ON PRSODIC STRUCTURE AND ITS
RELATION TO SYNTACTIC STRUCTURE*

by

Elisabeth O. Selkirk

UNIVERSITY OF MASSACHUSETTS

Reproduced by the
Indiana University Linguistics Club
310 Lindley Hall
Bloomington, Indiana 47405
November, 1980
ON PROSODIC STRUCTURE AND ITS RELATION TO SYNTACTIC STRUCTURE

In this paper a theory of phonological representation is presented which departs significantly from the standard generative theory (such as that outlined in Chomsky and Halle's *The Sound Patterns of English*) in imputing a suprasegmental, hierarchically arranged organization to the utterance, not a simple linear arrangement of segments and boundaries. This hierarchical organization will be referred to as PROSODIC STRUCTURE. The purpose of this paper is to give an indication of (i) the nature of this prosodic structure and (ii) the relation between it and syntactic structure. A fuller treatment of these same matters is given in Selkirk (forthcoming).

Answers are proposed to a number of basic questions: What are the units of prosodic structure, i.e. the linguistically significant sub-parts? What are the principles governing the internal structure and prominence relations within the different prosodic units? How do these prosodic units, or constituents, relate to constituents of syntactic structure? Evidence supporting these answers is adduced from a broad range of phonological/phonetic phenomena - including phonotactics, stress and rhythm, and the rules of segmental phonology. A further question must be asked throughout, one which represents a challenge to linguists and workers in speech production and perception alike: What is the relation between these prosodic units - units of linguistic analysis and the units of production and perception? It is suggested that they may be the same.

The theory presented here is an elaboration, or revision, of the theory of Liberman (1975) and Liberman and Prince (1977), who propose that the phonological representation of an utterance consists of a 'metrical' tree which is binary branching and between whose nodes the prominence relation strong/weak (s/w) is defined, and who argue that, above the level of the word, the branching of phonological representation is identical to that of syntactic representation as below:

```
  S
  / 
 W  S
 / 
N  W
```

Maria prefers the tango
It will be argued below that this conception of the suprasegmental representation is too impoverished, and that it is necessary to posit PROSODIC CATEGORIES, i.e. specific differentiated sub-units of prosodic structure, which 'label' the nodes of the tree. These prosodic categories are the syllable, the foot, the prosodic word, the phonological phrase, the intonational phrase and the utterance. It will also be argued that there is not an isomorphism between prosodic structure and syntactic structure, rather that prosodic structure is an entity distinct, and that a mapping of a non-trivial sort must be defined between it and syntactic structure.

1.0 In this first section, the prosodic units at the level of the syntactic word or lower are investigated. These include the syllable (ο), and foot (ώ), and the prosodic word (ω). Each is shown to have its own hierarchically arranged internal structure, as is exemplified in the representation of προσωποποιητική [prosopopoihtikí].

The reasons generally given in support of the syllable as a construct of phonological theory are threefold. First of all, it can be argued that the most general and explanatory statement of phonotactic constraints in a language can be made only by reference to the syllable structure of an utterance. Second, it can be argued that only via the syllable can one give the proper characterization of the domain of application of a wide range of rules of segmental phonology. And third, it can be argued that an adequate treatment of suprasegmental phenomena such as stress and tone requires that segments be grouped into units which are the size of the syllable. The same three reasons leading to the postulation of the syllable can be shown to motivate the existence of privileged groupings of segments within the syllable which must be thought of as constituent-like linguistic units themselves. The conception of the syllable that emerges, therefore, is one of a hierarchical unit; an internally structured tree quite analogous to a tree representing syntactic structure. (See also Pike (1967), Fudge (1969) and McCarthy (1972), among others.) Numerous scholars have argued for the unit syllable as a part of phonological representation on the grounds that the notion of possible word or morpheme in a language is definable, in large part, on the basis of syllables. To use an example from Kahn (1976:57ff) άκτιν is not a possible 'word' of English. Assuming a syllabic analysis of 'word' in English, the non-existence of άκτιν is predicted simply from the fact that it cannot be 'parsed' as a sequence of well-formed English syllables. (st cannot close a syllable, nor can st begin one.) The matter is not so simple in a theory not making use of the syllable, as Kahn demonstrates; the statements required to rule out such a form are cumbersome, and fail to express the appropriate generalizations. A theory incorporating the syllable is therefore preferable, the argument goes, for it allows for a natural and explanatory statement of the phonotactics of words.

In a similar fashion, the phonotactics of the syllable itself lead to the postulation of units of representation within the syllable, and thus to the conclusion that the syllable has an internal immediate constituent structure. In our view, the phonological representation of an English monosyllabic word like δύναμις is structured as in

There is a first major bipartite division of the syllable -- into ONSET (the initial consonant cluster) and RHYME (the rest). The rhyme in turn divides into two parts -- the PEAK (containing the syllabic nucleus) and CODA (the final consonant cluster) (Fudge 1969). Of course not all syllables in English are so rich in internal structure: a car [kær] has only a simple (non-branching) onset, a complex peak, and no coda at all; fat [fæt] has a simple onset, a simple peak and a simple coda; age [æ] has a complex peak, but no onset and no coda.

