Chapter 1
The Relation between Syntax and Phonology

1.1 An Overview of the Issues

A sentence of a language, uttered, is but a stream of sound, and that stream of sound has associated with it a certain meaning, or meanings. A grammar of a language is a characterization of the relation between sound and meaning for the sentences of that language. This relation between sound (phonetic representation) and meaning (semantic representation) is not a direct one. It is mediated by structure, or syntax, the arrangement of a sentence into parts. The meaning of a sentence is a function of the meaning of its (syntactic) parts, and, too, the sound of a sentence is a function of the sound of its parts. In both instances the function is rather complex. The purpose of this book is to contribute to our understanding of the ways in which the phonology of a sentence may be determined by its syntax, to attempt to characterize the relation between sound and structure in language.

A theory of grammar has the task of characterizing the set of possible relations between sound and meaning (the set of possible grammars) for language in general. The theory of grammar adopted here has developed within the framework of generative grammar—in the works laying the foundation of the so-called standard theory, notably Chomsky 1965 and Chomsky and Halle 1968, and in works that have since contributed to revisions of it. Basic to this theory (in its revised extended form; see Chomsky 1980, 1981) is the assumption that the linguistic description of a sentence involves assigning to it a set of phonological representations $P_1 \ldots P_n$, a set of syntactic representations $S_1 \ldots S_n$, and a logical form, LF. (The logical form of a sentence is not in fact a representation of its meaning, but only a representation serving as a crucial link between the syntactic representation and the semantic rep-
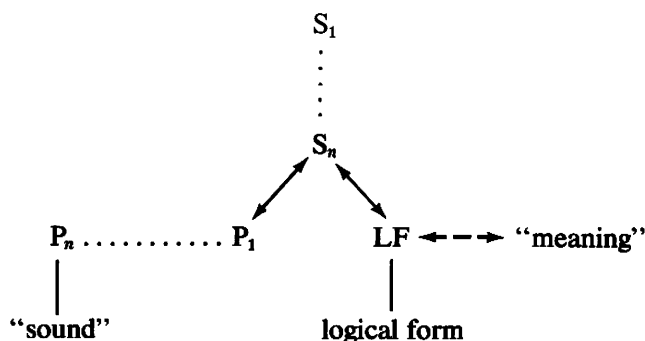
$$S_1$$
$$\vdots$$
$$S_n$$

P_n .... P_1　　　　LF ←– –→ "meaning"

| |
"sound"　　　　　logical form

Figure 1.1

resentation (see Chomsky 1981, for example). Since that distinction is for the most part immaterial to our concerns, however, we will in general ignore it.)[1] The representation of the pronunciation or sound of a sentence, P_n, is seen to relate to its logical form via a system of intermediate representations, where the syntactic representation, the surface structure S_n, is pivotal (figure 1.1).[2]

Thus the linguistic description of a sentence involves a *representation* of that sentence by three different components of the grammar: the phonological, the syntactic, and the "syntacticosemantic." A theory of grammar must specify the nature of the representation by each of these components and what types of rules, if any, define these representations (the *representation question*). It must also specify whether more than one representation by each of these components is relevant to linguistic description and, if so, specify the nature of the set of rules that relate these representations within a component (the *derivation question*). Finally, it must specify the relations between the components, i.e., between the representations defined by the different components (the *interpretation question*).

It is a fundamental assumption of the generative theory of phonology and syntax that only the *surface* syntactic representation, S_n, has a place in characterizing the sound-meaning relation for a language. We believe this assumption to be well founded and will, where appropriate, defend it against evidence (such as that offered by Bierwisch 1968 or Bresnan 1971a, 1972) that might seem to indicate that other, nonsurface representations of the syntax impinge in some way on the phonology. (See chapters 5 and 7.) It is also a fundamental assumption of the stan-

dard theory, one that we adopt in somewhat modified form here, that characterizing the relation between the surface syntactic representation S_n and the surface phonological (or phonetic) representation P_n involves (a) a set of rules defining the mapping between S_n and an underlying phonological representation P_1 and (b) a set of rules mapping P_1 into P_n. The first are rules that *interpret* the syntactic representation as a phonological representation—that is, translate one sort of representation into another. We will call these the *rules of the syntax-phonology mapping*. The second are rules that *derive* one phonological representation from another. These might appropriately be called the *phonological rules* of the grammar. In this book, we will advance and defend a particular theory of that relation between (surface) syntactic representation and (underlying) phonological representation (the syntax-phonology mapping) and will explore the implications of this theory for the theory of phonological derivation.

The conception of the syntactic level, or syntactic component, that doubtless prevails in the current context is the one born with the so-called *standard theory* (Chomsky 1965) and retained, in its most general lines, in the various extensions and revisions of this theory (see, for example, Chomsky 1981 and references therein). This *revised extended standard theory* specifies, as do other syntactic theories, that a syntactic representation S_i is a well-formed labeled bracketing, or tree. It specifies further, and here parts company with some, that a set of syntactic representations $S_1 \ldots S_n$ is associated with any particular sentence. It claims that S_1, the deep structure, can be characterized in terms of (or generated by) a set of context-free rewriting rules (the phrase structure rules) and a lexicon (Chomsky 1965).[3] Finally, it claims that S_1 is related to S_n, the surface structure, by a set of transformational rules. According to the revised extended standard theory, then, there is a transformational derivation within the syntactic level. In the present articulations of the theory, it should be noted, the surface syntactic representation S_n, which is phonologically interpreted (put in direct relation to P_1), is not identified with the only near-surface syntactic representation S-structure, which is put in relation to logical form (Chomsky and Lasnik 1977). The status of this distinction is a matter of debate at present, and we will for the most part ignore it, understanding "surface structure" in the standard sense, as the syntactic representation wherein the full transformational derivation is complete. In recent years, alternative theories of syntax have been proposed that entirely eliminate the transformational derivation, i.e.,

the positing of a distinction between deep and surface syntactic representations. (See, for example, Bresnan 1982 and the works cited therein, and Gazdar 1981.) According to such theories, there is but one syntactic tree representation, the surface structure.

In what follows, the term *syntactic representation* will refer not only to the representation of phrase structure, or syntactic structure, but also to the representation of word structure, or morphological structure. The two share essential formal properties (see Selkirk 1982). Yet it will prove useful to maintain a distinction between sentence grammar, on the one hand, and word grammar, on the other, the first characterizing the sound-meaning relation of the sentences of a language and the second the sound-meaning relation of its words (see Bach 1983). In this we depart somewhat from the standard theory. A word grammar might consist of a word-syntactic component, characterizing the possible word structures of the language (see Selkirk 1982, for example), a phonological component interpreting these word structures phonologically, and a semantic component. The rules and principles of these components, and their interaction, could conceivably be rather different from those of sentence grammar. While it is our contention that word grammar and sentence grammar are in fact parallel in many important ways (see the discussion in sections 3.1, 3.4, 8.3), establishing that parallelism is not a primary concern of this book. Rather, we seek to investigate questions primarily involving sentence grammar and the syntax-phonology interaction within it.

To make the focus on sentence grammar entirely clear, we will simply construe the syntactic surface structure of a sentence as consisting of a sequence of words, defined by a separate word grammar. The aspects of these words that are relevant to sentence grammar are (i) their (surface) word structure, (ii) their (derived) phonological representation (i.e., the output of the phonological rules of word grammar), which we will call the *word-level phonological representation,* and (iii) their semantic representation. The second, of course—and perhaps the first as well—is directly relevant to our main concern, which is the phonological interpretation of syntactic structure.

Whereas the answer to the representation question in syntax has remained more or less the same in the successive articulations of syntactic theory within the generative framework, this is not true of the answer to the representation question in phonology. The last decade has seen fundamental changes in the theory of phonological representation, changes whose implications are perhaps not fully understood,

but that clearly require radically different answers to the derivation question and the interpretation question (the relation between syntax and phonology) than the standard theory offered, especially as articulated in *The Sound Pattern of English* (Chomsky and Halle 1968; hereafter SPE).

Common to most approaches to the linguistic representation of the sound of sentences is the notion that the sound continuum must be analyzed as a sequence of discrete sound *segments*. Generative phonology has characterized a sound segment as a complex of distinctive features, and thus has construed the sequence of sound segments making up an utterance as a *distinctive feature matrix*. Within the standard theory, all phonological properties of an utterance, even those termed "suprasegmental," such as tonal contours and stress patterns, are held to be "segmental," in the sense that they can be reduced to a representation in terms of the distinctive feature complexes forming part of the unilinear sequence. A sequence of segments, then, is taken to be the representation of what is properly phonological—of what, ultimately, is pronounced. Yet, as has long been recognized, a sequence of phonological segments alone cannot permit an insightful description of the significant phonological properties of an utterance, for it can be shown that there exist different sorts of relations between the segments in sequence, relations that may be thought of as varying "degrees of connectedness." Standard generative theory, following in the steps of American structuralist linguistics, represents some of these relations between segments as "juncture" elements, or *boundaries,* proposing that these boundaries are themselves segments, occupying a place between the truly phonological segments in the strictly linear arrangement of phonological representation. Moreover, according to standard generative phonology, characterizing the relations between segments in a phonological representation of a sentence also involves the full range of information represented in the labeled bracketing or tree of its syntactic representation. A major contribution of SPE and other early works in the generative tradition was the demonstration that certain phonological properties of sentences, in particular their stress patterns, are determined in a rather direct way by their (surface) syntactic structure. Thus, in the standard framework, phonological rules were seen to apply to the segments of phonological representation in virtue of the syntactic constituent structure relations obtaining between the segments (or subsequences of them). For the standard theory, a phonological representation is a syntactic labeled tree or bracketing of

a terminal string consisting of a sequence of sound segments and boundaries.

