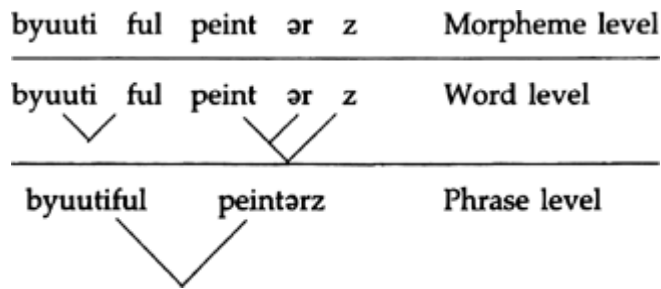
(3)



The labels PHR, WORD, and M stand for phrase, word, and morpheme respectively. Given such a representation, a principle that holds within a word but not across words can be specified as having WORD as its domain. In addition to the use of the symbol ## to indicate external word boundary, *SPE* makes use of this strategy to specify the domain of phonological rules in terms of syntactic constituency relations and lexical category labels such as N, ADJ, and V, as opposed to phrasal phonology, where representational labels like P-WORD and P-PHRASE correspond to wordhood and phrasehood relevant for the purpose of phonology.

The structure of the same phrase in terms of different levels of representation is:

(4)

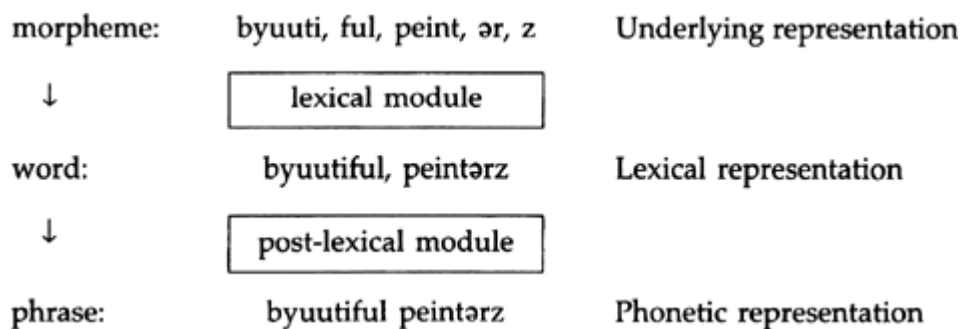


In (4), the information about wordhood is present not as labels on nodes, but as a unit along a distinct level of representation. Given such a representation, a principle that holds within words but not across words can be specified as holding at the word level.

Imposing sequentiality to the three levels of representation in (4) yields the fundamental premises of lexical phonology:

- (5) (a) The grammar consists of two modules, the lexical and the post-lexical.
 (b) The lexical module precedes the post-lexical module.
 (c) The domain of phonological principles can be specified as the lexical module, or the post-lexical module, or both.

(6)



The input-output metaphor in (6) has dominated generative grammar since its inception. If we abandon this input-output metaphor of procedure, and interpret (4) as parallel or simultaneous levels of representation, we have the conception of the organization of grammar in cognitive phonology (Lakoff 1993) and harmonic phonology (Goldsmith 1989, 1993a).

As pointed out in the introductory section, a large number of proposals in the recent literature have suggested that we abandon the sequentiality of rule application in phonological theory in favor of general laws for linguistic representations analogous to the laws of gravitation or subatomic phenomena. In a conception of linguistic organization that rejects sequentiality of rule application, the idea of sequentiality of levels of representation is an anomaly, if not a logical contradiction. This would require abandoning assumption (5b) above, while retaining (5a) and (5c), the "precede" of (5b) being replaced by the part-whole relation between words and phrases. We will come back to this issue at a later point.

In sum, there is reasonable consensus that principles of phonological organization must be able to access information about wordhood. This information can be made available to phonology either in terms of node labels on a tree structure, or in terms of different levels/dimensions of representation that factor out different aspects of structure. The second choice leads to the further choice of assuming either that these levels of representation are sequentially ordered in terms of the input-output relation, or that they are simultaneously present as parallel levels.

2 The Internal Structure of Words

There are few phonologists who would dispute the claim that principles of phonological organization need access to the notion WORD. Even classical phonemicists who believed in the strict autonomy of phonology allowed statements about the distribution of allophones to refer to notions like word final and word initial. However, word-internal structure is a different matter. Are principles of phonology allowed to refer to morphological structure? The last few decades have seen a number of debates on this question, where phonologists have tended to be divided into two camps.

2.1 Morphophonology

The question of the sensitivity of phonology to word-internal structure comes up in the description of both distribution and alternation. But it is most clearly manifested in the treatment of alternation. Let us begin with prototypical patterns of alternation in English which are acknowledged to belong to phonology proper in all approaches to phonology, as they make no reference to morphological information:

(7)

- | | |
|-------------------------------|-----------------------------|
| (a) rapid ['ræpid] | rapidity [rə'pɪdɪti] |
| (b) Pat is a wise man. | Pat is wise. |
| [wai.z] | [wai.ʒ] |

In (7a), there is an alternation between aspirated and unaspirated stops; in (7b), there is an alternation between fully voiced and devoiced obstruents. The statement of these patterns does not require any reference to morphological structure:

(8)

- (a) Voiceless stops are aspirated at the beginning of a stressed syllable.⁶
 (b) Obstruents are devoiced when adjacent to a pause.⁷

The crucial constructs required in (8), namely, distinctive features, stress, syllable, and pause, are part of phonological information. In contrast, the *t/s* alternation in (9a) and the *s/z* alternation in (9b) crucially require reference to morphological information:

(9)

- (a) resident residency
 [rezident] [rezidensi]
 (b) house (n) house (v)
 [haus] [hauz]

If we intend to capture these alternations in phonology, we need to make statements that are equivalent to (10a) and (10b), crucially appealing to morphological constructs such as stem and affix:

(10)

- (a) Underlying stem-final **t** is **s** before **i** in class 1 affixation in [+Linate] forms.

(b) Underlying stem-final *s* of the verb is *z* in a special class of morphemes.

The patterns in (8a, b) are purely phonological, while those in (10a, b) are morphophonological.⁸ At the center of debates on phonological theory has been the treatment of morphophonological patterns. Bloomfieldian phonemics separates phonology and morphophonology, while *SPE* makes no distinction between the two. A large number of phonological theories since *SPE*, such as natural phonology, natural generative phonology, upside-down phonology, lexical phonology, generative phonotactics, and so on have been attempts to express the distinction between phonological and morphophonological phenomena.

Before we proceed, it is important to separate the separate the distinction between phonological and morphophonological patterns from the distinction between allophonic and phonemic patterns. Consider the following examples of voiced/voiceless alternation in English:

(11) (a) **twelve** **twelfth** (b) **dogs** **cats** (c) **Bill's away.** **Pat's away.**
 [twelv] [twelfθ] [dɔgz] [kæts] [bilz] [pæts]

While the alternations in (7a) and (7b) are clearly allophonic in the sense that the alternating units are not contrastive in the language, the alternation in (11) is clearly phonemic, because in English, *f* and *s* contrast with *v* and *z* respectively. Until recently, both within classical phonemics and generative phonology, these alternations were considered morphophonemic (e.g., Anderson 1974; K. P. Mohanan 1982). As more recent studies have shown, however, the best way of stating the pattern in (11) is in terms of syllable structure, rather than morphological structure:

(12) Syllable-internally, adjacent obstruents agree on voice.⁹

If we accept (12), the phenomenon in (11) is an example of phonemic alternation that is not sensitive to morphology. Other examples of this type include optional external sandhi phenomena such as the *s/š* alternation in *horse* [hors] / *horseshoe* [horššu]. This shows that the distinction between purely phonological and morphophonological does not coincide with the distinction between allophonic and phonemic:

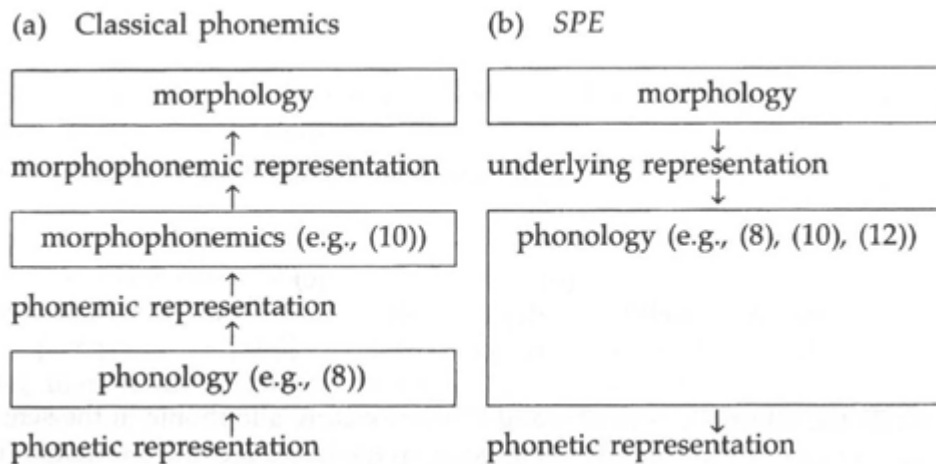
(13)

	Phonological	Morphophonological
Phonemic	(11)	(9)
Allophonic	(7)	?

2.2 The Classification of Principles

As is well known, phonology in classical phonemics meant purely phonological allophonic patterns. Morphophonemic patterns belonged to a separate component of the grammar, sandwiched between phonology and morphology. The status of purely phonological phonemic patterns was somewhat uncertain. *SPE* eliminated this distinction, allowing phonological principles to freely access morphological information:

(14)

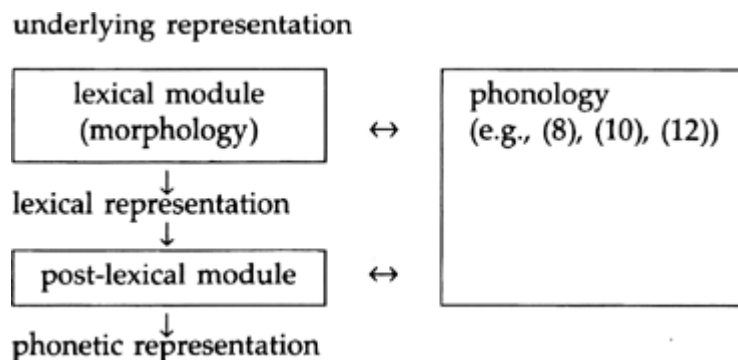


2.3 Sequential Modularity

Phonologists have always had a strong intuition that patterns such as in (9) are significantly different from patterns such as in (7). This involves the distinction between morphophonemic and phonemic. They have also recognized that the phonemic representation of classical phonemics tried to express a level of representation that reflected a language user's conscious access to certain structural elements of speech, including the number of "sounds," and their sameness and distinctness. This involves the distinction between phonemic and allophonic.

These two issues have often been conflated in discussions, including those within lexical phonology. Since *SPE* had erased the lines that demarcated these distinctions, the intuitions led to a series of proposals that were counter to *SPE*. The questions that these proposals have tried to come to grips with are: Where does morphophonology belong? How do we express the "contrast" that reflects the language user's consciousness? Most of the proposals tried to answer these questions by designing different classificatory systems for phonological rules (for instance, rules versus processes, automatic versus nonautomatic alternations). In the early eighties, lexical phonology proposed a different solution by separating the statement of the principles from the domains in which the principles hold, and deriving the differences in the effects of phonological principles from differences in domains:

(15) Lexical Phonology



The conception of the organization of phonology in (15) makes it possible for a phonological principle to have both morphophonemic and purely phonological effects. A telling example is that of the place assimilation of nasals in Malayalam which has been argued to hold in both the lexical (morphophonemic) and post-lexical (purely phonological) modules. Malayalam has an inventory of seven nasals which are contrastive on the surface: bilabial, dental, alveolar, alveopalatal, retroflex, palatal, and velar. Mohanan and Mohanan (1984) demonstrate that the otherwise idiosyncratic

distributional restrictions on these nasals can be accounted for by assuming an underlying inventory of three nasals, namely, bilabial, alveolar, and retroflex. The rules that yield the right results within this account are independently motivated:

(16)

- (a) A nasal assimilates in place to the following plosive.
- (b) A voiced plosive becomes nasal when preceded by a nasal, optionally in [-Dravidian] and obligatorily in [+Dravidian] words.