We will not defend this particular analysis of English here, but will address the question of syllable-internal structure in somewhat more general fashion. Pike (1967) has argued that it is a universal of syllable composition that a constituent break exists between a syllable nucleus and its margins (our 'onset' and 'coda'). The argument for
The possibility of substitution of one phoneme for another in a particular slot in the margin, for example, is likely to be more dependent upon the particular phonemes manifesting other slots in that margin than it is by the particular phonemes manifesting the nucleus of such syllables. I.e., if a formula CCV is manifested by /a/ in the first consonant slot, and the nucleus slot is filled by the phoneme /a/, the list of phonemes which fill the second consonant slot are more likely to be controlled by the presence of the /a/ than they are by the presence of the /a/ - e.g., they may be limited to voiceless consonants after the voiceless /f/.

Such considerations indicate that a closer relationship exists between the two consonants than exists between either consonant and the vowel. (Pike 1967:386-7)

The claim is that the likelihood of the existence of phonotactic constraints between two position slots in the syllable (as well as the strength of inviolability of those constraints, one might add) is a reflection of the immediate constituent (IC) structure relation between the two slots. The more closely related structurally (in the obvious sense), the more subject to phonotactic constraints the two position slots are. (We will call this the IC principle of phonotactics.) According to this principle, therefore, onset, peak and coda are units within which the tightest phonotactic constraints obtain.

The grouping of the peak and coda into a constituent is advocated as a universal of syllable composition by Kuryłowicz (1947), one reason for this being similar to Pike's. The claim made is that cooccurrence restrictions between peak and coda are always more likely to exist (and indeed are quite common) than are restrictions between either peak or coda and the onset. The explanation offered is that the former two comprise a constituent.

As any detailed analysis of the phonotactics of the English syllable shows (cf. Bloomfield 1933, Fudge 1969), it is within the onset, peak, and coda that the strongest collocational restrictions obtain. By contrast, there are no phonotactic restrictions at all for the language which involve onset and peak, for example. The existence of this array of restrictions follows from a single principle, the IC principle of phonotactics, if one assumes an IC analysis of the syllable.

One might next want to ask whether the labels syllable, onset, rhyme, peak and coda, which we have so far used merely as descriptive terms should be assigned to their respective nodes in a representation. That is, should the earlier representation of *flowne* be recast as follows?

![Diagram of syllable structure](image)

The preliminary answer we have to offer is that while the prosodic category 'syllable' is required in linguistic theory and the corresponding label necessary in the tree, support for naming its internal constituents in this way is currently lacking. The sort of evidence pointing to the need for the 'syllable' label will be reviewed below, where it will be seen that the label is necessary in order to identify the syllable as a unit distinct from any other branching structure (be it 'larger' or 'smaller' than the syllable).

Another suggestion concerning node-labelling within the syllable can be entertained. It builds on the observation that in a binary branching constituent of the syllable, one member always tends to the 'weaker' than the other. Pike and Pike (1947) described this relation as the 'subordination' of one to the other. Extending the theory of hierarchical prosodic structure of Liberman (1975) and Liberman and Prince (1977) to the syllable, this relation can be given a formal representation, by labelling the 'subordinate' node w and the other s. With nodes labelled in this fashion, the representation of *fliawn* would take on the shape below:

![Diagram of syllable structure](image)

The peak is of course strong (s), i.e., more sonorous, than the onset. And within each of the other constituents, the s has been assigned to the more sonorant element. This approach incorporates the suggestion made by Pike (1967:387ff.) and McCarthy (1977) that the assignment of w (= 'subordinate' status) vs. s can be made, at least partly, on the
basis of the relative ranking of the two segments (or constituents) on a universally defined sonority hierarchy.

The arguments in favor of a hierarchical conception of the syllable are not limited to phonotactics. It has been shown that an internal structural analysis makes available a superior treatment of phonological processes in language. For example, Pike (1947:142) argues for the nucleus (=peak) on the grounds that it serves as a unit with respect to suprasegmental phenomena such as pitch and stress. Pike and Pike (1947) adduce evidence from numerous phonological rules, suprasegmental and segmental, for a particular IC analysis of the syllable in Mazateco. In addition, as a point of favor of the rhyme constituent, Kuryłowicz (1947), Pike (1967:391) and Newman (1972:203) have observed that it makes possible a unified treatment of the heavy-light syllable distinction which plays such an important part in stress systems. The light syllable CV can be characterized as one whose rhyme is simple, non-branching, while the heavy syllable, be it CVC or CVV (a long vowel) is one with a complex, or branching, rhyme. Other phonological, or shall we say phonemic, phenomena such as duration and closeness of transition between segments might also be taken as revealing of the immediate constituent structure of the syllable. For example, work on the duration of English vowel-glide combinations, which in our analysis together comprise the peak of the syllable, seems to show that they function as a unit on a par with simple vowels, which in our theory are also peaks. Lehiste and Peterson (1960) report that the length of simple vowels and complex vowels (branching peaks) is affected in the same way by the voiceless/voiced property of a following obstruent: vowel length before voiceless vs. voiced is in the ratio of 2:3 in both cases. Fujimura and Lovins (1977) report on additional very promising durational evidence suggesting that vowel and glide form a unit. Evidence such as that provided by Chen (1970), who claims that there is a constancy (approximate) in the length of vowel plus stop combinations, could be taken as supporting the existence of the rhyme. According to Chen, a lengthening of the vowel (as before voiceless stops) coincides with a shortening of the consonant. That is, one could say that within a constituent like the rhyme the duration of one element is adjusted in function of another. Other areas which bear examination in the light of syllable-internal structure are nasalization, voicing assimilation, and so on.