Surface structure, in the standard theory, is a labeled bracketing of sound segments. The relation between surface structure and underlying phonological representation is defined by rules of a "readjustment component." On this theory the underlying phonological representation of a sentence differs little from surface structure: it contains the same sequence of segments (with some possible additions, made by readjustment rules that "spell out" "empty" morphemes),[4] and it has more or less the same labeled bracketing (though this may be modified somewhat by readjustment rules, in ways never made explicit). The essential difference between the two is the presence of boundaries in phonological representation. These grammatical formatives are said to be introduced by a set of conventions forming part of the readjustment component, which insert boundaries in the phonological representation on the basis of the surface structure (SPE; Stanley 1973; Selkirk 1972, 1974). The boundaries of the standard phonological representation constitute a very rough translation, into linear terms, of the hierarchical syntactic structure of the sentence. Whereas some rules of the phonological component apply directly in terms of the syntactic labeled bracketing of the sentences, others, according to the standard theory, appeal only to the relational information encoded in boundaries.

It is now more than a decade since what came to be called the standard theory of generative phonology was propounded in *The Sound Pattern of English*. And it is by now well established that a phonological representation is more than a mere string of segments (sound segments and boundaries) with an associated syntactic structure. It is known that a phonological representation consists of a sequence of *syllables,* and that the syllable has an internal constituent structure, its "terminal" positions coinciding in general with what we know as segments.[5] It is also known that there may be more than one *autosegmental tier* in a phonological representation, and that on each of these independent tiers phonological features or feature bundles are arranged (as segments, or "autosegments," to use Goldsmith's term) in linear fashion.[6] Following recent work on this topic (see, for example, Halle and Vergnaud 1980), we will view the sequence of syllables (or, perhaps, their terminal positions) as the *core,* or *axis,* of a three-dimensional object in which the autosegmental tiers are parallel to the axis, the (auto-)segments of the tiers being "connected" to one or more consecutive positions in the axis by "association lines." (See especially Halle and

Vergnaud 1979, McCarthy 1979a, Clements and Keyser 1981, Selkirk 1984.) Moreover, it is known that the syllables of phonological representation are arranged in some kind of hierarchical organization. By "hierarchical organization" we do not mean the complex of autosegmental tiers; the elements of the various tiers appear not to be related to each other hierarchically or in any direct fashion whatsoever, but to be directly related only to the syllable axis. (On one autosegmental tier will be represented the tonal segments consisting of the tonal or intonational contours, on another the features involved in vowel harmony, and so on.) By "hierarchical organization" we mean, very roughly speaking, the organization of the units of phonological analysis into layers, vertically arranged on the same plane. Just what the nature of that hierarchical phonological representation is will be a major focus of this book.

There are in fact two distinct sorts of hierarchical organization that form part of a phonological representation. One is what may be called *prosodic constituent structure* (a term that includes *metrical trees;* see section 1.2). It is a structure of the same general sort that is familiar from syntactic description, one in which linguistic units are grouped into yet larger units, constituting a well-formed labeled bracketing or tree. The syllables (and their internal constituents) are clearly units of this hierarchy, as are, above them, intonational phrases. What units, if any, may intervene between syllables and intonational phrases in this prosodic constituent structure is a matter of some debate (section 1.2). The other sort of hierarchical organization within the phonological representation of a sentence is a representation of its *rhythmic structure*. Rhythmic structure per se can be represented as a *metrical grid* (Liberman 1975). A metrical grid is a representation of a hierarchy of temporal periodicities. It consists of a hierarchy of metrical levels, each level in turn consisting of a sequence of positions (beats) that stand for points in (abstract) time and define the recurring periodicities of rhythm; it is not a tree. The rhythmic structure of a sentence is the alignment of its syllables with a metrical grid.

Why these hierarchical aspects of "nonlinear" phonological representation deserve special attention here should be obvious. This conception of phonological representation as having its own hierarchical structure(s) demands a radical rethinking of the relation between syntax and phonology. Phonological representation can no longer be seen simply as a "readjusted" surface structure. It has its own defining properties. Thus the interpretation question—the question of the map-

ping between phonological representation and syntactic representation—takes on a much greater importance than in the standard theory, and has an entirely different quality to it. It must be viewed as a characterization of the relation between the syntactic hierarchy, on the one hand, and the phonological hierarchy (or hierarchies), on the other.[7]

This emerging richer conception of a phonological representation has further implications for the theory of the relation between syntax and phonology. We believe that it is in terms of the hierarchical organization(s) of phonological representation that the "juncture" or the "degrees of connectedness" between the segments of phonological representation that may affect the application of phonological rules should be represented. Over the years, it has been argued (in a number of different ways) that the junctural properties of sentences should somehow be represented "suprasegmentally" rather than as the segmental boundaries of the standard theory (see, for example, McCawley 1968, Pyle 1972, Selkirk 1981a, Rotenberg 1978, Basbøll 1978). Here we take that line of thinking one step further and propose that these junctural properties be characterized in terms of the already independently motivated hierarchical structures of the representation. Thus the theory of phonological representation that we will advocate here eliminates segmental boundary elements altogether. (See chapters 3, 6, and 7 for discussion of the role of the phonological hierarchies in supplanting boundaries.)

In sum, the "revised theory" of the phonological representation is that it consists of (a) a *prosodic constituent structure* (including a sequence of syllables), (b) a set of *autosegmental tiers,* (c) a rhythmic structure, the *metrical grid,* and (d) a specification of the *associations or alignments* between these various aspects of the representation. The "revised theory" of the relation between syntax and phonology is that it is a mapping from a syntactic representation into a fully specified phonological representation with these properties.

Our theory departs from the standard theory of the syntax-phonology relation in another way as well. As mentioned earlier, we will assume that a grammar consists of a word grammar and a sentence grammar, and that the syntax-phonology relation must be characterized for both. Given our assumption that syntactic surface structure consists of a sequence of words (the outputs of word grammar), with their individual word-level phonological representations,[8] our goal of characterizing the syntax-phonology relation for sentence grammar strictly speaking commits us only to investigating the phonological

properties of the sentence that are governed by the rules and principles of sentence grammar. (Though, of course, since there is significant overlap in the rules and principles of the two subgrammars, there will be occasion to examine the syntax-phonology mapping in word grammar as well.)

As we will show, the properly syntactic aspects of the surface syntactic representation play a crucial role in determining just how the phonological representation of the sentence is hierarchically organized—in governing how the phonological representation is, in essence, "constructed." Whether the surface syntactic labeled bracketing has any greater role in phonological description—that is, whether or not it actually governs the application of phonological rules of sentence grammar (and thus has a direct role in a phonological derivation)—is debatable. We will argue that in the unmarked case, phonological rules of sentence grammar are affected by syntactic structure only indirectly, through the influence of syntactic structure on the hierarchical structure of phonological representation. (See especially chapter 6.) It is the latter sort of structure that appears to govern the application of the vast majority of phonological rules. Note that an important conceptual distinction is being made here between *phonological rules,* which apply in a derivation in terms of a phonological representation, hierarchically arranged, and *rules for constructing or defining the representation* (e.g., rules of syllabification and resyllabification (see section 1.2.3), rules of intonational phrasing, rules aligning syllables with the metrical grid, and so on), which apply (partly) in terms of surface syntactic representation. The latter rules define the mapping between syntactic and phonological representation, and will be the focus of concern in this book.[9]

1.2 Hierarchical Structures in Phonology

1.2.1 Rhythmic Structure
Liberman 1975 proposes a formal representation of rhythmic structure for language that embodies the claim that the rhythmic organization of speech is quite analogous to that of music. This representation is called the *metrical grid.* We will argue here that the metrical grid forms an integral part of the phonological representation of the sentence, that it is in terms of the grid that patterns of stress or prominence are to be represented, and that it is in terms of the grid that a theory of stress patterns in language must be couched. Prince 1981, 1983 argues con-

vincingly that a theory that views stress patterns as resulting from a set of rules for defining directly the alignment of the syllables of the sentence with the rhythmic structure of the metrical grid is not only possible, but highly desirable, and preferable to other approaches to the analysis of stress patterns. We view Prince's position as essentially correct and will elaborate on it here.