(16a) is a purely phonological rule in that it does not require morphological information. Furthermore, it applies across words: /**mad^hu^ram**/ “sweet” and /**ṭa^ru**/ “give (imp.)” → [**mad^hu^ranṭa^ru**] “give (me) sweets!”. (16b) is morphophonemic as it requires information about morphological subclassification ([±Dravidian]). Furthermore, it does not apply across words or across the stems of a compound. For (16a) and (16b) to correctly predict the distributional restrictions in a procedural account, it is necessary to assume that (16a) feeds (16b), assimilating the place of the nasal to the following plosive, and then assimilating the manner of the plosive to the preceding nasal (e.g., /**pandⁱ**/ → **pandⁱ** → [**panⁿi**] “pig”). In a declarative account, this would amount to saying that (16a) and (16b) hold simultaneously on the mapping between morphophonemic and phonemic representations. Either way, the main point is that (16a) should be recognized as holding between the underlying and phonemic levels as well as between phonemic and phonetic levels. This option is available to (15), but not to (14a). NOR is it available to other theories such as natural phonology and natural generative phonology, which draw a dividing line between two kinds of phonological principles rather than between two domains in which the principles take effect. Thus, it is important to realize that, as Anderson (1981) points out, a phonological principle that does not refer to morphological information in its formulation may nevertheless interact with morphophonemic principles in such a way that it creates the effect of a morphophonemic pattern.

Close cooperation between morphophonemic and purely phonological principles is also found in the phenomenon of voicing assimilation illustrated in (11). The general principle proposed for these alternations was that syllable-internally, obstruents agree in voice (12). When we put together the underlying representation of the stems /*twelv*/ and /*kæt*/ with the underlying representations of the affixes, namely, /*θ*/ and /*z*/, the result *twelvθ* and *kætz* violates (12). The ill-formedness can be repaired by changing these forms to either *twelvð* and *kædz*, or to *twelfθ* and *kæts*. Why does English choose the latter option, and not the former?

One way of avoiding this question is to revise (12) as (12'), stipulating that the winner in the conflict is [-voice].

(12') Syllable internally, an obstruent is [-voice] when adjacent to a [-voice].

This was the solution explored in K. P. Mohanan (1991). This formulation, though descriptively adequate, is less desirable on explanatory grounds. Observe that (12) is a universal principle that holds in (almost) all languages. In contrast, (12') is formulated in such a way that what is common to voicing assimilation in English, and say, in Hebrew, is not reflected in the formulation. In English, [-voice] spreads to [+voice] both left to right and right to left; in Hebrew both [+voice] and [-voice] spread, but only from right-to-left. The way they are formulated, (12) holds for both English and Hebrew, but (12') holds only for English.

The explanation, I suggest, lies in an independent principle:

- (17) Adjacent obstruents within a syllable are voiceless.
(Absolute morpheme-internally, weak across morphemes).

Principle (17) allows monomorphemic syllables like [ækt], [lift] and [risk], but disallows syllables like * [zbin], *[rizg] and *[lizb], which cannot be bimorphemic. It allows syllables like [bægd] and [livd], but only when the cluster is formed by the addition of an independent morpheme. These facts can be interpreted as follows. Constraint (17) holds strongly within morphemes and weakly across morphemes. (17) is not strong enough to repair concatenations like *bæg+d* and *liv+d* and make the clusters voiceless. However, it exerts a force in the choice between [+voice] and [-voice] in *twelvθ* and *kætz*, yielding *fθ* and *ts*, rather than *vŋ* and *dz*.¹⁰ (12) requires that the coda be changed. Given that *vθ* and *tz* are deemed to be undesirable by (17), the coda is repaired to *fθ* and *ts*.

If this line of explanation is accepted, then we must accept that (17) holds strongly in the morphophonemic domain (within morphemes, in the lexical module) and weakly elsewhere. The actual voicing assimilation in (11), which is a purely phonological alternation, is a combined result of (12) and (17). Hence the separation in (14a) is not a viable alternative.

2.4 Nonsequential Modularity

Lexical phonology incorporates the traditional sequential (i.e., derivational) conception of the organization of grammar in two senses. Firstly, the modules of the grammar and the consequent levels of representation are viewed in terms of the input–output relation, where each module or level of representation is seen as preceding or following another. Second, the output of one principle constitutes the input to another. The logical connection between two levels of representation is stated as a number of ordered steps in a derivation. The order in such a derivation is crucial, unlike in a logical or mathematical derivation where the order of application is irrelevant. As stated earlier, a large number of phonologists have started abandoning the sequential conception of linguistic derivation, and it stands to reason that these phonologists also reject the sequential conception of modularity and hence that of levels of representation. If we abandon the sequentiality in underlying, lexical, and phonetic representations, and hold that these three types of information are represented along parallel dimensions of structure, we have the organization of grammar in cognitive phonology and harmonic phonology:

(18) Harmonic/Cognitive phonology

M–level (corresponds to underlying level)
 W–level (corresponds to lexical level)
 P–level (corresponds to phonetic level)

From the discussion of the underlying, word, and phrase levels in Goldsmith (1990, p. 231), I conclude that harmonic phonology draws the line between two domains of phonology as in (15), rather than between two kinds of principles as in (13a). Factoring away the procedural implication in (15), what is common to (15) and (18) is the idea of three levels of representation, in contrast to (14b).

3 Morphology versus Morphologically Sensitive Phonology

3.1 Are all Nonautomatic Sound Patterns Part of Morphology?

In a series of articles, Singh (1984, in press) and Ford and Singh (1983, 1985) have put forward a view of the organization of phonology that departs from *SPE* in a more extreme form than lexical phonology, or even classical phonemics. According to this view, called generative phonotactics, alternations such as in (9a, b) should be handled not as part of phonology, but as part of the word formation rules in morphology. The formal statements that account for (9a) and (9b) are given as (19a) and (19b) respectively:¹¹

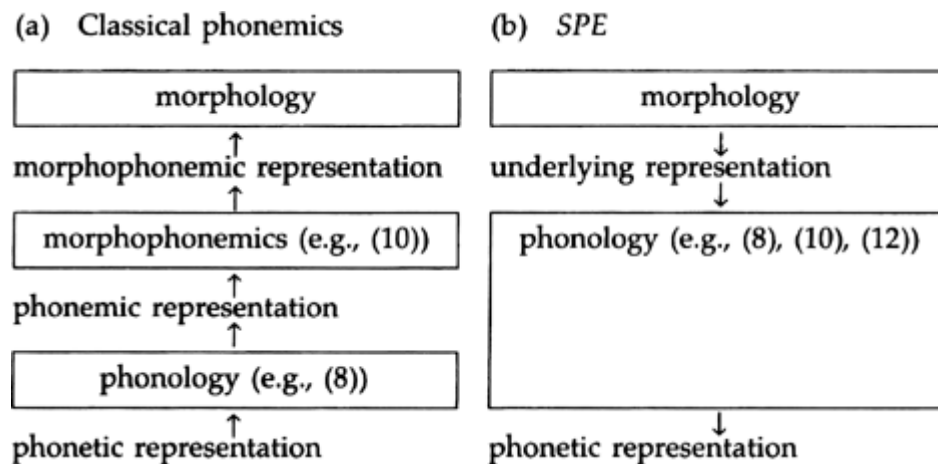
(19)

- (a) Word formation rule A
 $[Xt]_{\text{CONCR}} \leftrightarrow [Xsi]_{\text{ABSTR}}$
- (b) Word formation rule B
 $[Xs]_{\text{NOUN}} \leftrightarrow [Xz]_{\text{VERB}}$

This position denies a theoretical status to the construct morpheme, either in phonology or in morphology. Thus, affixation is the morphological operation of attaching a phonological string to a word to form a new word, not that of the attachment of an affix. As part of the morphological operation, the stem may undergo phonological changes as well.

Given the availability of the formal device in (19), generative phonotactics makes the extreme claim that no phonological principles can refer to morphological structure.¹² In this view, all phonological patterns that are sensitive to morphological information are part of morphology, rather than pure phonology or morphophonology.

(20)



In other words, there is no morphophonology in this view. If this claim is correct, then the question of phonological rules accessing information about morphological structure is a misguided one. Therefore it is necessary to respond to this claim before we proceed further.

It so happens that if we are willing to tolerate a certain degree of duplication, the formal device of correspondence statements illustrated in (20) can be extended to cover most examples of segmental alternation triggered by affixation, particularly those which have been at the heart of *SPE* phonology such as trisyllabic shortening, velar softening, and spirantization. There exist, however, many phonological patterns which cannot be dealt with in terms of correspondence relations between words. We have already seen one of them, namely, (17), which makes crucial reference to the construct morpheme. As pointed out earlier, the contrast between well-formed coinages like [nisp] and [nisk] and ill-formed coinages like *[nizb] and *[nizg] require the assumption that (17) holds strongly within a morpheme. Consider the additional principles (21a-c) that govern syllable structure in English, and (21d), a morphemeinternal regularity in Malayalam:

(21)

- (a) Morpheme-internally in English, a coda can have at most three consonants.
- (b) In English, a consonantal segment can be syllabic only at the end of a morpheme/stem.
- (c) Morpheme-internally in English, dental fricatives cannot occur after an obstruent in a coda.

- (d) In Malayalam, a single dental nasal followed by a vowel can only occur morpheme-initially; in this environment, an alveolar nasal cannot occur.

Principle (21 a) allows the four-consonant coda in bimorphemic words like *texts*, but does not allow a monomorphemic [beksts] (e.g., **Four bextses* [bekstsiz]). (21 b) allows a syllabic lateral in bimorphemic *fickleness* [fiklnəs], but disallows it in *[klnep]. (21 c) allows syllables like *fifth* [fifθ] and *depth* [depθ] only if they are bimorphemic. All these principles are nonautomatic in that they crucially refer to the morphological construct morpheme. Principle (21 d) allows morpheme-initial [n̥] in [n̥ayanam] “eye” and [sun̥ayani] “person with beautiful eyes”, but disallows forms like *[nayala] and monomorphemic *[pun̥ala]. None of the facts accounted for by these principles can be accounted for in terms of correspondence conditions (19). Hence the extreme claim that phonological principles are not sensitive to word-internal structure is untenable.¹³

Turning to patterns of alternation, we note that unlike the typical segmental morphophonemic alternations, many facts of word stress in English are not expressible in terms of the formalism in (19). Take, for example, the effect of suffixes like *-al*, *-ous*, and *-ant*. The productive generalization is that in words with these suffixes, stress falls on the penultimate syllable if it is heavy, and, if not, on the antepenultimate syllable.¹⁴ Thus, stress shifts from the first syllable in *parent* to the penultimate syllable in *pa'rental* because the penult is heavy, but from the first syllable in *medicine* to the antepenultimate syllable in *me'dicinal* because the penult is light. The reader will find a massive number of additional examples of this kind in *SPE*.