As for the higher unit SYLLABLE itself, it has been increasingly acknowledged that phonological rules may be sensitive to the syllable structure of an utterance. For example, the realization of particular allophones of English sound types may depend on the position of those segments within the syllable and thus, an investigation of such sorts of allophonic variation may give a key to syllabification in the language. This is especially true of the processes affecting the realization of the voiceless stops, p, t, k, and especially of the phenomenon of aspiration. A wide range of workers in the field have observed that voiceless stops are as-

pirated in syllable-initial position in English. Formalized, the rule could read as follows:

\[
\text{Aspiration} \\
\begin{align*}
\text{[+cons] } & \rightarrow \text{ [+aspirated] } / \text{ [..] }_0 \\
\text{[+cont] } & \rightarrow \text{ [+voice] }
\end{align*}
\]

This aspiration rule is able to capture in quite simple fashion the fact that aspiration is found in the following types of contexts:

a. initial in a word: Toronto, aesthetic, polynomy

b. before a stressed vowel, except if s precedes: hotel, repair, recent vs. estate, disaster, advantage

c. before a sonorant plus a stressed vowel, except if s precedes or it is a l followed by l (or any other combinations ruled out as an onset): stromes, apply, aggression, blonde, improve, beatific, acquaint vs. destroy, display, diagnose, or Atlantic, etc.

d. before a sonorant plus stressless vowel, except if s precedes, or if it is i followed by l, etc.: artigo, acromy, country, simplified, implication, escalation, April vs. mister, misgiant, explicit or atelier, rija, etc.

These are all contexts in which, according to normal syllabification, the stops are syllable-initial. An array of facts such as this provides a rather striking example of the advantages of a phonological theory which gives representation to the syllable, and allows the syllable to serve as a domain for phonological rules, without the syllable to refer to, the statement of the rule would be ungainly in the extreme, having to mention explicitly each of the contexts enumerated above. Processes such as these constitute evidence for labelling as a syllable (c) the node of a prosodic structure tree which corresponds to the prosodic unit syllable: the label 'syllable' identifies the appropriate domain of such rules within the larger prosodic tree.

It must be recognized that the question of what the syllable looks like in phonological representation is a distinct one, logically, from the question of how the grammar of a language is to give expression to the notion like phonological rules. In particular, the notion of syllable post, peak, and rhyme. That is, individual representations of syllables do not state generalizations about syllable structure any more than individual noun phrases in a syntactic representation state gen-
izations concerning the notion 'possible noun phrase' in a language. A grammar must therefore provide for some statement of the notion 'possible syllable of L', this statement being distinct from any phonological representation of the language. We will suppose that for each language this statement is in the form of a template and an accompanying set of phonotactic constraints somewhat in the spirit of Fudge (1969), and Hooper (1976), but with differences. These together specify all the possible syllable types of the language, and can be thought of as serving as well-formedness conditions on the syllabic structure of the phonological representations of a language.

Suppose, for example, that the syllable template for English looks something like this:

```
ENGLISH SYLLABLE TEMPLATE

-son (+son) +cons (-son) +syll

-syll
```

The function of the template is to encode the gross characteristics of syllable structure: (i) the composition of the syllable in terms of segment types identified by the major class features [syllabic], [+sonorant] [+consonantal], (ii) the order of these segment types within the syllable, (iii) the structural relations between the segment types (defined in IC terms), and (iv) the optionality of segments or groups of segments (= constituents) within the syllable. Sufficiently to say here that the role of the template is to contribute to defining the well-formedness of the syllable structure of particular phonological representations. A necessary condition for the well-formedness of a representation is that, within a syntactic domain specified for the language, the syntactic structure of the representation be NON-DISTINCT from the template.

The template on its own does not give a characterization of English syllable structure that is sufficiently restrictive, though. Formulated as above, it allows for more types than are even evidenced in English.

For example, in the onset, if the first consonant is a labial then it may not be followed by $w$, or if it is an $s$, it may not be followed by $r$. And in the peak, for example, only a limited variety of diphthongs are available: $i$ is not possible after any back vowel except $e$ and $a$. These further restrictions do not find a natural expression in templates like these. Following Fudge (1969), we acknowledge that another device is required in the grammar to express such phonotactic constraints. Fudge calls them collocational restrictions, and gives them the form of an implication on the order of "if a second position in onset is $w$, then first position is not [+labial]." (Alternatively, one could think of them as filters on the output of an 'overgenerating or overly permissive template'. They might have the form, eg. [+labial] $w$. We will take no firm position on this matter here.) Clearly, then, a second condition on the well-formedness of the syllable structure of a representation is that it not be ruled out by the collocational restrictions of the language.