Before introducing the metrical grid, we will review the general characteristics of musical rhythm that any formalized system must represent. In their discussion of the temporal organization of music, Cooper and Meyer 1960:3 define a *pulse* as "one of a series of regularly recurring, precisely equivalent stimuli," as for example the ticks of a clock (or a metronome). The existence of a regular succession of pulses is of course prerequisite to any organization into rhythmic patterns, the patterns themselves being impossible in the absence of this basic regularity. For Cooper and Meyer, the pulses of musical time are arranged into *metrical patterns:*

Meter is the measurement of the number of pulses between more or less regularly recurring accents. Therefore, in order for meter to exist, some of the pulses in a series must be accented—marked for consciousness—relative to others. When pulses are thus counted within a metric context, they are referred to as *beats*. Beats which are accented are called "strong"; those which are unaccented are called "weak." (Cooper and Meyer 1960:3)

To see this, consider the following diagram, where the x's are taken to be pulses (points in time) and the underlining indicates which pulses are "marked for consciousness."

(1.1)

$\underline{x}$ x $\underline{x}$ x $\underline{x}$ x $\underline{x}$ x $\underline{x}$ x

Cooper and Meyer will call the pulses here *beats,* because those that are marked for consciousness (accented) are organized into metrical patterns; that is, they recur in regular fashion—in this case in a strictly binary alternating pattern. (For Cooper and Meyer, the underlined beats are the strong beats; the others are weak.) Keeping in mind that what Cooper and Meyer call *meter* in music is what we, in referring to speech, are calling *rhythm,*[10] we will use this musical analogy, as Liberman (1975) has done, to pursue the analysis of speech rhythm and its implications for linguistic representation. Now, in the temporal organization of music, there is a hierarchy of levels (Cooper and Meyer 1960:4–5), each with its own beats, strong and weak; that is, there are

various levels of pulses, and on each level the pulses are organized into metrical patterns. In addition,

[in] the metrical schemes of Western classical music, each level of the hierarchy is periodically regular; the "pulse" at a given level is fixed (with some exceptions) at a periodicity which coincides with the periodicity of the next level up in a constant way, generally either two to one or three to one. (Liberman 1975:272)

Any representation of the rhythmic organization of music, or of any system with a rhythmic organization like that of music, must thus give representation to pulses, or beats, to the distinction between strong and weak beats, and to the various levels on which these strong-weak distinctions may obtain. The metrical grid, proposed by Liberman 1975, is just such a representation. We will outline its essential features here; for a discussion of its formal properties, see chapter 5 of Liberman 1975 and section 3.3 of Liberman and Prince 1977. It is a two-dimensional object consisting of parallel horizontal levels on which there are points, marking periodicities. Because the periodicities are hierarchically arranged, any point on a higher level will coincide (vertically speaking) with a point on a lower level (though not vice versa). (1.2) shows a well-formed metrical grid:

(1.2)

```
x                       x
x           x           x
x   x       x   x       x   x   x
x x x x x x x x x x x x x x x x x
```

The horizontal levels will be called *metrical grid levels,* or *metrical levels,* for short.[11] For reasons to be explained later, the points on the lowest metrical level will be referred to as *demibeats.* All points on the second metrical level and above will be referred to as *beats.* A beat (or demibeat) that does not coincide with a beat on the next higher metrical level will be referred to as a *weak* beat (or demibeat). A beat (or demibeat) that does coincide with a beat on a higher metrical level will be referred to as a *strong* beat (or demibeat).

Note that there is nothing in the grid representation itself that specifies the nature of the periodicities of the pulses (metrical patterns) on any metrical level. In principle, any sort of pattern "graspable by the mind"[12] could be employed in a rhythmic organization and be represented by the metrical grid. But in fact, in many sorts of rhythmically organized activity, whether it be Western classical music, dance,

marching in military style, or uttering the syllables of a language, there is a noted tendency to an alternation of strong and weak beats. As a variant on this binary organization, one may encounter ternary beats (a strong accompanied by a sequence of two weaks), but quaternary groups seem to be felt as two binary. Thus there may be some quite general *Principle of Rhythmic Alternation* lying behind the patterns attested.[13] We offer a provisional formulation of this principle: that between two successive strong beats there intervenes at least one and at most two weak beats. We take this principle to be a plausible, not entirely far-fetched hypothesis about the nature of rhythmic patterns. It is an empirical question, of course, whether such a principle is really at play in the diverse realms of human activity that may be subject to rhythmic organization.[14] It may be that there is no magic about the number two and its variant three, and that the patterns attested are more heterogeneous. But we will for the time being take this Principle of Rhythmic Alternation as a working hypothesis to guide our investigation, holding onto the idea that there is more to rhythmic organization than the mere existence of patterns of beats at various levels, that there may be something important to be learned about rhythm, and human cognitive capacities, in determining the types of patterns that are involved.

1.2.2 The Role of Rhythmic Structure in Linguistic Description

Liberman's claim, which we support and for which we will present additional evidence, is that the rhythmic organization of natural language is analogous to that of music. More specifically, the claim is that it is appropriate to represent the rhythm of an utterance as the alignment of its syllables with a metrical grid, in particular a metrical grid governed by something like the Principle of Rhythmic Alternation in its provisional form. Liberman's position on the place of rhythmic organization in linguistic description can be construed as one according to which the rhythmic organization of speech is a relatively superficial phenomenon, produced as part of the phonetic implementation of a more basic phonological representation having quite different properties. According to this theory, the metrical grid alignment of the sentence is a representation in terms of which such things as the tendency toward the isochrony of stressed syllables and, more generally, the relative durations of syllables might be expressed (see chapter 2). Liberman 1975 and Liberman and Prince 1977 also assign it a role in characterizing the conditions under which the stress shift rule of En-

glish (which gives *thírtèen mén* from *thìrtéen mén*) might apply.[15] However, they do not assign it a role in characterizing the basic patterns of stress, or prominence, in language. For Liberman 1975, the description of stress requires (first) a representation of the organizaton of the utterance into a "metrical pattern"[16] and (then) a representation of its organization with respect to the metrical grid. The former is the system of relations of relative prominence obtaining among the elements of the utterance (its syllables, words, and phrases). For Liberman, it is a representation in terms of a binary-branching tree structure whose nodes are labeled *s* (strong) and *w* (weak)—a *metrical tree*.[17] Thus Liberman 1975, Liberman and Prince 1977, and other more recent works in the "metrical" tradition posit as part of linguistic description an abstract stress pattern, which is independent of the metrical grid and which mediates between it and the syntactic structure of the sentence.[18]
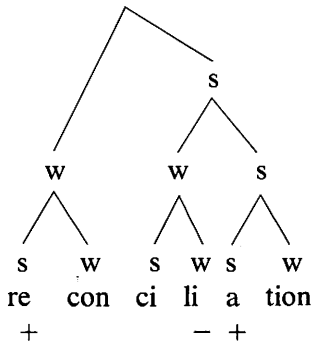
Liberman and Prince propose that relations of relative prominence above the level of the word be represented merely by annotating the nodes of the surface syntactic tree of the sentence with the labels *s* and *w*.[19] Under this analysis, the English Nuclear Stress Rule becomes simply "label the right-hand node *s*" (from which it follows that its sister will be *w*). The sentence *Mary's sister adores Russian novels* is thereby assigned the metrical pattern in (1.3):

(1.3)

R
w   w
w   s   w   w   s
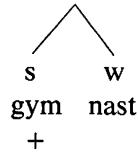   s

Mary's sister adores Russian novels
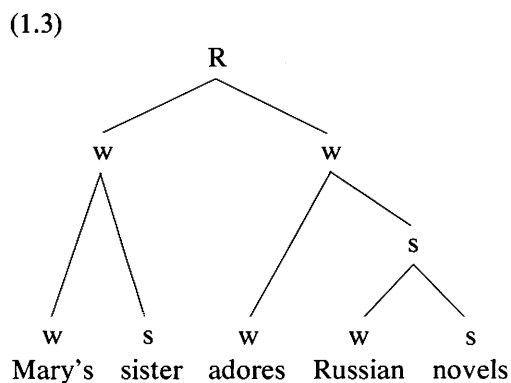
(*R* is the root of the tree.) As for patterns of word stress, Liberman and Prince do not represent them wholly as *s/w*-labeled trees, but retain the feature [stress], now merely binary, to express the distinction between stressed and unstressed syllables. The words *reconciliation, gymnast,* and *modest* are represented as follows:[20]