Now, as is well known, only class 1 suffixes in English (*-ous*, *-ant*, *-al*, *-ity*, *-ive*, *-ion*, *-ic*, etc.) are capable of shifting the primary stress of the stem to some other syllable; class 2 suffixes (*-ness*, *-less*, *-hood*, *-dom*, *-ful*, etc.) do not change stress. Thus, the addition of the class 2 affix *-less* to *parent* and *medicine* has no effect on stress: *parentless*, *medicineless*. Clearly, stress placement in English is sensitive to the morphological distinction between class 1 and class 2 affixes. It is also sensitive to the phonological distinction between light and heavy syllables. I see no way of combining the two into a morphological principle in terms of the formalism illustrated in (19).

Similar conclusions are to be drawn from patterns of alternation which do not involve affixation. Take, for example, the exceptionless patterns of stress and word melody assignment in subcompounds and co-compounds in Malayalam. Subcompounds have a single primary stress and word melody per compound, while co-compounds have a primary stress and word melody for each stem in the compound (K. P. Mohanan 1986). I see no way of expressing this generalization as a purely morphological rule, rather than as a phonological rule sensitive to morphology. We conclude, therefore, that the conventional wisdom of allowing phonological rules to access word-internal structure must indeed be retained.

3.2 Are all Sound Patterns Part of Phonology?

A word of caution is appropriate at this point. That there exist morphologically sensitive phonological principles does not entail that all relations between morphology and phonology should be stated as morphologically sensitive phonological rules. This point is particularly important in the context of the debate between affixal morphology and nonaffixal morphology.

As far as I can see, the central hypothesis that separates the two approaches to morphology is the following:

- (22) The only permissible morphological operation is that of combining affixes and stems.

Affixal morphology (Lieber 1980; Marantz 1982; Kiparsky 1983; and so on) subscribes to (22). Thus, alternations such as ablaut, spirantization, or deletion cannot be morphological operations, and hence must necessarily be treated as part of morphologically sensitive phonology. In contrast, nonaffixal morphology (Matthews 1974; Anderson 1992; Zwicky 1988) rejects (22).

The interpretation of the claim in (22) depends upon the interpretation of the term “affix.” I would like

to suggest that a great deal of unnecessary controversy can be eliminated if we (1) separate three levels of description, namely, features, morphemes, and formatives, and (2) define affix as a type of formative (T. Mohanan, 1993). The three-level terminology is as follows:

Features: entities like [PAST], [PERFECTIVE], [PASSIVE] and [PLURAL] which are part of a universal inventory

Morphemes: abstract language-particular morphological units; e.g., the English morphemes {SIT}, {GO}, {TABLE}, {ED}, and {EN}

Formatives: phonological strings which act as morphological units; e.g., the English formatives /sit /, / gou /, / d /, / n /, and / z /

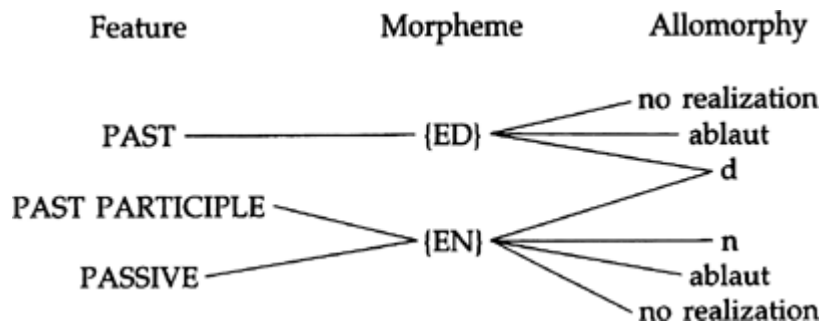
The three levels of abstraction are illustrated by the following examples:

(23)

	(a)	(b)	(c)	(d)	(e)	(f)
	<i>kissed</i>	<i>worte</i>	<i>kissed</i>	<i>wortten</i>	<i>wortten</i>	<i>went</i>
Feature	KISS PAST	WRITE PAST	KISS PASS	WRITE PASS	WRITEN PERF	GO PAST
Morpheme	KISS ED	WRITE ED	KISS EN	WRITE EN	WRITE EN	GO ED
Formative	kis d	rout	kis d	writ n	writ n	went

Why do we need to separate the three concepts? The morphological information relevant for syntax and semantics is encoded in features. The additional morphological information relevant for phonology is encoded in formatives. Within the three-level view illustrated in (23), morphemes mediate between features and formatives, providing the anchor for allomorphy. For example, the syntactico-semantic features [PERFECTIVE] and [PASSIVE] have the same morphemic realization in English. Thus, the allomorphic realization of the perfective form of the verb in *John has written/burnt/mailed the letter* is identical to the allomorphic realization of the passive form of the verb in *The letter was written/burnt/mailed by John*. In order to express this unity of allomorphy, it is useful to represent these forms as {WRITE, EN}, {BURN, EN} and {MAIL, EN}. These representations will be distinct from the past tense forms *wrote*, *burnt*, and *mailed* which will be represented as {WRITE, ED}, {BURN, ED}, and {MAIL, ED}. The mediation is schematized in (24):

(24)



The compositionality of words in terms of their feature composition is accepted within both affixal and nonaffixal approaches. I also assume that, within nonaffixal morphology, it would be legitimate to make statements such as: "The perfective forms *sung* and *mailed* consist of {SING, EN} and {MAIL, EN} respectively; {SING, EN} is realized as / sʌŋg /, while {MAIL, EN} is realized as /meil/ + / d /." Hence

the debate between the affixal and nonaffixal approaches is not located at what I have called the level of morphemes.

I also assume that, within nonaffixal morphology, statements such as the following would not be legitimate: “{SING, EN} is realized as / s^hɪŋg / + / Ø /, while {MAIL, EN} is realized as / meɪl / + / d /.” We must conclude, therefore, the debate is located at the level of formatives. If so, the interpretation of (22) is: “the only permissible morphological operation is that of combining formatives.” Under this interpretation, (22) forces the use of zero formatives in (25A). A theory that rejects (22) is free to have the representations in (25B):

(25)

	A		B	
	<i>spoke</i>	<i>cut</i>	<i>spoke</i>	<i>cut</i>
Feature	speak past	cut past	speak past	cut past
Morpheme	speak ed	cut ed	speak ed	cut ed
Formative	spiik ø	kʌt ø	spouk	kʌt

(25A), which incorporates (22), makes the claim that past tense forms like *spoke* and *cut* have two formatives each. In contrast, (25B), which rejects (22), views these forms as noncompositional at the level of formatives, though they are compositional at the level of morphemes and features.

If we accept (25A), the vowel alternation in *speak/spoke* will be formulated as a phonological principle sensitive to morphology (26a). If we accept (25B), it will be formulated as an allomorphic rule that states a relation between morphemes and formatives (26b):

(26)

- (a) *ii* is changed to *ou* in a verb formative when *ii* followed by a zero formative that marks {ED}.
- (b) *ii* is changed to *ou* as a marker of {ED}.

In the preceding section, I showed that phonological theory must allow principles of phonological organization to access morphological information. However, that some of the morphologically conditioned patterns of phonological distribution and alternation must be stated as principles of phonological organization does not entail that all morphologically conditioned patterns of speech must be viewed as phonological. It is true that (21a–d) must be viewed as principles of phonological organization, but nothing that I have said in support of the conception common to (15) and (18) forces a choice between (26a) and (26b). Within the theory of lexical phonology, for example, Kiparsky (1982) chooses the strategy illustrated in (26a), assuming affixal morphology, while K. P. Mohanan (1982) chooses (26b), rejecting affixal morphology.

In sum, I take it that there is persuasive evidence to show that a subset of phonological patterns of distribution and alternation are conditioned by the morphological structure of words. I also take it that we must separate morphophonology from the rest of phonology. The two options available for this purpose are to distinguish between (1) two types of principles, or (2) two domains for the application of principles. Lexical phonology makes the latter choice, as do most current theories of phonology (e.g., lexical prosodic phonology, nonlexical prosodic phonology). Whether or not all alternations of pronunciation should be dealt with in phonology, or whether we should allow morphology to take care of those alternations which are governed solely by morphological

information, is an issue that is not settled yet.

4 Construction Types

In the late seventies and early eighties, a number of studies argued that the use of boundary symbols is both too rich and too impoverished as a formal device for the representation of phonologically relevant morphological construction types (Rotenberg 1978; Pesetsky 1979; Selkirk 1980a; Strauss 1982; K. P. Mohanan 1982). As an alternative to the use of boundary symbols, the critics proposed two broad strategies. One was to encode the construction type in terms of labeled trees, as illustrated in (3). This was essentially the move made in Rotenberg (1978), Selkirk (1980), Strauss (1982), Aronoff and Sridhar (1983), and Sproat (1985). The other was to encode the information in terms of sequential modularity, following the path of (6) and (15). This was the move in Siegel (1974), Pesetsky (1979), Kiparsky (1982), K. P. Mohanan (1982), and Pulleyblank (1986a). Various combinations of the two approaches are found in Rubach (1984), Halle and Vergnaud (1987), and Inkelas (1989).

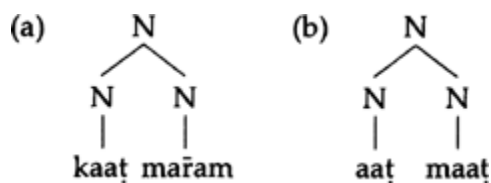
Let me illustrate these two approaches with the treatment of gemination in Malayalam compounds. In addition to the familiar type of compound in which one of the two stems is a head and the other is a modifier (subcompounds), Malayalam also has a compound construction in which a number of stems can be strung together, each of which is a head (co-compound). These two constructions are illustrated in (27b) and (27d) respectively.

(27)

(a)	(i)	kaaṭə	"forest"	(ii)	maṛam	"tree"	
(b)		kaaṭṭəmaṛam	"forest tree"				(subcompound)
(c)	(i)	aaṭə	"goat"	(ii)	maāṭə	"cow"	
(d)		aaṭəmaāṭəkalaṭə	"goats and cows"				(co-compound)

The stem-final / t / in (27b) undergoes gemination, but does not in (27d).¹⁵ The generalization illustrated here is that stem-final gemination holds in subcompounds, but not in co-compounds. Now, the phrase structure for (27b) and (27d) look the same:

(28)



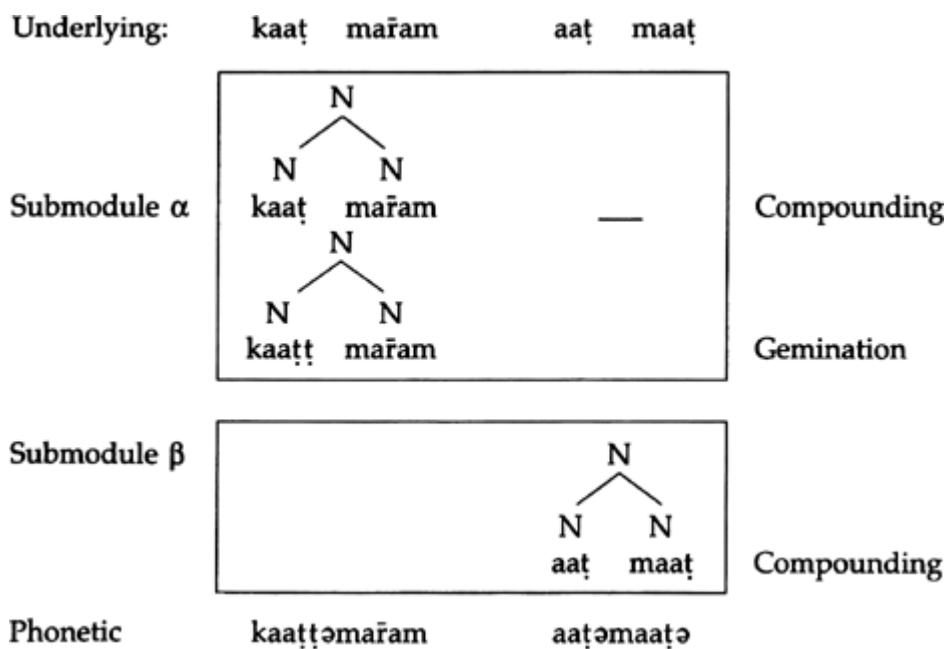
In order to account for the phonological distinction between the two types of compounds, therefore, one needs some other device. The strategy employed in lexical phonology is to assign the two types of compounding to two different lexical submodules, and define the domain of stem-final gemination in terms of these submodules:

(29)

- (a) The lexical module contains submodules α and β .
- (b) (i) Two nouns A and B can be put together to make a compound noun in which the head is the second one.
Domain: lexical submodule α .
- (ii) Two or more nouns A, B, ... can be put together to make a compound which means A and B and ...
Domain: lexical submodule β .
- (c) At the end of a form, before another, an obstruent is geminated.
Domain: submodule α .