An additional condition must be imposed in defining the well-formedness of the syllable structure of an utterance. Parsability with respect to the templates and satisfaction of the collocational restrictions is not enough. Consider the simple case of a word like allow. There are two syllable structure analyses of the word that are consistent with the English template: $\lambda.law$ or $\lambda.aw$. (Note that here we adopt the convenience of employing the period to indicate the limits of syllables, since the internal bracketing is irrelevant to our present point.) But only the first is correct (as can be seen from the fact that $l$ is pronounced with its light non-velarized, syllable-initial version). To cite a few additional examples, the syllabifications pro.strate, ao.treat, ar.can are clearly the only correct ones, though pro.trate, pro.strate, ao.treat, and ar.can would all be permissible, given the template and collocational restrictions of English. In general, when a medial consonant or consonant cluster may be analyzed as either a codas or an onset according to syllabification principles, it is the onset analysis which prevails. A number of scholars have proposed that in syllabification (in our terms, the determination of the well-formedness of some syllabic structure) the following universal principle be respected:

Maximal Syllable Onset Principle

In the syllable structure of an utterance, the onsets of syllables are maximized, in conformance with the principles of basic syllable composition of the language. (cf. Hoard (1971), Hooper (1976), Kahn (1976)).

The final question to be addressed in a linguistic description of syllabification concerns the relation between syntactic structure and syllable structure. It can be shown that for any language, syllabification will have a particular characteristic syntactic domain. In English, that domain is smaller than the syntactic word. As is well known, the so-called neutral
Affixes of English are phonologically set off in a way different from the non-neutral affixes. Compare the pronunciation of *rhythm* [rɪðm] before the neutral affix -y, as in *rhythm* [rɪðm] to that before the non-neutral affix -ic, as in *rhythm* [rɪθmɪk]. We would claim that neutral affixes are in a separate domain of syllabification from the stems to which they are attached, so that the limit between a neutral affix and a stem always coincides with the limits of two syllables. A non-neutral affix, on the other hand, is always contained within the same domain of syllabification as the root to which it is attached. Compare the representations below, where the curly braces delimit domains of syllabification:

a. rhythm  b. rhythm - y  c. rhythm - ic

When, as in (c), the m is in the same domain as the vowel-initial suffix, it syllabifies with it, and is not syllabic itself. But when the m is not followed by a vowel within the same domain, as in (a) and (b), it may serve itself as the nucleus of the syllable, and be realized as a 'syllabic' sound.

Assuming the theory of English morphological structure that is developed in Selkirk (forthcoming), the domain of syllabification can be given a simple straightforward characterization. It is the SIMPLE (non-branching) STEM, and any stem affixes. This is illustrated in the figures below.

Because neutral affixes can only be adjuncts to a stem, they never form part of a simple stem and hence form a distinct syllabification domain. Non-neutral affixes are root affixes, however, and as such are included within the simple stem, and hence are in a same syllabification domain with their root.

To sum up this approach to the syllable, then, it has been argued that the syllable is a unit with a place in the prosodic hierarchy, and that it has its own internal hierarchical structure. Evidence has been reviewed that the syllable is a domain for the application of phonological rules. And, finally, it has been shown that a syntactic domain for syllabification must be specified in the grammar of a language.

We turn next to consideration of the higher-order units, foot (ζ) and prosodic word (ω). Our discussion here will revolve primarily around the facts from English, and draws heavily on the Liberman and Prince treatment of English word stress. It differs from the Liberman and Prince approach, however, in positing the existence of prosodic categories like ζ and ω, giving them a role in 'stress placement.'

The foot is the prosodic unit immediately superordinate to the syllable. The definition of just what constitutes the foot may differ from language to language. In English, we claim, the notion possible foot can be defined by a set of templates:

**English Feet**

**The Basic Feet:**

(i) ζ  

(e.g., probate)

(V G = tense vowel or diphthong)
In our view, these templates have a role in the grammar similar to that of the syllable template — they function as well-formedness conditions on underlying representations. That is, given a particular syntactic domain (in English it is the simple stem), all syllables on that domain must be 'footed' into higher order structures that are parsable by the set of foot templates.

If a syllable is itself a foot, it will be 'stressed', cf. *probate*, *gymnast*. A syllable will also be 'stressed' if it is the strong element of a bisyllabic foot, cf. *Amaz*. An 'unstressed' syllable is one which is weak, and not also a foot. The claim is that being a foot will give an individual syllable a status in prosodic structure that is distinct from that of a syllable which is not a foot, that a syllable will be phonetically 'interpreted' differently depending on its footedness. In this way, we can explain, without appealing to a feature [stress], the examples that posed problems for Liberman and Prince: *modest* vs. *gymnast*. In a Liberman and Prince representation both words have identical prosodic structure: \( \sigma^w \). The presence or absence of secondary stress on the second syllable cannot be attributed to the prosodic tree, and they therefore (re-)introduce the feature [stress] to provide the distinction. Compare their representation with ours.