(1.4)

a.   b.   c.

```
              /\
             /  \
            /    s
           /    /\
          /    /  \
         w    w    s
        /\   /\   /\
       /  \ /  \ /  \
      s   w s  w s   w          s    w          s    w
      re con ci li a tion       gym nast        mo dest
      +        - +              +
```
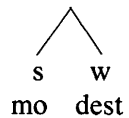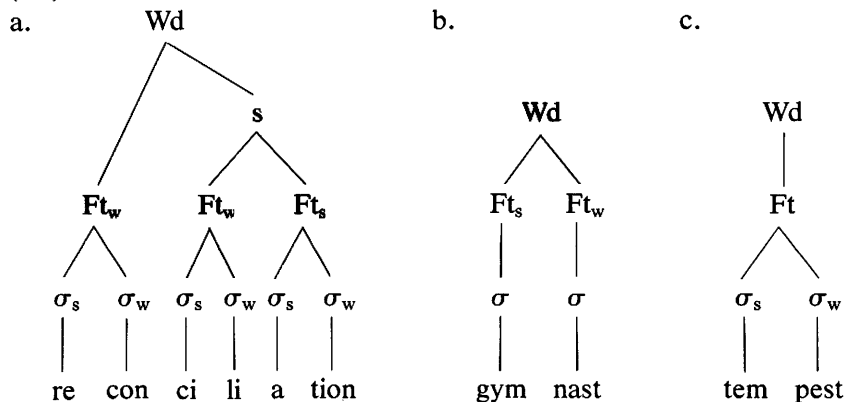
Thus prominence relations among stressed syllables, whether within the word or on the phrase, are uniformly represented with metrical trees. Liberman and Prince's proposals have given rise to an extremely fruitful line of research, into word stress in particular. (See Prince 1976, Halle and Vergnaud 1979, Kiparsky 1979, McCarthy 1979a, Safir 1979, Selkirk 1980b, Prince 1980, and Hayes 1980, among others.) From this work, most notably that of Halle, Vergnaud, and Hayes, very important insights have been gained into possible patterns of stress, and a theory of the parameters involved in a universal theory of stress (at the word level) has been evolving. This research has shown that a certain enrichment of the hierarchical branching tree representation permits a representation and characterization of word stress patterns that does away with the feature [stress] entirely. This enrichment involves introducing units of prosodic constituent structure[21] into the description— in particular, the units *syllable, foot,* and *(prosodic) word*. With this elaboration, the representation of the words in (1.4) is instead seen to be something like (1.5):
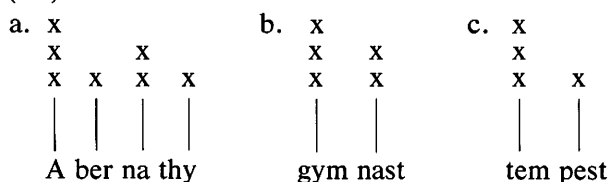
(1.5)

a.

$$Wd$$

Ft_w Ft_w Ft_s

σ_s σ_w σ_s σ_w σ_s σ_w

re   con   ci   li   a   tion

b.

$$Wd$$

Ft_s Ft_w

σ σ

gym   nast

c.

$$Wd$$

Ft

σ_s σ_w

tem   pest

A stressed syllable (σ) is here represented as the strong(est) syllable of a foot (Ft). (The sole syllable of a monosyllabic foot is by convention considered strong.) A stressless syllable is one that is weak. The description of the distribution of stressed and unstressed syllables in words is no longer a matter of rules assigning the feature [±stress], but of rules that indicate (among other things) what constitutes a well-formed foot in the language, often in terms of the nature of the component syllables. On this view, the foot is a unit of phonotactic description, much like the syllable.[22] In the more recent articulations of "metrical theory," then, the abstract stress pattern of words and sentences is represented in terms of prosodic constituent structure with an *s/w* labeling of the nodes. Though most research in the metrical framework (with the exception of Dell (to appear)) has not attended to the place of the metrical grid in linguistic description, Liberman's theory of an abstract stress pattern that is ultimately translated into a metrical grid representation is more or less presupposed.

Prince 1981, 1983 has argued, however, that metrical trees should be eliminated in a theory of stress, and that the metrical grid must be given the fundamental role in the representation of prominence relations and in the theory of patterns of prominence. This is the position that we will develop here.[23]

The notions "stressed," "unstressed," and "degree of stress" are straightforwardly represented in the alignment of syllables with the grid. In grid terms, a stressed syllable is one that is aligned with a beat (or basic beat, or strong demibeat—all are equivalent with respect to this definition); an unstressed syllable is one that is instead aligned with a weak demibeat. As for degrees of stress, one syllable has "more

stress" than another if the beat aligned with the first coincides with beats on a metrical level higher than that of the beats aligned with the second. Consider in this light the following stress patterns:

(1.6)

```
a.  x                   b.  x              c.  x
    x       x               x    x             x
    x   x   x   x           x    x             x    x
    |   |   |   |           |    |             |    |
    |   |   |   |           |    |             |    |
    A ber na thy           gym  nast          tem  pest
```

In all cases, the first syllable has the greatest stress (is the most promi-nent); it is the only one to be aligned with a beat on the highest metrical level. In (1.6a), the first and third syllables are stressed (associated with beats), and the others are stressless. In (1.6b), both the first and the second syllables are stressed, and there are no stressless ones. And in (1.6c), the first syllable is stressed, and the second stressless.[24] Clearly, the grid theory of stress has the means to represent the distinctions needed for an insightful analysis of stress, at the word level or higher.

In arguing, in general terms, for a "relational" representation of stress, as against the standard "numerical" representation, Liberman and Prince 1977:261–264 make the point that a relational theory of stress is to be preferred in that it makes understandable ("rationalizes") the array of special properties that characterize the stress feature and stress-assigning rules in the standard framework, e.g., the n-ary nature of the stress feature, the syntagmatic character of nonprimary stress, the nonlocal effects of the Stress Subordination Convention, and so on. We believe this is an argument of fundamental importance: with a rela-tional theory of stress, these special characteristics "follow directly from the way the phenomenon is represented, rather than being arbi-trary typological observations" (Liberman and Prince 1977:263).

In describing the advantages of a relational theory of the representa-tion of stress, Liberman and Prince had in mind a particular theory of relational representation: metrical trees. But, we submit, their argu-ment could just as well be taken as one in favor of representing stress with the metrical grid, for the alignment of the sentence with the metri-cal grid is, in part, a relational representation of stress, in Liberman and Prince's sense of the term. There is no upper limit on the number of metrical levels that contain beats aligned with syllables ("stress" is n-ary). A beat may be added, moved, or eliminated at some remove from others, but nonetheless depending on the presence of some

other(s) within the same grid ("stress" is nonlocal). And, as we will show in chapter 2, whether or not a syllable bears a certain "degree of stress" is a function, in part at least, of the "degrees of stress" of syllables in the surrounding context (nonprimary "stress" is syntagmatic).

But stress is not strictly relational. It is well known that an individual syllable may be stressed (rather than stressless) regardless of whether its neighbors are stressed. Within metrical grid theory, a syllable is "stressed" only by virtue of its alignment with a basic beat in the metrical grid, and there is nothing to prevent a sequence of syllables from being aligned with basic beats. In that sense, "stress" is not relational in metrical grid theory. It should be noted that a basic-beat-aligned syllable in the present theory is the analogue of the strong (or only) syllable of a foot in the revised metrical tree theory, which includes prosodic constituents as part of the representation of stress. According to the latter theory, it is possible for monosyllabic feet to succeed each other, and thus possible for a sequence of syllables to be stressed—that is, for a syllable to be stressed without regard to its neighbors (Selkirk 1980b). There is also a second sense in which stress is not relational: a syllable with "main word stress" is always more prominent than a syllable that is merely "stressed," (section 3.2.2). This greater prominence is represented in metrical grid theory as an alignment with metrical level three or higher. To capture this "inherent" greater prominence of main word stress in metrical/prosodic tree theory, it could perhaps be stipulated that a syllable that is the strongest within a prosodic word be interpreted as more prominent than one that is merely the strongest within a foot. (In a metrical tree theory without prosodic category levels, such a relation could not be expressed at all.)[25]

The metrical grid theory of stress thus gives a uniform representation of both the relational and the nonrelational aspects of the stress patterns of words (and phrases), while a metrical/prosodic tree theory expresses relational concepts by labeling trees with *s* and *w*, and nonrelational concepts by means of the organization of the tree into prosodic constituents. We will show in the course of what follows that the homogeneous and highly restrictive representation of stress patterns offered by the metrical grid is quite adequate to the descriptive task, and moreover that it provides the basis of an explanation for many stress-related phenomena.

We ask of any theory of stress not simply that it make available an appropriate representation of stress patterns and that it permit an insightful analysis of the stress patterns of particular languages, but also

that it constitute a theory of the notion "possible stress pattern in language." Actually, it is an empirical question whether there is anything of interest to be said about universals of stress at the word level or above, whether there is anything at those levels that calls for explanation. Research on word stress, which has recently become quite intensive, has led to the conclusion that there are indeed highly interesting things to be said in a theory of word stress. A first apparently universal property of word stress, one that distinguishes it from other phonological phenomena, is that there are patterns to it—there are discernable regularities in the occurrence of stressed and unstressed syllables, as well as regularities in the location of primary stress. It is not the case that the distribution of unstressed, stressed, and main-stressed syllables is random in language (see chapter 2). This observation has serious implications for any theory of stress.