Within the derivational conception of lexical phonology, the statements in (29a-c) yield the following results:

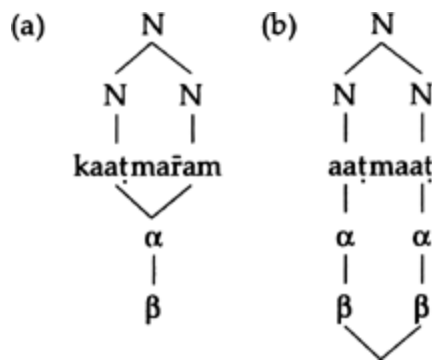
(30)



The conception underlying the derivation in (30) is that of building larger units from smaller units, allowing rules / constraints to take effect in the course of the building. Readers who are familiar with various theories in syntax will see that this is essentially the same as the conception in Montague grammar, and the device of generalized transformations in minimalist theory (Chomsky 1992). In addition, this part-whole relation is also viewed as involving a sequential ordering of modules.

The statements in (29) can be implemented in a representational approach, instead of a sequential approach, by making distinctions in representations in terms of appropriately labeled nodes and defining the domains of phonological principles in terms of these nodes. Suppose we use the representation in (31) instead of (28):

(31)



The statement of gemination can now be as given in (32):

- (32) (a) There are two lexical constructions α and β .
 (b) (i) Two nouns A and B can be put together to make an α .
 Semantic interpretation: the second constituent is the head.
 (ii) Two or more nouns A, B, ... can be put together to make a β .
 Semantic interpretation: A and B and ...
 (c) At the end of a form, before another, an obstruent is geminated.
 Domain: α .

The strategy in (32) is used in Selkirk (1980), Aronoff and Sridhar (1983), Rubach (1984), Sproat (1985), and Inkelas (1989). Sproat, for example, refers to the α in (32) as phonological stem (P-STEM) and the β as phonological word (P-WORD). Inkelas employs the labels α and β as diacritics.

It must be observed that there are two fundamental differences between the approaches in Sproat's prosodic nonlexical phonology and Inkelas's prosodic lexical phonology. In the former, a phonological rule can have access to the information in either the syntactic tree on top or the prosodic tree at the bottom in (31). As far as phonology is concerned, in other words, the two sets of structures are "co-present." In the latter approach, only the prosodic tree is visible for phonological rules, not the syntactic tree. This restriction is an extension of the indirect mapping hypothesis in post-lexical prosodic phonology (Selkirk 1984; Nespore and Vogel 1986). At the conceptual level, this is like *SPE*, where phonological rules can access only the readjusted syntactic structure: (2b) is visible to phonology, not (2a).

Another difference between the two approaches is that while the former relies purely on syntactic and prosodic representations, the latter relies on prosodic representations and sequentiality. Thus, the analyses in prosodic lexical phonology crucially depend upon the successive building and erasing of prosodic trees in different lexical strata, with phonological rules applying cyclically between building and erasing.

5 Distributional Asymmetries of Construction Types

This section deals with a phenomenon in morphology rather than phonology. However, it has a place in our discussion because of its interaction with phonology. Both sequential modularity ((30)) and labeled trees ((31)) have been employed in dealing with a problem that one may call the *distributional asymmetry* of morphological constructions. Take, for example, the relative ordering of the causative and passive affixes in Japanese and Malayalam. Japanese allows both V-CAUSE-PASS and V-PASS-CAUSE, with corresponding differences in their syntactic behaviour. Malayalam allows only V-CAUSE-PASS, blocking a potential syntactic option.

The asymmetry in the relative ordering of the causative and passive in Malayalam applies to individual morphemes. What is of interest to us are the asymmetries in morphological construction types, rather than in individual morphemes. The most frequently cited example of such an asymmetry of construction types in morphology is that of class 1 and class 2 affixes in English. It is well known that

the affixation of class 1 suffixes such as *-ity*, *-ion*, *-ic*, *-al*, *-ous*, *-ee*, *-aire*, *-ify*, *-ate*, and *-ion* can change the stress of the stem, while the affixation of class 2 suffixes such as *-nes*, *-less*, *-ful*, *-dom*, *-ed*, and *-ing* cannot do so. In terms of sequential modularity, this contrast can be expressed by assuming that stress assignment takes effect in the module in which class 1 affixes are attached. Now, it was pointed out by Siegel (1974) that while class 2 affixes can be attached to stems containing either class 1 affixes or class 2 affixes, and class 1 affixes can be attached to stems containing class 2 affixes, class 1 affixes cannot be attached to stems containing class 2 affixes. For example, the class 1 suffix *-ity* and the class 2 suffix *-ness*, can be productively attached to adjectives to derive nouns. However, only the latter can be attached to an adjective containing a class 2 affix, not the former: *paylessness* but **paylessnessity*. Siegel's Level Ordering Hypothesis accounts for this distributional asymmetry in morphology, and its correlation with the facts of stress, by proposing the following organization in terms of sequential modularity:

(33)

- (a) Class 1 affixes are attached at module α .
- (b) Class 2 affixes are attached at module β .
- (c) Module 2 precedes module β .
- (d) Domain for stress assignment: module α .

What is crucial for the account of the asymmetry is the relative ordering of the two modules in (33c). If, instead of sequential modularity, we resort to representational modularity, we can capture the morphological asymmetry using (34c):

(34)

- (a) Class 1 affixes are attached to α , and they yield α .
- (b) Class 2 affixes are attached to β , and they yield β .
- (c) A β contains α , but an α may not contain β .
- (d) Domain for stress assignment: α .

In various analyses in lexical phonology, the α and β of (34) are referred to as level / stratum 1 and level / stratum 2 respectively. Selkirk refers to the α and β of (34) as root and stem respectively, while Sproat calls them stem and word.¹⁶

Abstracting away the difference between sequential and parallel modularity, what is common to (33) and (34) are the claims that

- 1 there is a clustering of phonological patterns which motivate the distinction between two classes of affixes for the purposes of phonology;
- 2 there is a clustering of morphological asymmetries which motivate the distinction between two classes of affixes for the purposes of morphology;
- 3 there is a correspondence between the phonologically motivated classification and the morphologically motivated classification.

Now, the first claim is not without problems. There exist affixes like *-ist* which exhibit a fluctuating behavior with respect to phonological patterns (e.g., *columnist* can be either [kɔləmnɪst] or [kɔləmnɪst]). Furthermore, the convergence of phonological patterns on the dividing line between class 1 and class 2 affixes is not perfect (e.g., not all class 1 suffixes that begin with / i / trigger velar softening). Despite these problems, the claim about the need for two classes of affixes for

phonological purposes appears to be generally accepted.¹⁷

In contrast, various studies have pointed out far more serious problems with the morphological asymmetry claim. One is that of “bracketing paradoxes,” which give evidence for assuming that a class 1 suffix is attached to a stem that contains a class 2 prefix, as in *ungrammaticality* (Strauss 1982; Kiparsky 1982c; Sproat 1985).¹⁸ A second is the existence of morphological restrictions which need to be stated directly in terms of individual affixes. If there is a mechanism to state such restrictions, we do not need an additional mechanism of affix classification (Fabb 1988). A third problem is the existence of instances such as *governmental* where what must be regarded as a class 2 suffix is followed by a class 1 suffix (Goldsmith 1990).

One may take these three problems facing the morphological asymmetry claim as counterevidence that requires us to abandon the claim. Alternatively, one may treat them as serious problems of detail that require a solution, while retaining the asymmetry claim as capturing a pattern that is still broadly correct.

6 The Nonconvergence of Morphology and Phonology

The discussion in the previous section suggests that convergences in phonology, in morphology, and across phonology and morphology, are hardly perfect. In this section, I discuss some more significant instances of nonconvergence. These fall into two types. First, a distinction required by morphological asymmetry is not required by phonology. Second, a distinction required by phonology is not required by morphological asymmetry.

6.1 Multiple Domains and the Loop

6.1.1 The Phenomena

As an instance of the first problem of nonconvergence, consider the need to allow the same phonological rule to hold both within the morphophonemic module and the purely phonological module, corresponding to the lexical module and post-lexical module in lexical phonology. An example of this pattern is that of the place assimilation of nasals in Malayalam (16a) pointed out in section 2.3 (16a) needs to be part of the morphophonemic-phonemic mapping because its effect is relevant for the morphophonemic alternation between voiced plosives and nasals (16b), but it also holds across words. We may represent the situation diagrammatically as follows:

(35) Nonconvergence type 1: Phonological Patterns with Multiple domains

morphology-syntax	phonology
construction A	construction A
construction B	construction B

The phenomenon analyzed in K. P. Mohanan (1982, 1986) in terms of the loop presents the converse situation. In order to clarify the nature of what we are dealing with, I would like to separate the phenomenon of the loop from the formal device of the loop. The phenomenon is best illustrated by the morphological and phonological facts of subcompounding and co-compounding in Malayalam. As we saw in section 4, subcompounds and co-compounds behave differently for the purposes of stem-final gemination, which applies at the junction between two stems of a subcompound, but not those of a co-compound (27). This distinction also holds for stem-initial gemination:

(36)

- (a) (i) *meeṣa* “table” (ii) *petṭi* “box” (iii) *-kaḷ* “plural suffix”
 (b) *meeṣappettikaḷə* “table-boxes” (= boxes made out of tables) (subcompound)
 (c) *meeṣapettikaḷə* “tables and boxes” (co-compound)

Yet another phonological difference between the two constructions is that each stem in a co-compound receives its own primary stress and word melody, but the subcompound receives a single primary stress and word melody (the primary stressed syllables are underlined):¹⁹

(37)

- | | | | |
|---------|----------------------------------|------|---------------------------------|
| | Subcompound | | Co-compound |
| (a) (i) | <u>kaatṭəmaṛam</u> (= (27b)) | (ii) | <u>aaṭəmaatəkaḷə</u> (= (27d)) |
| | “forest tree” | | “goats and cows” |
| (b) (i) | <u>meeṣappettikaḷə</u> (= (36b)) | (ii) | <u>meeṣapettikaḷə</u> (= (36c)) |
| | “table-boxes” | | “tables and boxes” |

It is clear that the two constructions are phonologically distinct. Yet, unlike what happens with class 1 and class 2 affixation in English, there is no morphological asymmetry between the two kinds of compounds: a co-compound can have a subcompound as one of its stems, and a subcompound can have a co-compound as one of its stems (K. P. Mohanan 1986). This situation can be diagrammatically represented as follows:

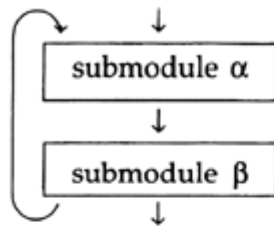
(38) Nonconvergence type 2: The Phenomenon of the Loop

morphology-syntax	phonology
construction A	construction A
construction B	construction B

The formal device of the loop, proposed in Mohanan (1982) as a way of capturing the loop phenomenon, is intended to make a distinction between the two constructions in terms of sequential modularity: one module precedes the other, and yet allows the output of the second module to be input to the first:

(39)

(a) Organization of the Lexicon



(b) Morphology

(i) subcompounding (Domain: submodule α)(ii) co-compounding (Domain: submodule β)

(c) Phonology

(i) Stem-initial gemination (Domain: submodule α)

When preceded by a vowel-final stem, an obstruent is geminated.