Positing the prosodic category \( \Sigma \), defining it for its templates like those above, requiring that all sequences of syllables on the same domain group into feet according to the templates and assuming that \( \Sigma \) has the phonetic interpretation just described is our alternative to the Liberman–Prince use of the feature [stress] and rules for assigning that feature. We argue at some length in Selkirk (1980, forthcoming) that the \( \Sigma \) approach to English stress allows for a superior analysis of English stress patterns. This being the case, our positing of the prosodic category \( \Sigma \) receives justification. Further justification would be provided if it could be shown that the \( \Sigma \) is a characteristic domain for the operation of phonological rules. Relatively little research has been done in this area, but already it is known that in English, it is within the \( \Sigma \) that the resyllabification takes place which creates the correct environment for the flapping of \( t \). For example, *total* \{ (tow) (twall) \} resyllabifies to \{ (towt) (twall) \}, and it as a consequence converts to a flap [tawl]. (See Selkirk (forthcoming) for a defense of this analysis.) Moreover, within a \( \Sigma \), the vowel of a weak syllable may have a tendency to delete; this is known to be the case in French (cf. Selkirk 1978), in Old English (cf. Keyser (forthcoming)).
in Egyptian Arabic, and other languages. The situation thus looks promising for establishing the \( \exists \) as a domain of phonological rules and thereby pointing to the necessity of the node label \( \exists \) in phonological representation.

Turning now to the prosodic word \( (\omega) \), we adapt, with relatively little modification, the Liberman and Prince analysis:

**The Prosodic Word: Constituency**

The \( \exists \) are joined in a right branching structure.

**The Prosodic Word: Prominence**

Given a pair of sister nodes \( \{N_1, N_2\} \), \( N_2 \) is \( \omega \) iff it branches.

(This is the Liberman and Prince Lexical Category Prominence Rule.)

The principles account for the internal structure of imperative, which we repeat here:

\[
\text{\textbf{im\textbf{per}e\text{\_a\text{\_tive}}}}
\]

\[
\text{\textbf{v\text{\_p\text{e\_k\_t}}}}
\]

\[
\text{\textbf{v\text{\_n\text{\_o\_m\_e\_t}}} \quad \text{\textbf{v\text{\_n\_o\_m\_e\_t}}}}
\]

The principles account for the internal structure of imperative, which we repeat here:

In addition, the syntactic domain of these principles must be specified. Because the neutral affixes are 'stress-neutral', i.e. do not affect the pattern of stress of the stem to which they attach, we assume that the simple stem is the domain of \( \omega \)-formation as well as of \( \exists \)-formation.

There seems to be a fair amount of evidence that the \( \omega \) is a privileged domain of phonological processes. Nakatani and Schaffer (1978) report, for example, that there is a significant elongation associated with a monosyllabic \( \omega \), one that contributes to the correct parsing of apparently ambiguous Adjective-Noun sequences (pronounced with nonsense syllables). \( \omega \)-initial consonant-lengthening is also generally attested in English (cf. Nakatani and Schaffer (op. cit.) and Lehiste (1960)). It is also a well-known fact that \( \omega \)-final non-low vowels are tensed in English (cf. Chomsky and Halle 1968). Clearly the \( \omega \) is a category of prosodic structure, one having a distinctive internal organization, giving rise to the proper patterns of stress prominence, and one serving as a distinctive domain for phonological rules.

To summarize, our examination of the organization of phonological representation has so far led to the conclusion that that representation is hierarchically arranged and that specific subunits of that hierarchy -- the \( \omega \), the \( \exists \), and the \( \omega \) -- have to be isolated. Each one of these prosodic categories has its own specific principles of internal constituency and prominence, and each one functions as a characteristic domain of phonological rules. The proper conception of prosodic structure is thus one that constitutes an enrichment of the Liberman and Prince conception.

It should of course be apparent from the preceding discussion that prosodic constituency does not reflect syntactic constituency. But the well-formedness of prosodic structure at this level must be defined with respect to some syntactic domain. It happens to be the case that in English, and in other languages we have investigated as well, the syntactic or morphological domain for \( \omega \), \( \exists \), and \( \omega \) is identical. In English it is the simple stem; in Egyptian Arabic it is the syntactic word; and so on. It would not be unreasonable to propose this as a universal constraint on possible prosodic structure -- that \( \omega \), \( \exists \), and \( \omega \) always share the same syntactic domain of well-formedness.

Above the level of \( \omega \), the question of the correspondence between syntactic structure and prosodic structure is something more of a live issue, and it is to that issue that we now turn.

2.0 Looking at the structure of the representation above the level of the prosodic word, we will see that prosodic words first group into phonological phrases \( (\phi) \), that the phonological phrases are grouped into intonational phrases \( (i) \), and finally, that the intonational phrases are what make up the utterance \( (u) \).

The principles governing the nature of the internal structure of the phonological phrase refer directly to the syntax of the sentence. The major principles regarding the constituency of \( \phi \) are two in number:

**The Phonological Phrase: Constituency**

(i) An item which is the specifier of a syntactic phrase joins with the head of the phrase.

(ii) An item belonging to a 'non-lexical' category (cf. Chomsky 1965), such as Det, Prep, Comp, Verbaux, Conjunction, joins with its sister constituent.

The effects of these two principles are illustrated in Figure 1 (pp 16-17) where in line 2 the grouping into \( \phi \) of the elements of the sentence The absent-minded professor has been avidly reading about the latest biography.
The absent-minded professor has been avidly reading about the latest biography of Marcel Proust.

By rhythm rule:

The absent-minded professor has been avidly reading about the latest biography of Marcel Proust.