Suppose that stress were a feature of vowels, as in the standard theory of generative phonology. Such a theory would have to accord the feature [±stress] a status different from that of any other feature characterizing vowels or other segments. If [±stress] were a feature just like [±high], for example, there would be no more reason to expect any particular word to contain a vowel specified [±stress] than to expect it to contain a [+high] vowel. Nor would there be any reason to expect vowels specified [+stress] to be arranged in any particular pattern with respect to each other within the word. Without additional stipulations, the standard theory can explain neither the reliable presence of stressed syllables nor their patterning. For the standard theory, the presence of stress patterns would have to be reflected in the stipulation that, universally, grammars include rules for assigning the feature [stress], rules stipulated to be of just the sort that give rise to the patterns attested. But the theory cannot explain why it should be stress, as opposed to some other phonological feature, that has this privileged status in the grammar, and it cannot explain why the word stress patterns should be as they are. For these and other reasons, the standard theory of stress has been recognized to be inadequate.[26]

The very existence of patterns as a fundamental property of stress could be taken as indicating that in the phonological representation of stress, patterns are somehow primitive. This point of view has influenced recent work on stress, which, following Liberman's lead, has viewed patterns of stress as reflecting hierarchical (patterned) arrangements of the syllables of the utterance, represented as metrical trees. But metrical tree theory is unable, without further stipulation, to pro-

vide any insight into the notion "possible pattern." There is nothing about metrical trees from which it follows that in stress-timed languages the number of stressless syllables intervening between stressed syllables is usually one and on occasion two, and that a pattern will never be based on intervals of two, or three, or more. In one articulation of the metrical tree theory of word stress, Hayes 1980 proposes that binary feet be stipulated as the only possible basic pattern, ternary feet being allowed as variants under special circumstances. Binarity *is* fundamental to patterns of stress, particularly at the lowest levels; yet in metrical tree theory this is merely stipulated. While acknowledging that such stipulations are not in principle objectionable, if part of a universal theory, we submit that a theory of stress from which it specifically follows that patterns would have this shape is to be preferred. The rhythmic theory of stress is just such a theory. If the alignment of syllables with a rhythmic structure such as the metrical grid is the representation, then these patterns are the expected ones, for rhythmic structure, be it in language or any other human activity, is governed by something like the Principle of Rhythmic Alternation (PRA). Roughly speaking, the PRA ensures that rhythmic clashes and lapses will be avoided, at all metrical levels, and that strong beats or demibeats will appear at regular intervals, two or three beats or demibeats away from a preceding or following strong. The patterns exhibited in well-documented languages appear to conform to this organization: on the third metrical level it is common for beats to be either two or three basic beats away from each other; at the basic beat level, it is even more common for beats to be two demibeats away from each other, ternary beats being allowed only in particular cases.[27] The claim, then, is that a metrical grid theory of stress is better able to explain these stress patterns.

The theory of stress patterns proposed here, to which chapter 2 provides a general introduction, is that they result from the conjoined effects of two sorts of rules: (i) *text-to-grid alignment rules* and (ii) *rules of grid euphony*. The *text* is surface structure as defined earlier, and text-to-grid alignment (TGA) rules construct a partial grid, aligning certain syllables with beats on various levels of the grid by virtue of their internal composition and/or their position within specified syntactic domains. The Nuclear Stress Rule is a rule of this sort. TGA rules establish fixed loci of prominence from which the alternations so characteristic of stress patterns emanate, introduced by the rules of grid euphony (GE). Rules of grid euphony thus complete the construction of

the grid. We suggest that they are defined solely in terms of the grid and apply at all metrical levels. Their role is to ensure that the grid is truly rhythmic, to make it conform as closely as possible with the PRA. The stress shift or "Rhythm Rule" of English is a rule of this sort. In this analysis, both GE rules and TGA rules belong to the component defining the mapping between surface syntactic representation and underlying phonological representation.

Now, the rhythmic structure of an utterance is in fact more than the representation of the prominence patterns of its syllables, more than beats of the grid that are aligned with syllables. Liberman 1975 suggests that the metrical grid of an utterance may also contain *silent grid positions*—positions not aligned with syllables, whose presence is determined in some way on the basis of the syntactic structure of the utterance. (Abercrombie 1968 refers to them as *silent stresses*.) Liberman takes these silent positions to be the means by which the apparently syntactically governed phenomena of pausing and final lengthening are to be explained. We will call these positions the *syntactic timing* or *juncture* of the sentence. As the book progresses, we will elaborate to a considerable extent the idea that there are silent positions within the metrical grid with which the syllables of an utterance are aligned. In chapter 6 we will show not only that this idea allows us to explain why the phenomenon of final lengthening should exist (alongside pausing) in the first place, but also that it provides the appropriate means of representing some of the junctural properties of the sentence and thus ultimately explains differences in applicability in different locations of the sentence of grid-based rules like stress shift, as well as of phonological rules of sandhi.

In sum, the rhythmic structure of a sentence consists of a metrical grid containing grid positions with which syllables are aligned, giving representation to *patterns of prominence*, and silent grid positions, giving representation to *syntactic timing* or *juncture*. The representation of the rhythmic structure of the sentence *Abernathy gesticulated*, for example, we claim to be as follows, where the underlined positions are the silent ones.[28]

(1.7)

```
                    x
x                   x
x     x             x   x
x  x  x  x  x x  x  x x x x  x x x
Abernathy      gesticulated
```

It is worth noting in this connection that the demonstration that there must exist positions of rhythmic structure that lack an alignment with syllables provides important evidence for construing rhythmic structure as independent of segments and their organization into syllables.

If it is true, as we claim, that the application of numerous phonological rules of sentence grammar is governed by the "adjacency" of syllables and segments defined with respect to the grid, then it must be concluded that the grid is present at an early point in the phonological derivation. We will assume that it is present in the (underlying) phonological representation, P_1, the output of the syntax-phonology mapping. This position, it should be noted, is the only one consistent with both the fact that (some of) the timing relations represented in the grid are determined directly by the surface constituent structure of an utterance (see chapter 6) and the interestingly restrictive assumption that syntactic structure is not available to the phonology (or phonetics), once the mapping from syntactic representation to phonological representation (via "construction rules") is complete (see section 1.3). Giving the metrical grid a place in the (underlying) representation P_1 does not force the conclusion that a metrical grid alignment is the one and only representation of the stress pattern of an utterance, however, for that representation also consists, we believe, of a prosodic constituent structure. Given this, it would be entirely possible to entertain a theory like Liberman's, according to which stress patterns are represented fundamentally in terms of the prosodic constituent structure (metrical trees), and according to which prosodic constituent structure has an immediate, simultaneous translation into a metrical grid alignment for the sentence. (This is the approach taken in very recent work by Halle and Vergnaud, for example.) Given this approach, the theory of stress— that is, the theory of the notion "possible stress pattern" in language— would be cast in terms of prosodic constituent structure. Our claim, with Prince 1981, 1983, is a different one. It is that: (i) stress patterns are represented only in terms of the metrical grid alignment, not at all in terms of prosodic constituent structure, and (ii) the theory of possible stress patterns is cast in terms of a theory of syllable-to-grid alignment. Much of this book is devoted to arguing for this position.

Chapter 2 reviews the general motivation for attributing a rhythmic structure to an utterance and outlines the proposed theory of metrical grid alignment for words and sentences. Chapter 3 sketches a theory of English word stress based on the metrical grid, drawing on insights gained by the earlier metrical framework concerning the nature of

stress patterns in general and English word stress patterns in particular. Chapter 4 shows that a metrical grid approach allows an insightful characterization of all phrasal rhythmic phenomena and argues against a metrical tree approach to phrase stress. Chapters 4 and 5 also present a grid-based theory of the relation between stress and intonation. With this, we hope to make clear that only the representation of relative prominence available in the grid is needed for an adequate, insightful description of the tunes and rhythms of the sentence. Thus we maintain that the metrical grid alignment of the sentence is the one and only representation of the stress or prominence relations of the sentence, as of the word.

1.2.3 Prosodic Structure: The Syllable

The syllable had no place in standard generative phonology (represented by SPE), though most other theories of phonology have recognized its fundamental importance.[29] Studies in the last decade in the generative framework have given the syllable an ever larger place in the theory, both as a unit of phonological representation and as a unit in terms of which many generalizations about phonological representation and phonological rules are expressed. The syllable is the paradigm case of a unit of prosodic constituent structure, and so will provide a point of reference in discussing the status of other such hierarchical units in the theory.

The syllable is now understood to be a "suprasegmental" unit. Departing from earlier works in the generative tradition such as Hooper 1972, 1976, Vennemann 1972, and Hoard 1971, which defined syllables in terms of syllable boundaries, Kahn 1976 and Anderson and Jones 1974 proposed that the syllable is a separate unit "standing above" the segmental string, to which the segments are "associated." Selkirk 1978b,c, Kiparsky 1979, McCarthy 1979a,b, Halle and Vergnaud 1979, and others have argued further that the syllable has an internal constituent structure, the segments being the structure's terminal string; their work thus rejoins earlier theories of the syllable such as those of Pike and Pike 1947, Kuryłowicz 1948, and Fudge 1969. Recent studies have proposed that the terminal positions of this syllable structure hierarchy are only "placeholders" of sorts, and that segmental material is represented on (one or more) autosegmental tiers, separate from those terminal positions, and associated with them by rules having a substantive character.[30] Given this more recent articulation of the theory, which we adopt here, the syllable and its internal structure form

the *core* or *axis* of phonological representation to which segments on the various autosegmental tiers are associated. What we are assuming, therefore, to use Halle and Vergnaud's 1980 term, is a theory of "three-dimensional phonology." On this theory, defining a possible syllable for a language and the possible segmental associations to it expresses basic phonotactic generalizations about the language.[31] It is in terms of the syllable structure of the language that the segmental composition of the utterance is "organized."