(ii) Stem-final gemination: (Domain: submodule α)When followed by a stem, t and r are geminated.(iii) Stress: (Domain: submodule α)

Primary stress on the second syllable if the second syllable has a long vowel; if not primary stress on the first.

(iv) Word melody: (Domain: submodule α)

The stressed syllable has a low tone; the final syllable has a high tone.

Given the looped organization in (39a), the domain specifications in terms of the two modules in (39b) make the right predictions for gemination, stress, and word melody. The derivation in (40) provides an illustration of how the four-stratum organization of the Malayalam lexicon works (Mohanani 1986).²⁰

(40)

Stratum 1		
[meeša], [pet̪ti], [ka]	[meeša], [pet̪ti], [ka]	
Stratum 2 (= α in (39))		
[meeša], [pet̪ti], [ka]	[meeša], [pet̪ti], [ka]	
[[meeša][pet̪ti]]	—	Compounding
[[meeša][ppet̪ti]]	—	Gemination
[meešappet̪ti]	[meeša], [pet̪ti]	Stress, melody
[meešappet̪ti]	—	BEC ²¹
Stratum 3 (= β in (39))		
[meešappet̪ti] [ka]	[meeša], [pet̪ti], [ka]	
—	[[meeša][pet̪ti]]	compounding
—	[meešappet̪ti]	BEC
Stratum 4		
[meešappet̪ti], [ka]	[meešappet̪ti], [ka]	
[[meešappet̪ti][ka]]	[[meešappet̪ti][ka]]	Affixation
[meešappet̪tika]	[meešappet̪tika]	BEC
...		
[meešappet̪tika]	[meešappet̪tika]	Phonetic

6.1.2 Evidence for the Phenomenon of the Loop

The four-stratum hypothesis for Malayalam forces the formal device of the loop. This formal device appears to have gained disapproval in many quarters, and the hypothesis itself has evoked a great deal of negative reaction in the phonology community (e.g., Sproat 1985; Christdas 1988; Inkelas 1989). It may therefore be useful to clarify what the analysis has successfully done, which its alternatives have failed to do. First, as pointed out in K.P. Mohanan (1986, pp. 120–2), it provides a solution for an ordering paradox (or its equivalent in non-rule-ordering frameworks). The paradox is that in subcompounds, the rules of vowel sandhi and vowel lengthening must apply prior to stress and melody assignment, while in co-compounds they have to apply after stress and word melody. The gist of the relevant phenomenon is as follows. When a vowel-final stem is followed by a vowel-initial stem in a subcompound or co-compound, the two vowels merge into a single long vowel (Vowel Sandhi). When a vowel ending stem is followed by a consonant initial stem in a subcompound or cocompound, the vowel is lengthened (Vowel Lengthening). As for stress, primary stress is assigned to the second syllable if the first syllable has a short vowel and the second syllable has a long one; if not, the primary stress is assigned to the first syllable. The primary stressed syllable gets a low tone as part of its word melody, while the last syllable gets a high tone. Now consider the following data which shows the interaction of vowel lengthening and stress:

(41)

- (a) (i) waḍ^hu “bride” (ii) grāham “house”
 (b) waḍ^huugrāham “wife’s house”

In the examples in (41a), the vowels in the second syllable are short, and hence primary stress falls on the first syllable in both words. In (41b), the vowel in the second syllable is long, and hence primary stress falls on the second syllable. We must therefore assume that stress assignment takes place after vowel lengthening.

Now consider the interaction between the three rules in co-compounds:

(42)

- (a) (i) waṛan "bridegroom" (ii) -maar (plural)
 (b) waḍ^huuwaṛanmaar "bride and bridegroom"

Since the first syllable in (42b) receives primary stress despite the second vowel being long, we must assume that stress is assigned before the vowel gets lengthened. The same paradox appears with respect to vowel sandhi. The four-stratum hypothesis explains these paradoxes by specifying the domain of vowel sandhi and vowel lengthening as strata 2 and 3, and ordering stress and word melody after vowel lengthening and vowel sandhi. The subcompounding in (41) takes place at stratum 2, where vowel lengthening precedes stress and word melody. The individual stems that are put together to form a co-compound are assigned stress and word melody in stratum 2, before they are compounded. As a result, co-compounding takes place after stress and word melody, and vowel lengthening takes place after compounding. As far as I know, these facts have not been counteranalyzed in the literature that rejects the four-stratum analysis of Malayalam (e.g., Sproat 1986; Inkelas 1989; Wiltshire 1992).

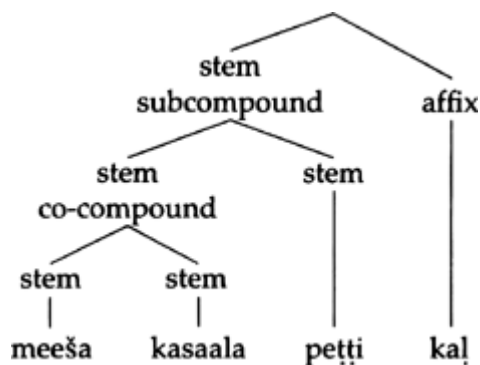
The second piece of evidence that favors (39) involves the combination of segmental and suprasegmental facts in the interaction between subcompounding and co-compounding. As pointed out above, one of the stems of a subcompound can itself be a co-compound, as illustrated in (43):

(43)

- (a) (i) meeṣa "table" (ii) kaṣaala "chair" (iii) peṭṭi "box"
 (b) meeṣakaṣaalappeṭṭikaḷ "boxes made from tables and chairs"
 (c) meeṣappeṭṭikkaṣaalakaḷ "chairs made from tables and boxes"

The structure of (43b) is given in (44) as an illustration:

(44)



As predicted by (39c), stem-initial gemination does not appear at the junction in the co-compound, but it does at the junction in the subcompound. What is interesting is that the second stem of the subcompound carries its own primary stress and word melody. This follows from the model in (39a) if we assume that the default principles of building suprasegmental structure (syllable, stress, word melody) do not disturb the prior assignment of structure, but assign it to the stretch that does not carry suprasegmental information (K. P. Mohanan 1986). In (43b), the substructure [[meeṣa][kasaala]] is assigned stress and word melody in its first pass through submodule 3. It then loops to submodule 2, to derive the structure [[meeṣakasaala][petṭi]], where stem-initial gemination takes place. In its second pass through submodule 3, the form [meeṣakasaalappetti] has stress and word melody in the

stretch *meešakasaala* but not in *peṭṭi*. Hence the latter stretch receives stress and word melody.

6.1.3 The Loop in the Representational Formalism

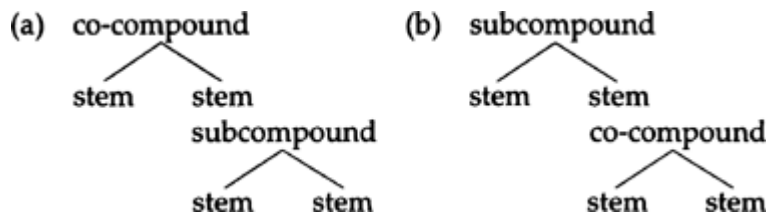
We are dealing with two distinct issues here. One is the need to recognize two kinds of nonconvergence between morphology and phonology: two constructions which need to be kept separate for morphological purposes may need to be unified for the purposes of phonology (nonconvergence type 1), and, conversely, two constructions which need to be kept separate for phonological purposes may need to be unified for the purposes of morphology (nonconvergence type 2). The other is the issue of the formal device that permits this nonconvergence. The loop is precisely such a device. I would like to point out that this device is equivalent to a recursive rule in a phrase structure grammar. Consider the rules in (45):

(45)

- (a) $\beta \rightarrow \beta \beta$
- (b) $\alpha \rightarrow \alpha \alpha$
- (c) $\beta \rightarrow \alpha$
- (d) $\alpha \rightarrow \beta \beta$

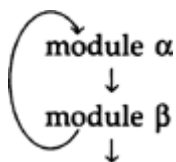
If we interpret α and β as subcompounds and co-compounds, these rules correctly generate the recursive structures illustrated in (46):

(46)



The use of representations that resemble (46) has been proposed as an alternative to the formal device of the loop. Note that rule (45d) is the representational counterpart of the loop in sequential modularity.

(47)

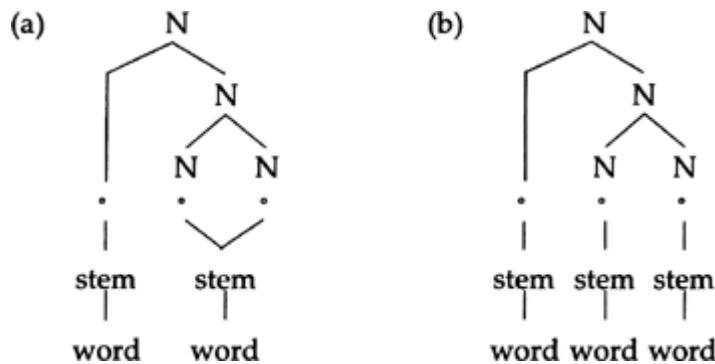


I fail to see how the looping in (47) is more unconstrained than the recursion in (45d). As pointed out earlier, no analysis can escape from the phenomenon of the loop in (38). As for the formal devices, whether the phenomenon of the loop should be captured in terms of representations as in (45d) or in terms of sequential modularity as in (47) is a secondary issue of detail.

Another representational alternative to (39) is that of prosodic phonology in Sproat (1985) and Inkelas (1989). In this approach, prosodic nodes carry the information necessary to distinguish between the phonology of subcompounding and co-compounding. Thus, the structures that correspond to (46a)

and (46b) are represented as (48a) and (48b) respectively:

(48)



The compound in (43b) will have the structure in (48b). The domain of stress and word melody in the prosodic approach are specified in terms of the domain (P-)WORD. This correctly predicts that (46a) will have two primary stresses and two word melodies, while (46b) will have three each. However, the difficulty arises when it comes to principles such as gemination, which hold between the stems of a subcompound but not a co-compound. The representations in (48) incorrectly predict that there cannot be any gemination between the two stems of the subcompound in (46b). In addition, they also fail to explain the phenomenon that we identified earlier as an ordering paradox between the principles of stress and word melody on the one hand, and vowel lengthening and vowel sandhi on the other. In the light of these observations, we must conclude that while the representations in (46a, b) are legitimate as specifications of the domains of the principles of phonological organization, those in (48a, b) are not. These phenomena pose precisely the same challenges to the alternative analysis of the Malayalam facts in terms of harmonic phonology in Wiltshire (1992), though problems of space do not permit me to go into the details.

