THE INTONATIONAL SYSTEM OF ENGLISH

by

MARK YOFFE LIBERMAN

M.S., Massachusetts Institute of Technology
1972

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

September, 1975

Signature of Author..
Department of Linguistics,
August, 1975

Certified by..
Thesis Supervisor

Accepted by..
Chairman, Departmental Committee
on Graduate Students

THE INTONATIONAL SYSTEM OF ENGLISH

by

MARK LIBERMAN

Submitted to the Department of Foreign Literature and Linguistics on August 15, 1975 in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

ABSTRACT

The task of a theory of English intonation is defined as the representation of stress, tune, and phrasing, and the explication of their interactions with each other and with the rest of the grammatical system.

It is argued that the association of text (tone-less linguistic material) and tune is accomplished by a metrical system, which assigns metrical patterns to text and tune, establishes a congruence between these patterns in any given case, and specifies possible alignments of the congruent patterns with a metrical grid. Metrical patterns are viewed as abstract structurings of complex events; metrical grids as abstract structurings of time.

A preliminary account of the phonology and morphology of the English tonal system is given: a number of specific tunes are identified and discussed. It is argued that the tonal lexicon of English is ideophonic in character, and that certain of its important properties follow from this fact. The investigation of complex tunes and tune-sequences is observed to raise important issues for syntax and semantics.

The theory of metrical patterns is argued to be of value in accounting for English stress patterns. A metrical reformulation of the phrasal stress rules, and a partial reformulation of the word-stress rules, seem to call into question the theory of the phonological cycle. Results of a preliminary formalization of the idea of metrical grids suggest the resurrection of the doctrine of stress-timing.

The role of a metrical system in defining English stress patterns, and in coordinating tune and text, is related to a very general hypothesis, first suggested by St. Augustine, about the organization of temporally structured behavior.

Thesis Supervisor: Morris Halle

Title: Professor of Linguistics

Whatever in this thesis is true, is dedicated to my father, from whom I learned that what is true is simple; to Morris Halle, from whom I learned the value of expressing simple ideas in a formal way; and to a little girl in a yellow dress, from whom I learned a simple truth about intonation.

What remains, is dedicated to those who improve on it.

TABLE OF CONTENTS

0. Introduction

0.1 WHY?

In the 200 years since Joshua Steele published his Essay Towards Establishing the Melody and Measure of Speech, many works on this subject have come into the world. A number of them bear titles rather similar to that of this thesis -- The Intonational System of American English, Prosodic Systems and Intonation in English, etc. Why, then, have I added another tract to this (already bewilderingly diverse and extensive) pile?

My interest in intonation stems originally from work in syntax and semantics. As anyone who has ever labored in those vineyards knows, it often happens that the way a sentence is said has a crucial effect on its value as an example of some given principle. I was frustrated at not being able to factor intonation out of syntactic and semantic arguments with any confidence, not knowing what it really was. I was equally frustrated by my attempts to learn what it was by reading the existing literature in the field, which is split into a number of separate and mutually unintelligible traditions, none of which seemed to provide satisfactory answers to the questions that interested me.

I therefore set out to construct a theory of intonation from first principles. My idea of the goals of the enterprise has remained fixed from the beginning -- it is set out in chapter 1, under the heading of <u>The</u> <u>Problem</u>. My idea of the underlying principles of the intonational system has undergone considerable metamorphosis, however. I will not bore the reader with a detailed history of the various errors I have fallen into at one time or another in the course of this two-year-long chutes-and-ladders game.

I think it is worthwhile, however, to make a general comment on two properties that most of these errors have shared -- insufficient abstractness and excessive complexity. For example, I associated the tonal aspect of an utterance too closely with its "textual" aspect -- one breakthrough came when I became willing to conceive of the "tune" as an entity which is in origin completely independent of the "text"; not an aspect of the features of the segmental string, not a set of suprasegmental diacritics, not even a separate string of segments, but a <u>completely</u> <u>independent</u> <u>structure</u>. Each step in this progression brought progress; each represented a more abstract conception of the nature of tonal phenomena in English.