In this book, the particulars of the segment-to-syllable association will usually not be relevant. We will therefore often represent the syllabic and segmental content of the sentence as a sequence of syllables and (written below them in standard orthography) a sequence of segments. For example:
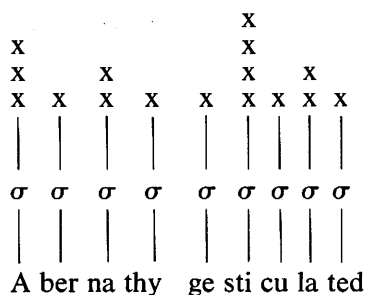
(1.8)

σ σ   σ   σ

A ber na thy

It has been widely recognized that tonal phenomena require an auto-segmental theory according to which tones are represented "supra-segmentally" as a sequence of elements on a tier separate from the segmental or the syllabic (see especially Goldsmith 1976a,b, Williams 1971, and Leben 1973, 1978, as well as the references in Fromkin 1978 and note 6). The tone-syllable relation may now be viewed as simply a special case of a more general set of relations between autosegmental tiers and the syllabic axis. In some languages individual morphemes or words may have their own tonal "melody"; in others the tonal melody may be defined only with respect to some larger domain, such as the, intonational phrase (see sections 5.3, 5.4); in still others the pitch contours may consist of both a "tonal" and an "intonational" contribution. But in any case, it is with respect to the syllabic composition of the utterance that the autosegmental tonal units are "realized." The tone-to-syllable association is governed by universal and language-particular well-formedness conditions. Among these are language-particular conditions stating how many tones may associate with a syllable, conditions that may in so doing make crucial reference to the internal structure of the syllable (see Clements and Ford 1979 and Halle and Vergnaud (in preparation)).

The alignment of the syllables of an utterance with a metrical grid, which we take here to be the representation of the prominence patterns of the utterance, may be viewed as a special instance of the association of syllables with an autosegmental tier (though we hasten to point out that other autosegmental tiers are apparently not hierarchical structures). Strictly speaking, then, the representation of the rhythmic structure of *Abernathy gesticulated* is as follows, where the syllable sequence mediates the relation between segments and the positions of the metrical grid.

(1.9)

```
                        x
x                       x
x       x               x   x
x   x   x   x       x   x   x   x   x
|   |   |   |       |   |   |   |   |

σ   σ   σ   σ       σ   σ   σ   σ   σ
|   |   |   |       |   |   |   |   |
A  ber na thy    ge sti cu la ted
```

(In discussing rhythmic structure, we will often omit the syllabic axis from the representation, for the sake of typographical simplicity, representing rhythmic structure simply as in (1.7).)

In sum, then, we see that the syllable has a crucial place in the theory of phonological representation and a crucial role in a theory defining the notion "possible phonological representation" for language in general. It is in terms of the syllable sequence of the core or axis that many rules governing the range of possible phonological representations of particular sentences are defined. In other words, the syllable has a central place in the mapping from surface syntactic representation to underlying phonological representation.

The syllable also has a major role in the phonological derivation, in governing the application of phonological rules. For example, the syllable serves as a *domain* for phonological rules; that is, it defines subsequences of the utterance within which phonological rules may be restricted to applying. It is now known that the notions "syllable-initial position," "syllable-final position," and "within the same syllable as," among others, are necessary in a theory of phonological rules, in order to express generalizations about a great number of phonological phenomena. Thus the syllabic structure of an utterance serves to define

some of its junctural properties, i.e., some of the relations between segments in sequence that are ultimately relevant to pronunciation.

Given the central role of syllable structure in phonological representation, it is important to determine how that organization into syllables may be governed by the syntactic representation of the sentence. (By *syntactic representation,* we refer to both word structure and phrase structure.) We offer two related proposals in this regard. The first is that morphemes are syllabified either as lexical items or in the course of the first cycle, and that there are language-particular well-formedness conditions, which we call *rules of basic syllable composition,* that either serve as redundancy rules on these lexical representations or introduce the initial syllabification, and define the notion "possible syllable" for the language. (This position is outlined in Selkirk 1978b, 1984.) The second is that, in cyclic fashion, this original syllable structure is rearranged at the limits of morphemes and successively higher units of morphological and syntactic structure. This amounts to saying that there is a partial "resyllabification" on successively higher cyclic domains (cf. Kiparsky 1979). Two sorts of *resyllabification* probably must be distinguished: resyllabification according to the basic syllable composition (BSC) rules of the language, and resyllabification according to some sort of universal principles. In English, for example, the (recursive) category level Root of word syntax is the domain of the syllabification and resyllabification according to the BSC rules of English. (The Root in English includes the so-called nonneutral affixes (Selkirk 1982).) On higher domains within the word—specifically, on the domain of the (recursive) Word category—resyllabification does not follow the BSC rules, but only certain restricted universal principles, such as the principle that makes a coda consonant the onset of a following onsetless syllable. And on domains higher than the word, it is debatable whether any resyllabification takes place at all.[32] English would thereby seem to contrast with French, for example, for which all word-internal cyclic domains are domains of resyllabification according to the BSC rules of the language, and for which resyllabification takes place, according to universal principles, on (certain) phrasal domains (Delattre 1940, Schane 1978).

In all cases we have encountered, BSC resyllabification is restricted to word-internal domains. Moreover, it seems quite likely that the possibilities of resyllabification *between words,* in phrasal contexts, are not defined directly with respect to syntactic structure (by syntactic-prosodic correspondence rules), but rather are determined by the syn-

tactic timing of the sentence—by the adjacency of syllables defined
with respect to the grid. If this were indeed shown to be true, then
phrasal resyllabification must probably be construed as a "late-level"
phenomenon, applying, along with rules of external sandhi, to a pho-
nological representation fully defined; it would not form part of the
syntax-phonology mapping.

   Because the limits of syllables may coincide with the limits of syn-
tactic constituents, and because some phonological rules may have
syllable structure domains, the application of phonological rules may
reflect the surface constituent structure of the sentence—but only indi-
rectly. Syllable structure thus provides one of the crucial, intermediate
links between syntax and phonology. It was originally our intention to
make the investigation of the syntactic structure–syllable structure
relation—in particular, the study of the syntax of resyllabification—an
integral part of the present work. It appears now that an adequate
treatment of this extremely important question is beyond the scope of
the book, in part because it would require more research than we have
carried out so far. And so it is with regret that we leave a general con-
sideration of this topic for a later time and place.

1.2.4 Prosodic Structure: Suprasyllabic Constituents

In previous work, we have assumed that a fairly rich hierarchy of pro-
sodic constituents or prosodic categories forms part of phonological
representation (see, for example, Selkirk 1978c, 1980a,b, 1981a). We
have suggested that the hierarchy for English includes at least the fol-
lowing categories:[33]

(1.10)
intonational phrase (IP)
phonological phrase (PhP)
prosodic word (Wd)
foot (Ft)
syllable (Syl)

We have proposed that a category of level i in the hierarchy im-
mediately dominates a (sequence of) categories of level $i-1$ (Selkirk
1981a). (Assuming *syllable* to be level 1, the others will be levels
2, ..., n.) We will call this the *strict layer hypothesis,* and will take it as
a useful working hypothesis here. In earlier work, we presumed that,
like the syllable, each of the suprasyllabic units in this hierarchy had
the potential for playing a role in the description of the phonotactics of

words and/or phrases (including their stress patterns) and in the description of tonal patterns and the domains of application of phonological rules. It is necessary now, we believe, to reassess the claims for the existence of those suprasyllabic prosodic constituents, for it is clear that some of the phonological phenomena that were thought to provide motivation for these higher units of structure are better explained in terms of the metrical grid alignment of the sentence. Some categories will disappear entirely from the prosodic structure repertoire; others will be given a much reduced role in phonological description, once the role of rhythmic structure in phonology is fully understood. For each of the units listed in (1.10), we will briefly review what is at stake.

1.2.4.1 Intonational phrase   This unit corresponds to a span of the sentence associated with a characteristic intonational contour or melody (see section 5.4). A sentence may correspond to one or more intonational phrases. An intonational phrase typically contains material belonging to a sequence of words and/or phrases, and it is not necessarily isomorphic to any constituent of syntactic structure (Selkirk 1978c). There are two possible intonational phrasings for *Abernathy gesticulated:*

(1.11)
$_{IP}$(Abernathy gesticulated)$_{IP}$
$_{IP}$(Abernathy)$_{IP}$ $_{IP}$(gesticulated)$_{IP}$

The existence of the intonational phrase is motivated primarily by the necessity of defining intonational contours with respect to some unit of representation that is both larger than the word and variable in extent. That unit cannot be a syntactic one, because the syntactic sequence with which an intonational contour is associated may not be a constituent of syntactic structure. And the metrical grid alignment of the sentence defines no such unit in the representation. In languages with characteristic intonational contours, then, we are led to posit intonational phrases as part of the prosodic constituent structure of phonological representation. (See section 5.4.)