Recall that the problem we have been dealing with is that of the non-convergence of morphologically motivated and phonologically motivated construction types, where a construction type would mean such things as class 1 and class 2 affixation in English, or subcompounding and co-compounding in Malayalam. The phenomena that we have examined so far can be correctly described in terms of the sequential modularity of lexical phonology, but not in terms of the proposed representational alternatives in non-lexical phonology or prosodic lexical phonology. Needless to say, that the alternatives to sequential modularity have not been successful in accounting for these phenomena should not be taken to mean that there are no nonsequential accounts. I am hopeful that such an account can be found, but I have no intelligent suggestions to make at present. As we shall see in the following section, the apparent advantage of lexical phonology in accounting for the rule ordering paradox and the facts of gemination in subcompounds containing co-compounds is challenged by the picture that emerges when we consider the facts of verbal compounding in Malayalam.

6.2 A More Complex Nonconvergence: Verbal Compounding in Malayalam

The crucial phenomena relevant for the analysis of subcompounds and co-compounds in Malayalam are stress, word melody, stem-initial gemination, stem-final gemination, nasal deletion, onset fusion, sonorant degemination, vowel sandhi and vowel lengthening. We have already seen how subcompounds and co-compounds exhibit differences in stress, word melody, and the two types of gemination. T. Mohanan (1989) shows that they are different with respect to sonorant degemination and onset fusion. Despite these differences, the two types of compounds are also similar in that they both undergo nasal deletion, vowel sandhi, and vowel lengthening (K. P. Mohanan 1986).

In addition to these two types of compounds, Malayalam also has verbal compounds, as illustrated in (49c).

(49)

- (a) (i) maṛam "tree" (ii) kuṭīra "horse" (iii) kayar- "climb"
 (b) maṛakkuṭīra "wooden horse" (subcompound)
 (c) maṛaṅkayari "tree climber" (verbal compound)

In a subcompound, the first stem is a modifier (= adjunct) of the second stem; in a verbal compound, the first stem is a complement of the second stem. In a co-compound, every stem is a head.

As far as stress and word melody are concerned, verbal compounds and subcompound are identical. Both types of constructions carry a single primary stress and word melody. On the basis of this behavior, K. P. Mohanan (1986) incorrectly concluded that subcompounds and verbal compounds are concatenated in the same lexical stratum, namely, stratum 2. In an interesting unpublished paper, Yatabe (1991) shows that verbal compounds behave differently from subcompounds with respect to gemination, onset fusion, and sonorant degemination. They are different from both subcompounds and co-compounds with respect to nasal deletion, vowel sandhi, and vowel lengthening. The example in (49c) illustrates that verbal compounds do not undergo nasal deletion, unlike subcompounds. The following examples illustrate the absence of gemination in verbal compounds:²²

(50)

- (a) (i) kaatə "forest" (ii) aana "elephant" (iii) ootum "will run"
 (b) kaatṭaana "wild elephant"
 (c) kaatooti "jungle roamer"

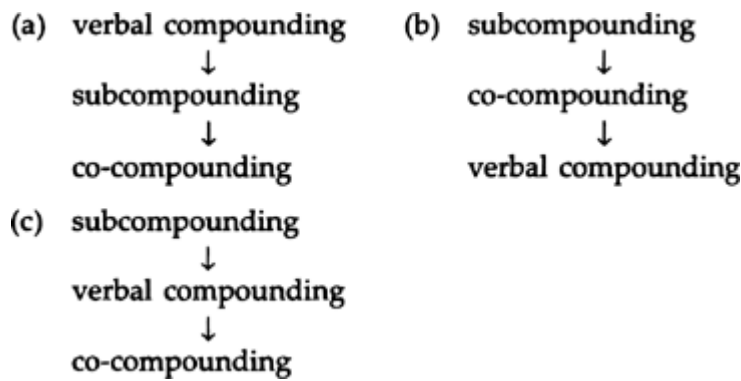
The phonologically motivated classifications of the three types of constructions are given in (51):

(51)

Phenomena	Classification
(a) Stress; word melody	Verbal compounds and subcompounds Co-compounds
(b) Nasal deletion; vowel sandhi, vowel lengthening	Verbal compounds Subcompounds and co-compounds
(c) Gemination; onset fusion, sonorant degemination	Verbal compounds and co-compounds Subcompounds

Given this situation, one may think of assigning verbal compounding to an independent stratum, and ordering it before subcompounding as in (52a), after co-compounding as in (52b), or in between, as in (52c):

(52)



There are two problems with this solution. First, since the three types of compounds are inputs to one another (i.e., any one of them can contain the others), the organization, whether (52a), (52b), or (52c), will have to allow loops that connect nonadjacent strata, a situation that has so far been disallowed in lexical phonology. Second, even if we loop nonadjacent strata, none of these solutions will yield an analysis that correctly captures the situation in (51). Verbal compounds carry a single primary stress and word melody: the solution in (52b) will incorrectly predict that each stem in a verbal compound will have its own primary stress and word melody. Since onset fusion and sonorant degemination do not hold in verbal compounds, the ordering in (52a) will yield incorrect results. The solution in (52c) will not be able to account for nasal deletion, vowel sandhi, and vowel lengthening, which apply in subcompounds and co-compounds, but not in verbal compounds. Thus, no possible ordering of the modules will yield the correct predictions. This dilemma is equally true for the alternative strategies in Sproat, Inkelas, and Wiltshire. It appears, then, that none of the theories discussed above, lexical phonology, prosodic lexical phonology, non-lexical prosodic phonology, and harmonic phonology, entirely succeeds in describing the facts of the three types of compounding.

In an attempt to provide an account, Yatabe suggests that phonological principles have direct access to the morphosyntactic construction types. Details of formalism aside, Yatabe's analysis requires that we recognise the structure of the three types of compound as follows:

(53)

- (a) Verbal compound: complement, head
- (b) Subcompound: modifier, head
- (c) Co-compound: head, head,...

The domains of phonological principles are stated directly in terms of these construction types, rather than in terms of sequential submodules, prosodic nodes, or a combination.

In addition to being descriptively adequate, Yatabe's analysis has the advantage that the morphological constructs employed in the specification of phonological domains ((53a-c)) are imbued with universal substance. For example, on the basis of parallel contrasts between subcompounds and co-compounds in Japanese and English, yatabe claims that there is a general universal tendency for verbal compounds to undergo fewer phonological principles than subcompounds. Unlike the previous proposals in the phonology morphology interaction, the proposal in (53) lends itself to the pursuit of universal statements of this kind.

Most proposals for the interaction between morphology and phonology since *SPE* have suffered from the weakness that they employ diacritic devices to refer to construction types. Thus, *SPE* employs diacritics like + and #, while lexical phonology employs diacritics like stratum 1 and stratum 2. Proposals for constructs like phonological stem (versus phonological word) have also remained nothing more than disguised diacritical markings, in spite of the sense of security induced by the

terminology. The diacritic status of prosodic categories is transparent in prosodic lexical phonology, where the terminology of α and β replaces that of P-STEMS and P-WORDS. I would like to submit that the distrust of the four-stratum analysis of Malayalam and English has its basis in the intuitive distrust of the unconstrained use of diacritics that carry no universal substance.

As I see it, what is valuable about Sproat's alternative is the use of phonological word as a construct for the statement of the domains of phonological principles. His hypothesis that each conjunct in a coordinate construction must contain at least one phonological word is a way of imbuing the notion of phonological word with substance. As Sproat himself acknowledges, this substance is absent in the proposal to use phonological stem on a par with phonological word. Yatabe escapes the use of diacritics by referring directly to construction types involving different relations of head, modifier, and complement ((53)). We observed earlier that phonological principles need to refer to the nondiacritic constructs morpheme and word. Let us hope that other phonologically motivated differences in construction types, such as that between class 1 and class 2 affixation in English, will also prove to be universally grounded.

It must be pointed out that if we accept yatabe's proposal, we are forced to acknowledge that word-internal phonology must refer directly to word-internal syntax, contra the claims made in prosodic lexical phonology. We have already seen that phonological principles must have access to the unit morpheme. If the crucial contrast between verbal compounds and sub-compounds is the syntactic relation of modifier versus complement, and the effects of this relation are repeatedly found across languages, then lexicalprosodic approaches to the morphology-phonology interface cannot provide a satisfactory account of the contrast. Whether or not what is true of the relation between word-internal syntax and phonology is also true of phrasal syntax and phonology remains to be seen. If it is, it would challenge the claims of prosodic phonology as well.

To summarize, the use of prosodic representations has not yet been able to provide successful analyses of certain phenomena (rule ordering paradoxes, interaction between two types of compounds) which do have an account within the sequential modularity of lexical phonology. This is not to recommend lexical phonology, but simply to point out the problems which need to be solved in future work. Evidence also seems to suggest that phonological principles need to have access to the following kinds of information on morphological construction types: morpheme, word, affix versus stem, two types of affixes, head, modifier, and complement. The question that arises is: if we develop a formalism that allows such direct reference, independently of sequential modularity or prosodic trees, do we still need sequential modularity or prosodic trees? I leave this question open.

7 Lexical Phonology, Rule Systems, and Representations

The topic of this chapter is the organization of the grammar. It is not the formal properties of the rules/constraints that capture the patterns of distribution and alternation in phonology, or the formal representation of phonological information. However, a number of claims about the behavior of rule-constraint systems and representations crucially depend upon the modular organization of grammar. Therefore, let me now briefly review some of these claims.

7.1 Rule-Constraint Systems

The division of grammar in lexical phonology into lexical and post-lexical modules has given rise to the following claims about the application of phonological rules, most of which have been challenged:

(54)

- (a) Only lexical rule applications may refer to word-internal structure.
- (b) Only post-lexical rule applications may apply across words.
- (c) Only lexical rule applications may be cyclic.
- (d) Lexical rule applications are structure preserving.
- (e) Only lexical rule applications may have lexical exceptions.
- (f) Lexical rule applications must precede post-lexical rule applications.

Claims (54a) and (54b) are in a sense criterial for deciding whether a rule application is lexical or not. However, given the general consensus that noncategorical gradient phenomena are a characteristic of phonetic implementation, not lexical rule application, we can derive from (54a) the prediction that morphologically sensitive patterns cannot be gradient. Sproat and Fujumura (1993) and Sproat (1993) argue that this prediction is incorrect, since both the gradient rhyme lengthening and the velarization of /l/ in English distinguish between the concatenation of words in a phrase and the concatenation of stems within a compound word, and both of these from other junctures and absence of junctures. As far as I can see, these facts can equally well be accounted for by appealing, in the post-lexical module, to the distinctions between (1) single words and phrases, and (2) phonological words as units composed of phonological feet. Both these distinctions are available in the post-lexical module. Hence, even though I acknowledge the possibility of (54a) turning out to be false, I do not see any persuasive evidence for abandoning it at the present.

The current status of (54c) is unclear, given that research within lexical phonology itself has challenged it. For example, Rubach (1984), Booij and Rubach (1984), and Kiparsky (1985) assume that the last stratum in the lexicon is or can be noncyclic; Halle and Mohanan (1985) assume that each stratum, including the first, can be specified as being cyclic or noncyclic. As pointed out in Kaisse and Shaw (1985), there are also examples of cyclic application in post-lexical phonology. Liu (1980) argues that the rule in Mandarin that changes tone 3 to tone 2 when followed by a 3 must apply cyclically across words. Drescher (1983) makes a similar argument for cyclic post-lexical rules. I refer the reader to [chapter 3](#), this volume, for a detailed discussion of cyclicity.