By monosyllable rule:

(The nodes circled have undergone the change introduced by the rule.)
of Marcel Proust is shown. In the first case, by (i), the adjective _abstemious_ is joined with the head noun _professor_, and by (ii) the determiner _the_ joins them. In the second case, by (i) the adverb _silently_ joins with the head noun _reading_, and by (ii) the two auxiliary verbs _has_ and _been_ form part of that _be_ as well. And so on. The phonological phrases formed in this way are right-branching.

The prominence relations within the phonological phrase are governed by what is essentially the Liberman-Prince formulation of the Nuclear Stress Rule:

The Phonological Phrase: Prominence

Given the two sister nodes of prosodic structure \([N_1 N_2]\) within \(\varphi\), \(N_2\) is \(\varphi\) (and \(N_1\) hence weak).

Evidence for the existence of a prosodic category \(\varphi\) in English, with the syntactic composition just described, can be deduced from the behavior of a number of phonological rules. The first of these is the Rhythm Rule. Liberman and Prince (1977) give their version of this rule the name thematic Reversal. It is the Rhythm Rule which is responsible for changing the w s w s original w s patterns of _abstemious_ and _Marcel_ to w w, as illustrated in line 3 of Figure 1. The rule may be formulated as follows:

Rhythm Rule

\[
\begin{align*}
\text{w} & \quad \text{s} \\
\text{w} & \quad \text{s} \\
\end{align*}
\]

\[
\begin{align*}
\text{w} & \quad \text{s} \\
\text{s} & \quad \text{w} \\
\end{align*}
\]

It says that when a weak left sister has the internal composition \(w\) \(s\), those internal prominence relations are switched to \(s\) \(w\). Note that the inversion only takes place when the right sister is strong: if _Marcel_ were being focused or contrasted (in comparison to _Antéine Proust_, for example), it would retain its original internal pattern:

\[
\begin{align*}
\text{R}\text{s} & \quad \text{w}\text{w} \\
\text{R}\text{s} & \quad \text{w}\text{w} \\
\end{align*}
\]

\[
\begin{align*}
\text{R}\text{s} & \quad \text{w}\text{w} \\
\text{w}\text{s} & \quad \text{w}\text{w} \\
\end{align*}
\]

We would claim that the Rhythm Rule operates only within the domain of what we are calling the phonological phrase and thus that it gives support to the hypothesis that there does exist such a prosodic category. Note, for example, that the rule does not apply when _Marcel_ is in subject position in the sentence: _That's one of the theorems that Marcel proved._

In a Liberman-Prince representation, the sequence _Marcel proved_ has a prosodic structure identical to that of _Marcel Proust_. In ours, it does not: the subject noun phrase is \(\varphi\) on its own.

By specifying \(w\) as the domain within which the Rhythm Rule applies, it can be prevented from applying in such environments.

A second phonological rule of English which has \(\varphi\) as its domain is the so-called Monosyllabic Rule (cf. Selkirk (1972)). In earlier discussions, this rule was seen as one which destressed monosyllabic words belonging to non-lexical categories, when they appeared in certain syntactic contexts. Given the theory of prosodic structure now being entertained, the rule can be given a simpler, more explanatory, formulation:

Monosyllabic Rule

\[
\begin{align*}
\text{w} & \quad \text{w} \\
\text{w} & \quad \text{w} \\
\end{align*}
\]

\[
\begin{align*}
\text{w} & \quad \text{w} \\
\text{w} & \quad \text{w} \\
\end{align*}
\]

\[
\begin{align*}
\text{w} & \quad \text{w} \\
\text{w} & \quad \text{w} \\
\end{align*}
\]

The rule says that a monosyllabic prosodic word \(\varphi\) is 'deworded' (and hence 'defooted') if it is weak and corresponds to a non-lexical item in syntactic structure. The rule is responsible for the changes exhibited in line 4 of Figure 1. The, has, been and of are all turned into simple weak syllables by the rule, with the result that they may undergo vowel reduction and other such rules which apply only to weak syllables. Note that the bisyllabic about, though a \(\varphi\) which corresponds to a non-lexical item, does not reduce.

There is abundant evidence that a weak monosyllabic non-lexical item will not reduce to \(\sigma\) unless it is weak with respect to a strong CONTAINED WITHIN THE SAME \(\varphi\). Consider, for example, the contrast between verb particles and identical prepositions, as illustrated here:
The cops boxed [pp[fnp]pp [the crowd]] = (the cops) (boxed) (in) (the crowd) (boxed) (in the crowd)

Because a verb particle is a prepositional phrase on its own (cf. Emmonds (1976)), it forms a Φ on its own, and hence is never subject to the Monosyllabic Rule -- as long as the rule is restricted to the Φ domain.

What we are calling the phonological phrase is thus the privileged domain of two rules of English which apply to and modify prosodic structure. In other languages such as French and Modern Greek, there is evidence from rules of the SEGMENTAL phonology that what we define as the phonological phrase constitutes a special domain. In French, for example, there is obligatory LIAISON within Φ. (cf. Selkirk (1974).) Contrast the following pairs:

(le petit enfant) [la pet[i]t enfant] 'the little child'

vs. (le petit) (en mange) [lpet[i]t mɑnɡ] 'the little one is eating some'

(allez-vous-en) (parler) [alevuzɔparle] 'go away to talk'

vs. (allez-vous) (en parlez) [alevuzɔparlez] 'Are you going to talk about it?'