To give another example, I once sought to merge stress and tone into a single system -- progress came when I recognized that they are entirely distinct phenomena. In reaching this conclusion, I came to an understanding of the nature of stress patterns which, I believe, allows a more natural and successful treatment of English stress rules than is otherwise possible -- this understanding was essentially a move in the direction of greater abstractness, treating stress patterns as structures defined on strings of segments, rather than sets of features inherent in the segments themselves.

With each of these moves in the direction of abstractness came a concomitant reduction in the size and complexity of the apparatus (particularly the language-particular apparatus) necessary to make the system work descriptively. Once the abstract concepts themselves are defined, many apparently arbitrary aspects of the descriptive system begin to fall into place.

This is all to the good, not primarily because order and simplicity are aesthetically desirable properties of a theory (although in my opinion this is true), but because the main goal of a theory of language is to explain why language is so humanly natural, a goal which can only be approached by imposing the most rigorous possible demand that the facts of human language should be orderly and simple when expressed in terms of that theory.

In endeavors of this sort, my experience is that you find what you look for. If your taste or creed inclines that way, you will find plenty of wild facts to tangle with, and reasons to be dissatisfied with the domesticated examples in almost anyone's garden. The surfaces of nature are always rough; our everyday communication generally bears as little relation to the linguist's constructs as a mountain does to those of a crystallographer.

On the other hand, if you refuse to be content with surfaces, and insist that the things we do so simply and naturally in everyday speech really are simple and natural, if only they were properly understood, then you will find your effort vindicated. A mountain is not a crystal, but there is much in it that is crystalline.

0.2 HOW?

This is my third introduction in as many months, a fact which points to the rapidly evolving character of what follows. Its evolution is now, I hope, complete except for details of stripe and spot, but the discerning reader will note certain evidences of immaturity here and there in the text, for which I beg indulgence.

Whenever I have been confronted with someone else's newish theory, like everybody else I always ask various versions of the two questions "well, what about phenomenon X?" and "what do you mean by concept Y," where a wide range of constants are substituted for the variables X and Y. Whenever I myself have submitted infant theories for the inspection of others, I get asked the same questions. The only real answer to either kind of question, of course, is "whatever's right;" this is not immediately satisfactory to either party, but if the discussion then attempts to thrash out what "whatever's right" actually is, progress generally results.

A number of people have helped me in putting the ideas in this thesis through such a process -- I might single out for special mention the members of my committee, Morris Halle, Noam Chomsky, and Paul Kiparsky, who have also inspired me by the example of their work. Haj Ross found space, in a busy time of year, to read an early and somewhat incoherent draft of chapters 2 and 3. Other helpful discussion was provided by Mary-Louise Kean, who was kind enough to read the final version as it came off the typewriter. In earlier stages of the development of my ideas about intonation, I profited from discussions with

John Goldsmith, Shosuke Haraguchi, Will Leben, Jacqueline Vaissiere, and Shinji Maeda.

A different kind of thanks is due to those who have related to my pilgrimage not as discussants but as participants. I have written a pair of papers on aspects of intonation with Ivan Sag, an experience to which I owe most of my original (and present) conception of the nature and scope of the problem. Much of the proposal for a new treatment of English stress rules (in chapter 4) was worked out in cooperation with Alan Prince, without whose help it would never have seen the light of day (although he should not be held responsible for any errors in the particular formulation which I present). My wife Ida made the enterprise possible, in more ways than I can express; the most minor of her contributions was to type the final draft.

1. The Problem.

1.1 Everyone knows that there are many ways in which a "sentence" can be altered in its effect by the way it is said. In some cases the alteration is such that we are inclined to say that the different "ways of saying" a string of words reflect fundamentally different sentences, that is, different syntactic structures. In other cases, we would conclude that the same syntactic structure is being performed in systematically different ways. In very many cases, it would not be at all clear what the correct analysis ought to be.

As a pretheoretical expedient, we might bow to popular usage and translate the "way of saying" a string of words as its intonation. Then the kind of issue raised in the preceding paragraph becomes a part of the more general problem of the status of intonation in grammar. The issue for linguistic theory is, how and where should intonation be represented?