Often tied up with cyclicity is the issue of the Strict Cycle Condition (SCC). As Cole ([chapter 3](#), this volume) points out, SCC has two parts. One is the formal need to prevent rules from “reaching” back in order to preserve rule ordering in cyclic domains. I will not deal with the problem of reaching back in this chapter, because it strikes me as a purely formal issue that arises out of certain kinds of formal mechanics. I agree with Cole's judgement that there is no empirical support for this constraint. The other part of SCC is the Derived Environment Condition (DEC), which asserts (1) that certain kinds of rules apply only in derived environments, where “derived” means derived either through morphemic composition or through the application of a phonological rule, and (2) that the domains which exhibit DEC are cyclic domains. The classic example of this effect is the nonapplication of trisyllabic shortening in monomorphemic forms like *nightingale* in English. We have various versions of DEC, the weakest of them being that a structure-changing rule cannot apply in underived environments in nonfinal strata in the lexicon.²³ As pointed out in K.P. Mohanan (1986), even this weak formulation is untenable as shown by the facts of nasal spreading in Malayalam (principle (16b)). This rule applies to monomorphemic forms in stratum 1, changing underlying forms like /pandi/ to [paⁿⁿi] “pig”. Since the place-assimilation rule that applies before nasal spreading applies only if the trigger is [-son, -cont], the post-nasal voiced plosive that becomes nasal must be underlyingly specified as [-son, -cont].²⁴ Since the specification changes to [+son] as a result of nasal spreading, we must take it as a structure-changing rule. Thus, nasal spreading is a structure-changing rule that applies to underived forms in stratum 1.

It appears to me that the various attempts at formulating a principle of DEC have run into problems because we are pursuing a nonissue. If we decide to formulate trisyllabic shortening as a phonological rule, examples like *nightingale* can simply be treated as lexical exceptions. Independently of DEC effects, we must stipulate examples like *obese* and *scene* as lexical exceptions to shortening, as the derived forms *obesity* and *scenic* do not undergo the rule. Observe that if the rule did apply to *nightingale*, deriving the surface form [nitingeɪl], there would be no evidence for postulating an underlying long vowel in the first syllable.

Let us consider the exceptions to obstruent voicing assimilation in English. Even though voicing agreement is a robust phenomenon in English, there are lexical exceptions to it in both derived environments (e.g., *midst* [midst]) and underived environments (e.g., *svelt* [svelt]). What special advantage do we gain by claiming that the nonapplication of voicing agreement in underived environments is due to a special principle, and that even the post-lexical application of voicing agreement is subject to this condition, rather than saying that the rule/constraint has lexical

exceptions?

How can anyone produce counterevidence to the claim that trisyllabic shortening and voicing agreement cannot apply in underived environments? Suppose there is a dialect of English in which *svelt* is pronounced as [sfelt], and *nightingale* is pronounced as [nitingeil]. Would this constitute evidence for assuming that the underlying representations of the two words are /svelt/ and /naitingeil/, and that these underived forms undergo voicing agreement and shortening? Certainly not. If such phonetic forms were indeed observed, they would be analyzed as having the underlying forms /sfelt/ and /nitingeil/, as there is no evidence to assume a more abstract underlying form. Thus, the absence of examples in which trisyllabic shortening applies in underived environments is simply the result of Occam's Razor which requires us not to postulate underlying representations that are more abstract than necessary. The illusion of the need for a linguistic principle of DEC, then, is the result of the combination of (1) lexical exceptionality, and (2) patterns of alternation that are not motivated by distribution.

Let us turn to (54d). At the conceptual level, structure preservation is the *phenomenon* of the preservation of distributional regularities in the patterns of alternation. As pointed out in Goldsmith (1990) and K. P. Mohanan (1991), distributional regularities are repeatedly found to inhibit an otherwise regular alternation (i.e., block the application of rule) or motivate an alternation or a particular part of it (i.e., trigger rules, or repair the output of rules). The traditional formulation of the Structure Preservation Constraint (SPC) elevates the blocking effect into a universal constraint, and turns out to be either false or empirically vacuous.

The intuition that SPC tries to express is that certain patterns of alternation, which we would expect otherwise to hold in certain forms, are blocked because they would violate some other independent principle of the grammar. For example, Kiparsky (1985) argues that place assimilation in Catalan allows the lexical spreading of backness to an underlying /n/ from a following velar plosive, but the similar spreading of dentality from a following dental plosive does not occur, since [+distributed] is disallowed in underlying representations, and hence no lexical rule can introduce it.

As pointed out by Spencer (1991), the idea (though not the terminology) of structure preservation in phonology–morphology was first proposed in Aronoff (1976, p. 98). Aronoff defines an allomorphic rule as one which “effects a phonological change, but which only applies to certain morphemes in the immediate environment of certain other morphemes ...” Such rules “cannot introduce segments which are not otherwise motivated as underlying phonological segments of the language.” Kiparsky (1985) says:

“In English..., voicing is distinctive for obstruents but not for sonorants. We express this by a marking condition which prohibits voicing from being marked on sonorants in the lexicon:

(16)

$$* \left[\begin{array}{l} \alpha \text{ voiced} \\ + \text{son} \end{array} \right]$$

A language in which voicing is entirely non–distinctive would have the marking condition

(17) * $[\alpha \text{ voiced}]$

By STRUCTURE PRESERVATION I mean that marking conditions such as (16), (17) must be applicable not only to underived lexical representations but also to derived lexical representations, including the output of word–level rules.”

Kaisse and Shaw (1985) interpret this as a statement that prohibits phonological rules in the lexicon from creating segments that are not part of the underlying inventory.²⁵ The analysis of the seven

points of articulation among the nasals in Malayalam in terms of three underlying nasals (Mohan and Mohan 1984) is a clear counterexample to this claim. Take, for example, the complementary distribution of alveolar and dental nasals in the language. The dental nasal occurs only (1) at the beginning of a morpheme, (2) when followed by a dental stop, or (3) when adjacent to another dental nasal. In these environments the alveolar nasal cannot occur. These facts follow from the analysis that assumes that dental nasals are not part of the underlying inventory. The dental geminate is derived through assimilation from / nd/ through (16a) and (16b); the morpheme-initial dental nasal is derived from / n / by a rule that makes alveolar nasals dental at the beginning of a morpheme ((21d)). Both these rules create the structure [+nasal, +coronal, +anterior, +distributed], in spite of the fact that [distributed] is not a feature that is contrastive underlyingly in the language. For similar examples of the violation in German and Welsh, see Hall (1989) and Sproat (1985).

Thus, as a universal principle of interaction between rules and constraints, SPC need not be obeyed even in the first stratum of a grammar. Conversely, we see the effects of structure preservation in the post-lexical module as well (K. P. Mohan 1991). What this means is that structure preservation is not related to the division of the grammar into the lexical and post-lexical modules. It is not a principle of the grammar, but a recurrent phenomenon of the preservation of distributional patterns in alternation (T. Mohan 1989; Goldsmith 1990; K. P. Mohan 1991).

Perhaps it would be instructive to ask what the intuition of Structure Preservation would correspond to in a grammar that does not subscribe to the sequential conception of rule applications and levels of representation. Suppose a grammar has three parallel levels of representation as in (6). In such a grammar, principles of phonological organization can hold within any of the three levels, or across two or more levels. In such a situation, what is analyzed as SPC will be the effect of a principle that holds within a level winning over a principle that holds within another level, or across two levels. If a condition holds on both the morpheme and word levels, no pattern of alternation that is valid between the morpheme and word levels would violate it. This would correspond to the attested effects of SPC within the lexical module. On the other hand, if a condition holds on the morpheme level, but not on the word level, we will see violations of SPC such as those observed in Malayalam.

I would like to suggest that the only universal claim that we can make about the appearance of nonunderlying contrasts is the one that Aronoff made about allomorphic patterns. Suppose we define an allomorphic alternation as one that is conditioned *purely* by the properties of particular morphemes, i.e., an alternation that is not triggered by phonological content or morphophonological structure. By this definition, the voicing alternation in the English pair *believe/bereft* will not be allomorphic because its trigger includes phonological content (the voiceless /t/), and the morphologically conditioned [n]/[ŋ] alternation in Malayalam mentioned above will not be allomorphic because it is conditioned by morphological structure, not by individual morphemes. In contrast, the vowel alternation in *sit/sat* will be an example of allomorphy. Such an allomorphic alternation will be a relation between two different formatives of a morpheme in the sense discussed in section 3.2. If so, an allomorphic alternation will be expressed as a relation between two different underlying forms, while a phonological alternation will be expressed as a relation between an underlying form and its phonetic form. If a segment is not part of the underlying inventory, it cannot appear in either of the two underlying forms related through an allomorphic pattern. Aronoff's generalization follows from this view of allomorphy. For example, we predict that there will be no language in which a morphological feature such as a plural or past tense is phonologically realized solely as, say, a front rounded vowel, or a nasalized vowel, where front rounded vowels and nasalized vowels are not underlying segments.

Finally, we turn to the claim of post-lexical rule applications not having lexical exceptions (54e). This claim is challenged by examples like the post-lexical voicing assimilation in English which has exceptions like *midst* [midst] and *svelt* [svelt]. Even a pervasive phenomenon such as the place assimilation of nasals in Malayalam, which applies both lexically and post-lexically, fails to apply post-lexically in the form [anpə] "kindness" (K. P. Mohan 1986). Therefore, what (54e) states is not an absolute condition, but a recurrent tendency for lexical patterns to have more lexical exceptions than post-lexical patterns.

In sum, we conclude that most of the claims in (54) on the interaction between rule-constraint systems and modules of the grammar are questionable. Particularly dubious are the claims of cyclic

application of rules, strict cyclicity, structure preservation, and lexical exceptionality. I have suggested that strict cyclicity is an illusion created by the lexical exceptionality of the patterns of phonological alternation that are not motivated by distribution, that structure preservation is simply an observed effect of the conflict between two contradictory principles of phonology, and that the claim of post-lexical rules having no exceptions as an absolute universal is simply false.

7.2 Representations

In recent years, most discussions of the modular organization of phonology, particularly those in lexical phonology, have involved a representational issue, that of underspecification. The fundamental claim of underspecification theories is that it is desirable to omit certain kinds of information from underlying representations. One of the central questions that theories of underspecification have tried to grapple with has been the recurrent asymmetries of phonological content (Archangeli 1984; K. P. Mohanan 1991; and Steriade, [chapter 4](#), this volume): Why do certain feature values spread, while their complements do not? Why do segments with certain feature values undergo phonological change, while others resist the change? Why do certain feature values block spreading, while others are transparent? Is there a correlation among these different kinds of asymmetries?²⁶ Is there any correlation between the asymmetric effects of spreading, blocking, and resistance on the one hand, and contrastive and noncontrastive information on the other?

I have argued elsewhere that while the pursuit of these issues has led to a deepened understanding of phonological patterns in natural languages, the mechanics of underspecified representations, structure-building rules, and structure-changing rules have actually obscured the patterns rather than illuminated them (K. P. Mohanan 1991). For example, it has been argued that asymmetries in the content of features argued for in underspecification theories have also been found in language acquisition, processing, and phonetic implementation (e.g., Stemberger 1991, 1992; Keating 1988). This has been taken as evidence for the formalism of underspecification. However, that there is a systematic asymmetry between coronals and noncoronals in language acquisition, language production, and the like, does not necessarily show that the asymmetry is to be formally expressed as specified versus unspecified information.

As in other issues in phonology, I expect that the current trend toward abandoning the sequential input-output conception of phonological derivation will have a profound effect on underspecification theories. This issue is anticipated in recent works such as Archangeli and Pulleyblank (in press). However, the potential problems of combining underspecification with a nonsequential conception remain unexplored. The conclusions will emerge more clearly only when the specifics of nonsequential analyses have been fully articulated.