Studies in the generative tradition have usually held that the surface syntactic structure of a sentence determines, in some fashion or other, the division of the sentence into intonational phrases (Downing 1970, 1973, Bing 1979a,b, Selkirk 1978c, 1980b, 1981a). Here we reject this idea, in favor of one that has its roots in earlier work on the topic: the idea that the definition of what may constitute an intonational phrase is

essentially semantic in character. For Halliday 1967a, for example, intonational phrases are units of "information structure."

Our specific hypothesis, defended in chapter 5, is that the immediate constituents of an intonational phrase must bear either a head-argument relation or a head–(restrictive) modifier relation to each other. This hypothesis may be seen as an attempt to spell out what it means to say that an intonational phrase is a "sense unit." As an implementation of this basic hypothesis, we suggest that the intonational phrasing of a sentence is assigned (freely) to the surface structure of the sentence, and that particular phrasings are subject to a well-formedness condition (or filter) that encodes the aforementioned constraints on the semantic relations obtaining among the constituents within the successive intonational phrases. This well-formedness condition, which we will call the *Sense Unit Condition,* may be stated either on an intonationally phrased surface structure or on (intonationally phrased) logical form, depending on where the semantically relevant information is considered to be available.[34] Thus the statement of the possible relations between syntactic constituent structure and intonational phrasing—the syntactic-prosodic correspondence rule for the intonational phrase—is quite trivial. It need merely be stated that a (highest) sentence corresponds to a sequence of one or more intonational phrases. Any general further constraints on the constituent membership of intonational phrases are claimed to follow from the semantically based Sense Unit Condition.

We will show in chapter 5 that the free assignment of intonational phrasing to a sentence and the subjecting of this phrasing to the Sense Unit Condition are entirely consistent with the approach that must be taken to assigning intonational contours to the sentence. We will argue that the tonal elements making up the pitch contour of the intonational phrase are assigned directly (and freely) to surface syntactic structure, and that it is on the basis of this assignment that the essentially semantic properties of the focus of the sentence are defined.

In our investigation of phrasal rhythm, we will show that the intonational phrase serves as a domain with respect to which patterns of rhythmic prominence are defined. It has also been thought to serve as a characteristic domain of rules of segmental phonology, especially rules of external sandhi.[35] However, caution is needed here in assessing the role of the intonational phrase with respect to phonological rules of sandhi. The limits of intonational phrases often coincide with substantial pauses, which our theory represents as silent positions in the metri-

cal grid. Thus it may be that juncture-sensitive rules that have been thought to have the intonational phrase as their domain are simply rules whose application is governed by the adjacency of segments and/or syllables defined with respect to the metrical grid.

In general, it will be necessary to adjudicate the respective roles of the metrical grid and of prosodic constituent structure in characterizing the junctural properties that are relevant to the application of phonological rules. We will take the position that a phonological rule may be sensitive to either of the two types of junctural representation; that is, it will be sensitive either to prosodic structure domains or to adjacency defined on the grid. Because segmental phonology above the level of the word is (lamentably) still grossly underinvestigated, it is not possible at this point to know whether rules appealing to the same sorts of junctural information share other properties as well.

1.2.4.2 Phonological Phrase   Let us use the term *phonological phrase* for any level of prosodic constituent structure that may include one or more major category words.[36] In principle, the proposed analysis will allow for the possibility that language may exhibit more than one level of phonological phrase, in which case finer terminological distinctions can be made: PhP^1, PhP^2, ..., PhP^n. With this terminology then, an intonational phrase is a special case of a phonological phrase, one that is associated with a characteristic tonal contour and that has an important function in representing the "information structure" of the sentence. The unit *utterance,* if it existed, would also be a phonological phrase in this sense.

The term *phonological phrase* has been used to apply to a (putative) level of English prosodic structure falling between the intonational phrase and the prosodic word (see Selkirk 1978c, 1981a). The English phonological phrase has been thought to have a role in the timing of the utterance, with an influence both on its rhythmic properties (Selkirk 1978c) and on its division into pauses (Gee and Grosjean 1981). We now think that the existence of this unit in English is highly suspect, for syntactic timing (silent positions in the grid) gives a representation of the disjuncture or separation between syllables that is more appropriate to the description of such rhythmic phenomena (see chapters 4 and 6). Indeed, we would now explicitly deny that the existence of a level of phonological phrase below that of the intonational phrase is well motivated in English.

1.2.4.3 Prosodic Word   A variety of linguists have thought it necessary to isolate a roughly word-sized unit of phonological representation.[37] Such a unit could serve to define such phonologically relevant notions as "word-initial," "word-final," and "word-internal," and it would seem to be required particularly when the words of the sentence defined in syntactic terms fail to correspond exactly to the "words" playing a role in the phonology. In Selkirk 1980a,b we suggested that there exists a unit of prosodic constituent structure, the *prosodic word,* for English, Sanskrit, and other languages. We claimed that it was within the (prosodic) word that prominence relations particular to word-sized units were defined—that is, that the unit "(prosodic) word" played a role in metrical theory in permitting the characterization of "main word stress." (See also Halle and Vergnaud 1979 and Hayes 1980.) In addition we claimed that the unit that serves to define main word stress is the unit in terms of which the notion "word" relevant to the application of phonological rules is defined. The hypothesis was that the *domains* for both principles governing prominence relations and rules for segmental phonology systematically coincide.

Given a metrical grid approach to characterizing word stress, of course, there is no obvious motivation for a phonological constituent Word for the realm of word-internal prominence relations (see chapter 3). "Main word stress" is a metrical grid alignment at a certain level of the grid that is established within a domain characterized in syntactic terms. As for junctural notions like "word-internal," "word-initial," and "word-final," we submit that they may be expressed, and more appropriately so, either in terms of adjacency defined with respect to the grid, or directly with respect to word-syntactic structure. Our proposed theory of syntactic timing (see chapter 6) has the result that there are no silent grid positions inside the syntactic word; hence, syllables internal to the same word are strictly adjacent with respect to the metrical grid. We suggest, then, that "grid-adjacent" may substitute for at least some instances of the notion "word-internal." Between words, the theory of syntactic timing assigns varying degrees of rhythmic disjuncture—that is, varying degrees of closeness with respect to the grid—in that differing numbers of silent grid positions are assigned between words in the representation, depending on the syntactic constituent structure of the sentence. Thus one could claim that certain appeals to the notions "word-initial" and "word-final" should be supplanted by appeals to "lack of grid-adjacency to what precedes" and "lack of grid-adjacency to what follows." Other appeals to the former

notions might well be appeals to syntactic structure itself. Thus we see little if any need for a prosodic word in phonological description.

1.2.4.4 Foot   The foot is a suprasyllabic unit, usually smaller in size than the word, that has played a central role in the description of stress patterns in the framework of "metrical phonology." It has rendered service in representing the distinction between stressed and stressless syllables, and as a device for computing the distribution of stressed and stressless syllables within specified domains. In this function, as Prince 1983 argues quite effectively, the foot has been supplanted by the metrical grid theory of stress. It is important to note, moreover, that there is relatively little evidence that the foot itself serves as a domain for phonological rules. Most alleged foot-sensitive rules can be easily and with no loss of generalization recast as rules sensitive to the stressed-stressless distinction.[38] In the present theory such rules would be recast as rules sensitive to the metrical grid alignment of syllables. We hypothesize, therefore, that there is no prosodic constituent *foot*.

   The particular claim we are making about the prosodic constituent structure of phonological representations in English, then, is that the phonological phrase, the prosodic word, and the foot are not units in the hierarchy, but that the syllable and the intonational phrase are. With the syllable sequence as the lowest layer, the prosodic constituent structure of *Abernathy gesticulated* is then either (1.12a) or (1.12b).

(1.12)

a.   IP   IP   b.   IP

σ σ σ σ σ σ σ σ σ σ σ σ σ σ σ σ
Abernathy gesticulated   Abernathy gesticulated
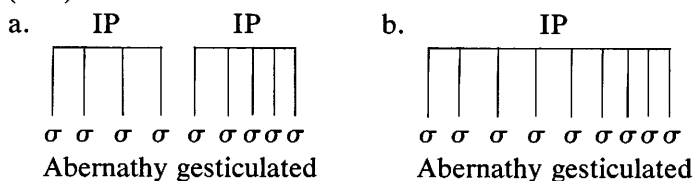
1.3 The Mapping between Syntax and Phonology

According to the theory sketched thus far, there are three principal stages in the mapping from syntax to phonology. The first is *surface syntactic structure*, S_n (figure 1.1). (Recall from section 1.1 that this includes a sequence of word-level phonological representations.) The second is surface structure *cum* intonational structure, or *intonated surface structure*, S'_n. (The term *intonational structure* designates the intonational phrasing of the sentence, the (autosegmentally represented)

tonal contours of the intonational phrases, and the assignment of certain of these tonal elements to particular constituents of surface structure; see chapter 5.) The third is surface structure *cum* intonational structure *cum* metrical grid, or *intonated and rhythmated surface structure*. It may be called S_n'' or, more appropriately perhaps, P_1. This is what should be thought of as the underlying phonological representation of the sentence.