In the first examples the word-final consonant is internal to a Φ (and precedes a vowel); it is therefore not deleted, as are the consonants of the second examples, which are in Φ-final position. All in all, then, the Φ is a necessary unit of linguistic analysis; it forms a necessary part of prosodic structure in phonological representations.

It must be noted now that the Φ is not by definition isomorphic to any constituent of syntactic structure. What is contained in Φ may not correspond to any syntactic constituent at all. A comparison of lines 1 and 2 in Figure 1 should make this clear. Corresponding to the uniformly right-branching structure of the VP of syntactic representation are three discrete constituents of prosodic structure: has been avidly reading, about the latest biography, of Marcel Proust. If our claims about the existence and nature of Φ are well-founded, one cannot conceive of prosodic by Liberman and Prince.

There is one syntactic generalization concerning the Φ which may be important in a model of speech perception. Every Φ ends in the HEAD of logical phrases provides fairly direct and easy access to a certain type of information -- that about 'headhood', a type of information which seems particularly important in unravelling the syntax and semantics of a sentence.

We turn next to a consideration of the intonational phrase. In our discussion we will be assuming the correctness of the contention that the intonational contour of an entire utterance may be segmented into a set of smaller primitive intonational contours, each with its own defining characteristics (cf. Liberman 1975, Bing, 1979) and we will be assuming furthermore that each of these contours is associated with an intonational phrase (see Bing 1979). One of the characteristics of the prosodic unit intonational phrase (I), then, is that it is the domain over which an intonational contour is 'spread'.

The internal composition of the intonational phrase can be described in very simple terms: it is made up of one or more phonological phrases. For the moment, we will assume that the choice of just which or how many Φ go into an I is free, though further research on the matter will undoubtedly reveal that considerations of length, and perhaps even syntax, are involved. What is at issue here is the hypothesis that Φ's are the constituent units of I's, and that they may group in whatever fashion in variable phrasing arise in the pronunciation of a sentence: The minimal intonational phrase is a single Φ, the maximal one is that which includes as illustrated in Figure 2.
Figure 2

VARIABLE PHrasing

U

I

s

s

The absent-minded professor has been avidly reading the latest biography of Marcel Proust

U

I

s

s

The absent-minded professor has been avidly reading the latest biography of Marcel Proust

FIGURE 2 (continued)

U

I

s

s

The absent-minded professor has been avidly reading the latest biography of Marcel Proust

U

I

s

s

The absent-minded professor has been avidly reading the latest biography of Marcel Proust
As is well-known, there is not always such an optionality to the designation of intonational phrases in a sentence. Certain syntactic configurations require that the sentence be broken down into intonational phrases in a particular way (cf. Downing (1971)). So, for example, as shown in Figure 3, preposed adverbials, non-restrictive relative clauses, and parenthetical expressions, must each correspond to an intonational phrase in prosodic structure. (Note that giving the status of intonational phrase to a sentence-medial constituent like a parenthetical has as an immediate consequence the fact that the sentence contains at least three intonational phrases (and intonation contours). A medial intonational phrase will be bounded by one on either side.)

We see then that the I is defined only partially in terms of syntactic factors. A certain class of syntactic constituents must correspond to an intonational phrase. In all other cases, the intonational phrasing is done in terms of phonological phrases, that is, in terms of prosodic structure, without direct reference to syntax. Note that it is not being denied that one could always give a syntactic characterization of what was a possible intonational phrase. This could certainly be done—by enumerating the contents, syntactically defined, of phonological phrases. What is being claimed is that if this were done, the generalization would be missed that the I is simply a sequence of possible $\phi$'s.

In any case, what should be apparent is the fact that an I does not necessarily correspond to a constituent of syntactic structure.

Before proceeding to briefly review some of the evidence provided by phonology that the I is indeed a hierarchical unit, a phrase of prosodic structure, it would be desirable to make more precise the definition of I as a prosodic unit. In particular, the question to be asked is whether the $\phi$'s constituting the I are joined in a binary-branching tree, labelled s/w, or whether the structure is flatter, with, possibly, no s/w prominence relations being defined. Both would seem to be possible, e.g.

But we will opt for the second, insofar as it is consistent with the overall Liberman and Prince-type approach we have been taking to prosodic structure. The principles governing I could therefore be put as follows: (p. 26)

- **Figure 3**

OBLIGATORY PHRASING

- **Preposed Adverbials**

- **Non-Restrictive Relatives**

- **Parentheticals**

Tuesday is, Jane said, a holiday.
The Intonational Phrase

Constituency

The I is composed of \( \psi \) joined in a right-branching structure.

Prominence

In I, the nodes \([N_1, N_2]\) are in the relation w/s.

Syntactic Domain

a) Parentheticals, preposed adverbials, non-restrictive relative clauses, etc. are I's.

b) Otherwise, the choice is free.