8 Summary

To summarize, the central problem that we have been concerned with in this chapter has been the relation between phonology and other aspects of the organization of language. In dealing with issues in this domain, we have found it useful to separate the issue of modularity from the issue of sequentiality. Thus, one can subscribe to the hypothesis that phonological theory needs to separate the module of word-internal structure from the module of structure across words, without necessarily assuming that the former module precedes the latter in a procedural sense. In a nonsequential conception, the modules and the levels of representation that are associated with them are "co-present," as structures along a multidimensional space, where information from different "levels" or dimensions of organization is simultaneously accessible to principles of the grammar.

Most recent approaches to syntax have converged on the need for different types of information being co-present. In the *Aspects*-type syntax (Chomsky 1965), for example, information about theta roles was available only at the level of deep structure. Information about grammatical functions was available only at a later stage in the derivation where information about theta roles was no longer available. *SPE* phonology inherited this conception of information distributed along various stages in a derivation, and subsequently passed it on to lexical phonology. With the advent of trace theory, the role of particular stages in a derivation became irrelevant, move alpha being simply a way of thinking about relating one level of representation to another. Furthermore, traces also allow all information to be simultaneously present. Thus, unlike the kinds of representations in generative semantics, a single

tree structure in the minimalist descendants of government-binding theory can contain information about theta roles, grammatical functions, quantifier scope, and topichood (Chomsky 1992).

This conception of co-present information is present in phonology in Halle and Vergnaud's (1980) metaphor of the ring-bound notebook. In addition to co-presence, this conception also opened up the possibility of distributing different types of information along different dimensions of structure. This move towards co-presence of information and multidimensionality of structure has led to the convergence of a number of proposals that abandon the input-output metaphor in generative grammar. These proposals state principles of the grammar not as rules, but as laws that state relations between entities within or across different co-present levels or dimensions of representation. In various stages of development and detail, this approach can be found in Chomsky (1981), Hale (1983), and K. P. Mohanan (1983), who suggest that lexical and configurational structures are a pair of parallel structures; Zubizarreta (1987), who makes a similar proposal for virtual and actual structures; Marantz (1984), who suggests that all levels of representation are "linked"; Sadock (1985, 1991) who proposes that morphological and syntactic structures are parallel-linked dimensions of structure; T. Mohanan (1990, in press) who proposes a multidimensional view of the representation that includes the dimensions of semantic structure, argument structure, grammatical function structure, and grammatical category structure.

This conception of simultaneous modularity conflicts with the earlier conception of sequential modularity in lexical phonology and prosodic lexical phonology. Those who subscribe to a multidimensional view of linguistic organization in syntax and phonology are therefore faced with the challenge of identifying the crucial insights expressed in terms of the mechanics of sequential modularity and sequential rule application, and exploring how they can be incorporated into a conception where different levels of representation are co-present.

Restricting the scope of this chapter to the role of morphological information in phonology, I suggested in the preceding sections that the information required by phonology is that of part-whole relations, categories, and morphological construction types. The construction types that I have demonstrated to be relevant for phonology include the following types of information: morpheme (formatives, features), word, stem, affix, type of affixation, head, modifier, and complement. Many phenomena that I reviewed above can be handled either in terms of sequential or simultaneous / representational modularity. However, there are phenomena for which we only have a sequential analysis at present (e.g., ordering paradox and interaction between two types of compounding). Conversely, there also exist phenomena that can be analyzed only by making direct reference to morphological construction types, challenging alternative analyses within lexical or prosodic phonology. This conflict between competing conclusions will have to be resolved by future research.

Prosodic phonology and prosodic lexical phonology make two claims about phonological organization. First, phonological principles are sensitive to prosodic structures. Second, even though prosodic structures are sensitive to morphological and syntactic information, phonological principles are not directly sensitive to morphology and syntax. The evidence I reviewed above suggests that phonological principles need to refer directly to morphosyntactic constructs such as the morpheme (formative, feature), head, complement, and modifier. These findings challenge the second hypothesis in prosodic (lexical) phonology.

Evidence also points to the conclusion that even though it is necessary to distinguish between different levels of representation and different modules of organization, it is not possible to have any classification of formal principles that correspond to the levels of representation and modules. The same principle can hold in different modules of organization, yielding somewhat different results.

I have suggested that there is a possibility that the submodularity within the lexical module in lexical phonology will have to be replaced by phonological rules having direct access to morphological construction types (with or without word internal prosodic units). Whether or not lexical and post-lexical modularity should also be abandoned is an independent question. My guess at this point is that it will be impossible to state the relation between underlying and phonetic representations without an intermediate level that corresponds to lexical representations. Abstracting away from the sequentiality of lexical phonology, it is possible that the underlying, lexical, and phonetic levels will correspond to the morpheme-, word-, and phrase- levels of representation respectively. It may take

another couple of decades for the dust to settle.

1 I have very little to say in this article about the relation between phonology and syntax; I leave this issue to the articles dealing with the phonology–syntax interface in this volume (chaps. 15 and 16).

2 I use the term “principles” so as to be neutral between the different ways in which phonological regularities are formulated, e.g., in terms of “rules” versus “constraints.” I will use the term “phonological patterns” to refer to both distribution and alternation. “Distribution” is the relation between a unit and the environment in which it occurs: the statement “[h] cannot occur at the end of a word in English” identifies a pattern of distribution. “Alternation” refers to the relation between two corresponding units in two related forms: the statement “The nucleus [ai] in *divine* corresponds to the nucleus [i] in *divinity*” identifies a pattern of alternation. It is important to distinguish the classification of patterns of distribution and alternation from the classification of formal devices that capture the patterns into rules and constraints. Either rules or constraints can be used for the statement of both distribution and alternation.

3 This article is the result of a personal struggle to make sense of a number of debates involving the phonology–morphology interaction that have occurred over the last few decades. In presenting a sketch of one's conception of the state of the art, there is always the danger of a skewed perspective, colored by one's own research and preoccupations. I have tried my best to eliminate autobiography, but total objectivity is impossible when presenting an overall assessment.

4 The flat pseudo–syntactic tree in (2b) resembles the structures in current theories of prosodic structure.

5 The information about stems and affixes is represented in the *SPE* theory and subsequent work in term of bracket. Thus, [[x][y]] denotes compounding, while [[x]y] denotes affixation. This bracket notation has no corresponding translation in the tree notation. See the objections to this notation raised in Halle and Mohanan (1985).

6 Here and in what follows, I formulate phonological principles in ordinary prose rather than in formal notation, because the use of prose statements helps me focus the discussion on the issues of the morphology–phonology interface, without getting sidetracked into debates on the formal devices for the statement of phonological regularities. Another reason is the faint hope that ten years from now, when the formalism for the statement of rules/constraints has changed radically, students and colleagues will still find the discussion useful and accessible.

7 I use the term “devoiced” to refer to an underlying voiced consonant which exhibits very little or no vocal cord vibrations, but still acts like a voiced consonant in its lenis (rather than fortis) articulation and its ability to lengthen the preceding vowel. The symbol of the voiced consonant with the circle at the bottom can represent such a consonant even if the devoicing is complete, i.e., if there is no vocal cord vibration.

8 By morphophonological pattern, I mean a phonological regularity that requires morphological information.

9 In addition to accounting for the alternation in (11), (12) also accounts for the ill–formedness of syllables like *[lisb], *[lizb], *[lifd], *[livt], *[sbin] and *[zbin], in contrast to the well–formedness of syllables like [lisp], [lift], and [spin].

10 This situation can be conceptualized more clearly through the following metaphor (K. P. Mohanan 1993). A magnetic field has a stronger hold on a piece of iron closer to its center than one farther away. When close, a magnet can make a piece of iron on a table jump towards it, but when farther away, its effect is not visible. A weak magnetic field (e.g., (17)) may nevertheless influence the path of a piece of iron which is set in motion by some other field (e.g., (12)).

11 In these formulations, X refers to variables, and the double arrow to a correspondence relation between two words in the lexicon. These authors refer to rules like (19) as Word Formation Strategies in their model of projection morphology.

12 The claim is that all examples of “non–automatic” alternation are part of morphology, not phonology. For Sommerstein (1974), an alternation is automatic if it “occurs in every morpheme of a given general phonological form in a given phonemic environment,” otherwise it is nonautomatic (1974, p. 45). This means that an alternation is “non–automatic” if (1) it has lexical exceptions, or (2) requires morphological information for its application. Under criterion (1), the alternation in (11a–c) is nonautomatic, and hence not part of phonology, as it has lexical exceptions in words like *midst* [midst] and *svelt* [svelt]. Since Singh

considers (11a–c) clearly to be part of phonology, I will not consider (1) to be a defining feature of nonautomaticity.

13 In addition to showing that the claim in (19) is untenable, these examples also show that the claim that there are no morpheme structure constraints is false (e.g., Kiparsky 1982). See Christdas (1988) and T. Mohanan (1989) for additional evidence.

14 For readers who are not familiar with metrical theories of stress: a light syllable is an open syllable (i.e., without a coda) with a short vowel. A syllable with a coda, a long vowel, or a diphthong is a heavy syllable.

15 The [ə] in these forms is epenthetic.

16 Both (33) and (34) express the asymmetry by classifying morphological constructions into two types. Fabb (1988) has argued against this, on the grounds that one needs independent statements on the co-occurrence of individual affixes. See Sproat (1985) for a response to this position, and also for instances of violation of (33c) and (34c).

17 See, however, the discussion in Goldsmith (1990) for objections.

18 In *ungrammaticality*, the hypothesis that *un-* must be attached to adjectives, not nouns, dictates the bracketing [[ungrammatical][ity]], while the hypothesis that *-ity* cannot be attached to a stem containing a class 2 prefix dictates the bracketing [[un][grammaticality]]. Another type of bracketing paradox involves the mismatch between semantically motivated and morphologically motivated constituency. Thus, in the English word *unhappier*, the semantically motivated constituency is [[unhappy][er]], while the morphologically motivated constituency is [[un][happier]]. Neither of these two types of paradoxes are relevant for the issue of the interaction between phonology and morphology.

19 Stress in Malayalam is discussed in detail in K. P. Mohanan (1986).

20 Assuming that all lexical strata are universally required to be cyclic, K. P. Mohanan (1982) was forced to assume that stress and word melody assignment take effect in the co-compounding stratum, thereby making incorrect predictions of the kind pointed out in Sproat (1985, 1986). The formulation given in (39) is taken from K. P. Mohanan (1986), which assumes that all lexical strata in Malayalam are non-cyclic, and that all morphological operations within a stratum precede all the phonological operations within that stratum.

21 Bracket Erasure Convention (BEC) has the effect of erasing the internal brackets of one submodule when the form exits the module.

22 For more extensive data, see Yatabe (1991).

23 See Cole ([chapter 3](#), this volume) for discussion.

24 See Mohanan and Mohanan (1984) and K. P. Mohanan (1986) for details.

25 See Sproat (1985) for the problems of arriving at this interpretation in a theory that combines structure preservation with radical underspecification.

26 The asymmetries mentioned above are not restricted to phonological features. For example, a segment in the coda is typically “weaker” than one in the onset, in the sense that the former undergoes assimilation, neutralization, and deletion more readily. Similarly, word-final segments are weaker than word-initial segments. As far as I know, underspecification theories do not have a unified account of these asymmetries.

Cite this article

MOHANAN, K. P. "The Organization of the Grammar." *The Handbook of Phonological Theory*. Glodsmith, John A. Blackwell Publishing, 1996. Blackwell Reference Online. 31 December 2007
<http://www.blackwellreference.com/subscriber/tocnode?id=g9780631201267_chunk_g97806312012674>

Bibliographic Details

The Handbook of Phonological Theory

Edited by: John A. Glodsmith

eISBN: 9780631201267

Print publication date: 1996