S_n'', or P_1, is a representation in which the hierarchical aspects of phonological representation are essentially fully established (except, possibly, for some phrasal resyllabification). It includes as well all the segmental aspects of phonological representation, represented on the various autosegmental tiers. This representation P_1 is mapped by what we have been calling *phonological rules* of the sentence into a surface phonetic representation P_n that shares many properties with P_1 (but not all).[39] We will have little to say about the rules participating in the derivation from P_1 to P_n in what follows. But it does seem worth pointing out (again) that this class of rules appears to be defined only in terms of those aspects of the representation that are strictly phonological. In the unmarked case at least, they do not appear to be sensitive to syntactic structure. It also appears that this class of rules does not apply cyclically. Both these characteristics of phonological rules may be seen as reflections of a single general condition: that phonological rules are blind to syntactic structure. In the general case, then, phonological representation, richly structured itself, mediates between syntactic structure and the phonological rules that ultimately specify the details of the phonetic realization of the sentence. Cases where phonological rules may appear to appeal directly to surface syntactic structure are highly marked and may even be surface suppletions.

The representation S_n', the intonated surface structure, is one whose properties we are only beginning to investigate here. The mapping from surface structure S_n to S_n' is quite trivial, it seems. As mentioned earlier, we adopt the hypothesis that intonational phrasing is freely assigned to surface structure, the only structural restriction being that the entire sentence be parsed into a sequence of one or more nonoverlapping intonational phrases. (An additional, independent hypothesis is that the tonal elements constituting the intonational contours of the phrases are also freely assigned in this mapping from S_n to S_n'.) We claim that it is at the level S_n' (or a mapping of it into logical form) that certain well-formedness conditions governing the intonational structure-meaning relation are defined. These include the Sense Unit Con-

dition, governing the semantic composition of intonational phrases. (They include as well the focus rules, which govern the relation between the intonational contour of the sentence and its focus structure.) These conditions, and this general approach to intonation, are defended in chapter 5. It may turn out to be the case that the intonated surface structure, S'_n, has quite a large role in grammatical description, as the level at which generalizations concerning the relation between prosody and word order, or prosody and ellipsis, are to be expressed. Unfortunately, we cannot explore these possibilities here.

The mapping from the intonated surface structure S'_n to the intonated-and-rhythmated surface structure S''_n (= P_1) is one of the central concerns of this work. A theory of that mapping is a theory of how the various aspects of S'_n—the syntactic labeled bracketing, the organization into syllables, the intonational phrasing, and the assignment of focus-relevant tonal elements—determine a rhythmic structure for the sentence. The theory defended here is that four components are involved in this mapping: text-to-grid alignment, grid euphony, syntactic timing, and destressing. The rules of two components, text-to-grid alignment and syntactic timing, appeal directly to the syntactic structure of the sentence, as well as to its intonational phrasing (chapters 2–6). Rules of destressing and text-to-grid alignment appeal directly to the syllable composition of the sentence (chapters 2, 3, and 7). Finally, text-to-grid alignment takes into account as well the tonal associations of the syllables (chapters 4 and 5).

A significant result of these investigations into the S'_n-P_1 mapping is the discovery that the rules involved apply in *cyclic fashion* (see sections 3.4, 4.4, 7.2.2, and 8.2). It is important to note that, given the present conceptual framework, these cyclic rules are not rules "of the phonology"; rather, they are rules that collectively *construct* a phonological representation on the basis of syntactic representation. We may speculate that, in sentence grammar at least, the cycle is a principle governing only the *interpretation* of syntactic representation *as a* phonological representation, rather than a principle governing the relation of one phonological representation (syntactically structured) to another. (Of course, restricting the cycle to the mapping constructing the underlying phonological representation P_1 would be unnecessary if syntactic structure were simply "deleted" at P_1, a possibility certainly worth considering.)

We sum up this view of the organization of the grammar in figure 1.2. Though still not complete in all details, this diagram is a fleshed-out
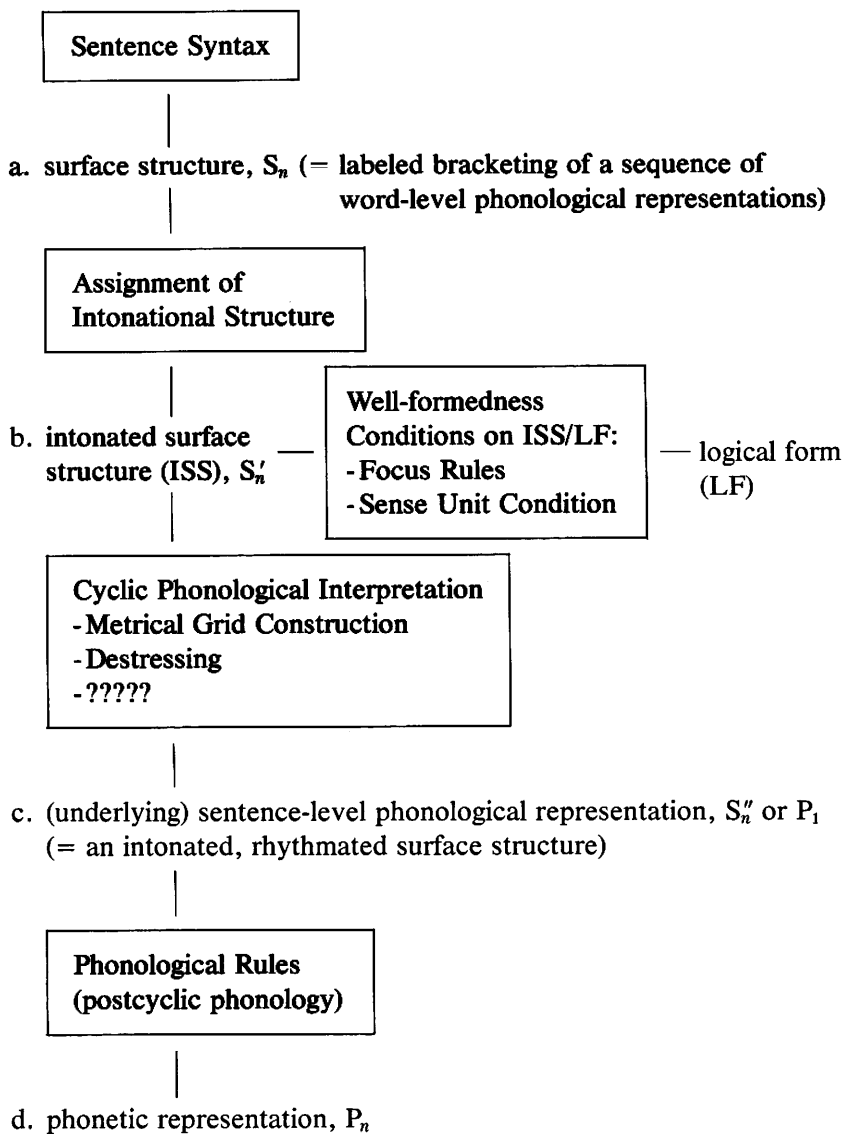
```
┌─────────────────────────┐
│    Sentence Syntax      │
└─────────────────────────┘
              │
```

a. surface structure, S_n (= labeled bracketing of a sequence of
                          word-level phonological representations)
              │

```
┌─────────────────────────┐
│   Assignment of         │
│   Intonational Structure│
└─────────────────────────┘
              │
```

```
                    ┌──────────────────────────┐
                    │  Well-formedness         │
b. intoned surface  │  Conditions on ISS/LF:   │ ── logical form
   structure (ISS), $S_n'$  │  - Focus Rules           │    (LF)
                    │  - Sense Unit Condition  │
                    └──────────────────────────┘
              │
```

```
┌──────────────────────────────────────┐
│  Cyclic Phonological Interpretation   │
│  - Metrical Grid Construction         │
│  - Destressing                        │
│  - ?????                              │
└──────────────────────────────────────┘
              │
```

c. (underlying) sentence-level phonological representation, S_n'' or P_1
   (= an intonated, rhythmated surface structure)
              │

```
┌─────────────────────────┐
│   Phonological Rules    │
│   (postcyclic phonology)│
└─────────────────────────┘
              │
```

d. phonetic representation, P_n

Figure 1.2

version of figure 1.1, constructed by adding a theory of the mapping between the surface syntactic representation S_n and the (underlying) phonological representation P_1 of the sentence. The general theory of the organization of the grammar being assumed here is therefore to be understood as a revision and extension of the standard theory of generative phonology.

It is worth mentioning again that, as should be clear from figure 1.2, our model of the syntax-phonology relation in sentence grammar is based on the assumption that the output of word grammar forms part of the surface syntactic representation to which this mapping applies. Included in this surface structure is a concatenation of the derived phonological representations of the "word level." This notion that the phonological output of a separate word grammar forms a discrete level of representation in the overall grammar (the "word level") is suggested by Kiparsky 1983a,b, Mohanan 1982, and others developing the framework of so-called lexical phonology and morphology. Although we are not committed to certain other details of that theory (see sections 3.1, 3.4, 8.2, 8.3), this assumption seems a useful one to adopt.