By "how," I mean to raise the same questions which we consider in representing any linguistic phenomenon, that is, the questions of underlying form (the most systematic, abstract mode of representation), surface form (the most particular and concrete mode of representation which is of systematic significance), and derivation (how to relate underlying and surface forms).

By "where," I mean to raise a question which again is familiar to us from the study of other linguistic phenomena, namely, at what point in the derivation of a sentence should intonational representations be introduced? Two related issues are, what other linguistic processes have access to information about intonational representations, and what kinds of information can condition intonational derivations?

1.2 The nature of the problem will become clearer if we look at a few simple examples.

1.2.1 Stress Differences.

There are a large number of cases in which differing intonations, different ways of saying a string of words, can be described as differing patterns of stress. For example, different placement of primary stress in the noun phrase "English teacher," as represented in 1.2.1/1, results (normally) in a difference in meaning:

1.2.1/1 a $\overset{1}{\text{English}}\ \overset{2}{\text{teacher}}$

b $\overset{2}{\text{English}}\ \overset{1}{\text{teacher}}$

By 1.2.1/1a, we would normally mean a teacher of English; by 1.2.1/1b, we would normally mean a teacher who is English (the fact that this pattern can be violated by "contrastive stress" does not in any way affect the conclusion that such a pattern exists).

The pattern just exemplified has been explained in the following way: the two versions of phrases like "English teacher" differ in syntactic structure, the version in 1a being of the form noun + noun, while the version in 1b is of the form adjective + noun; there are two different rules which can assign stress to such phrases, the "Nuclear Stress Rule" (NSR) and the "Compound Stress Rule" (CSR); the different constituent structures attributed to 1a and 1b will lead to different semantic interpretations, and will also condition the assignment of primary stress by ensuring that the CSR will apply to 1a while the NSR applies to 1b.

This account has been called into question in the literature,[1] but it seems to me to be fundamentally correct. The argument most frequently used against it is that many *noun* + *noun* phrases go by the NSR instead of the predicted CSR, while many *adjective* + *noun* phrases take the CSR instead of the predicted NSR. The examples cited fall into three classes: first, those that arise because of what might be termed the "information structure" of the examples, involving contrast, anaphora etc.; second, those that involve some systematic difference in the semantic relationship between modifier and head (e.g., *steel wárehouse*, "a warehouse made of steel," vs. *stéel warehouse*, "a warehouse for storing steel"); and third, those that seem to reflect idiosyncratic properties of particular lexical items (e.g., *Madison Ávenue* vs. *Mádison Street*).

We will return to this issue in a later section. For now, I will simply state my conviction that examples of the first kind are irrelevant to the question of whether there are rules of sentence stress which depend on syntactic structure, demonstrating only that other factors are also involved in determining stress patterns, while examples of the second and third kinds actually support the view that stress assignment is (in part, at least) structure-dependent

Thus, if we accept the traditional analysis, we have for this case of intonational differentiation both a description (in terms of stress pattern) and an explanation (in terms of the differential application of the NSR and CSR). We could cite many cases of intonational differentiation by stress pattern in which the explanation is much less clear. However, as long as we have a mode of description (in terms of relative levels of stress) which is adequate to differentiate among the cases which we are attempting to explain, then we are in a good position to look for the explanation.

An example of a case where this situation obtains is the following:

1.2.1/2a John called Mary a Republican, and then shé insulted hím.

b John called Mary a Republican, and then she insúlted him.

Such examples have been discussed by Green, Lakoff and (most systematically) Williams[2]; many of the properties of these cases remain somewhat puzzling, but there is little difficulty in determining what the facts are, since almost any notation in which levels of stress are distinguished will suffice to define the body of data to be accounted for.