The intonational phrase I has been assumed to be a structural unit which has phonological characteristics relating to the fundamental frequency of the utterance. It is the unit with which the primitive intonational contours are associated. It is also the unit which defines the domain of the declination effect (cf. Breckenridge 1977): at the start of a new intonational phrase, the pitch range is reset, across that phrase the declination effect is observed, and at the beginning of the next I phrase, pitch is reset again. And it can also be shown that it is the privileged domain of certain rhythmic or timing effects. For example, between elements contained within a same I there are certain, albeit restricted, possibilities of Refooting, i.e. of associating a weak syllable from a following \( \omega \) or \( \phi \) to the foot of a preceding \( \omega \). The flapping of \( t \) in such contexts as Go to the store, go today, give the money to Mary is evidence for Refooting, for if the \( t \) were \( \alpha \)-initial, and hence \( \alpha \)-initial, it would be aspirated instead. The result of flapping the weak syllable containing \( t \) is that the \( t \) may become part of the preceding syllable and be thus in the proper position for Flapping (see Selkirk forthcoming for details of this analysis.) Compare now, the pronunciation of the \( t \) in today when it is initial in an intonational phrase: According to Mary, today is a holiday. The flapped pronunciation seems distinctly unnatural. This would follow if Refooting were restricted to applying within the I domain. In other languages, such as Spanish, rules of the segmental phonology are known to apply only within the I-phrase. (See Harris (1968) and Kahane and Behm (1948) on this point.)

Important phonological phenomena are known to be associated with the limits of the I-phrase. It is at the beginnings or ends of the I that the so-called 'boundary tones' are found (cf. Liberman (1975), Bing (1979)). It is BETWEEN intonational phrases (and only between them, we would claim) that one finds potential pauses. And it is at the end of I that so-called pre-pausal lengthening is found. (Note that Klatt (1976) correlates

the location of pre-pausal lengthening with the location of possible pauses.)

There is thus a whole complex of phonological phenomena which take the intonational phrase as their domain. The intonational phrase is not merely a sequence over which an intonational contour is distributed; it is a rhythmic entity as well, and one which has a special status with respect to other segmental and suprasegmental rules. This means of course that where one finds variable phrasing, one expects to encounter the entire host of related phenomena working in tandem: if the \( \psi \) corresponding to the subject noun phrase is an I, it will have an intonational melody associated with it, have prepausal lengthening at the end, and so on. By postulating the I as a structural unit, as a category of prosodic structure which defines a particular type of domain one expects this sort of correspondence of seemingly disparate phenomena. The convergence is, in this sense, explained. It should go without saying, of course, that a prosodic tree unmarked for its decomposition into I's is inadequate as a representation of the utterance. The rules of phonology must 'know' what is an I and what isn't.

Finally we turn to the highest-order category of prosodic structure -- the utterance. We will not have much to say about this unit, except that it consists of one or more intonational phrases, that these are obviously in any relation of subordination to each other (though, in the spirit of Liberman and Prince, may be represented as such), and that the utterance usually coincides with the 'highest' sentence in syntactic structure. See Figure 4, where a sentence with a multitude of intonational phrases is represented in this way. (It is to be noted that the presence of so many obligatory I's (the circled ones) has the result that every other constituent is also an I, each with its own primitive intonational contour, in particular a nucleus.) From this example it should also be quite apparent that the prosodic structure of the sentence does not replicate its syntactic structure. The two are indeed distinct.

There are doubtless phonological phenomena which take the utterance as their domain. \( \Omega \) is not unheard of for certain phonological alternations to be restricted to either utterance-initial or utterance-final positions, and, though no examples come to mind at present, we would expect there to be rules which could apply across the entire span of the utterance.

3.0 In summary, a particular conception of prosodic structure has been argued for, and it has been put in relation with syntactic structure. While assuming the essential correctness of the Liberman-Prince notion that prosodic structure is essentially a binary branching tree with nodes labelled s/w, we have argued that this representation must be enriched by the addition of prosodic categories. Prosodic categories are the constitutive units of prosodic structure, the units in terms of which the phonotactics of an utterance may be defined. They are also the units or domains with respect to which phonological phenomena apply.
It should be noted that once prosodic categories form part of the phonological representation, the motivation for boundaries (cf. SPS, Selkirk, 1972) as part of a phonological representation disappears. Boundaries are none other than an encoding in the string of segments (the standard generative representation) of the higher order prosodic structure organizing that string. At one time thought of as the appropriate formal mechanism for delimiting domains of application of phonological rules, boundaries are rendered obsolete by the introduction of prosodic structure into phonological representation and, in particular, the inclusion of prosodic categories which effectively designate those domains. Looking at it in a different light, we can think of the focus being taken away from boundaries, which were originally viewed as markers of juncture between syllables and other prosodic units, and being put on the prosodic units themselves.

The final point is that this review of prosodic structure has shown quite clearly that prosodic structure is not syntactic structure, nor is it isomorphic to it. The two are quite distinct as formal objects. A mapping from one to another can, and must, be defined, however, for the prosodic structure reflects syntactic structure in certain ways. We would suggest that the mapping from syntactic to phonological representation in a generative grammar is precisely this mapping — the mapping between syntactic and prosodic structure.

Thinking now in terms of speech production and perception, we would hypothesize that the units of prosodic structure we have discussed here in linguistic terms are indeed the appropriate units in production and perception models, that the effect of syntactic phrasing in production, or access to that phrasing in perception, are crucially mediated by these units of the prosodic hierarchy.

REFERENCES


Nakatani, Lloyd H. and Judith A. Schaffer. 1978. 'Hearing words' without words: Prosodic cues for word perception. JASA 63(1).