Although the notation of patterns of stress, however it is accomplished, is able to support a great deal of interesting research, and obviously corresponds (at least in part) to some real linguistic property, it does not answer all the questions about the "way of saying" a string of words which we raised in 1.1. It falls short in two ways -- first, differentiation by stress pattern is given an underlying form, but we are left far short of its surface form. Nothing is said about how a given stress pattern will be realized phonetically, how it will affect (as it obviously does) the pitch, intensity and timing of the utterance of which it is an abstract property. Secondly, no account of any sort is given for those cases in which utterances with identical stress patterns are given different "meanings" by variations in pitch contour, in intonational phrasing, etc.

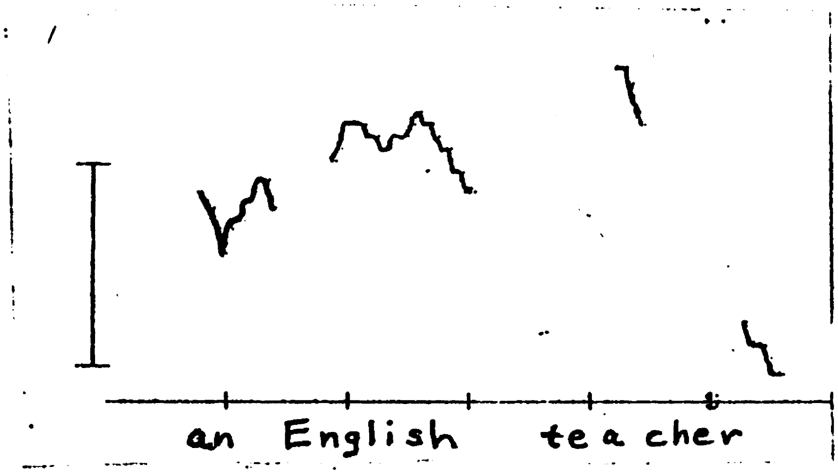
These inadequacies are not arguments against the theory of stress patterns, but merely a demonstration that it is insufficient, and must be supplemented (at least) by theories of pitch contour and intonational phrasing. In fact, investigation of these additional phenomena will show all the more clearly that a theory of stress patterns, in some form, must be a proper sub-part of the theory of intonation as a whole.

1.2.2 Tune Differences.

A second set of cases where identical strings of words are intonationally differentiated, consists of example-pairs whose stress patterns are identical, but whose F_0 contours are different in a way that affects meaning. In these cases we could say that the same words are being said with different "tunes."

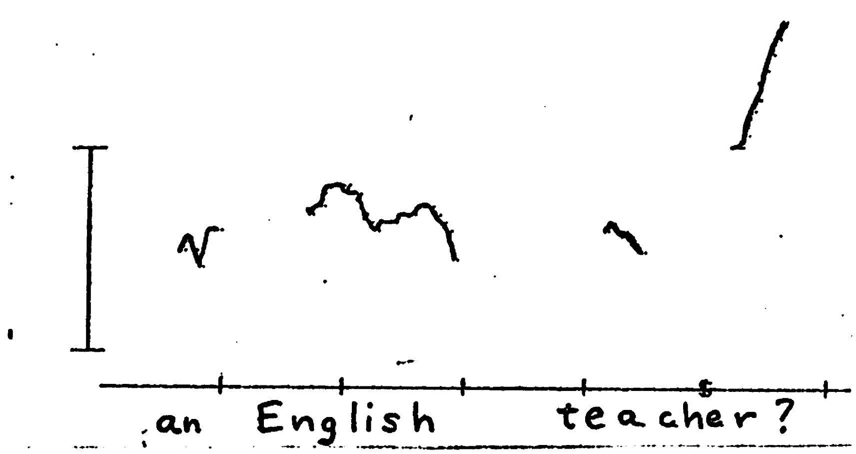
Some simple examples of tune differences are given below,[3] specifically, four different versions of the noun phrase "an English teacher." All four versions have the stress pattern $\overset{2}{\text{English}}$ $\overset{1}{\text{teacher}}$ (in SPE[4] notation), and all of them retain the meaning previously noted to be characteristic of this stress pattern, "a teacher who is English." Nevertheless, they differ among themselves quite considerably in other aspects of their "meaning," at least with respect to the circumstances in which it would be appropriate to use them.

1.2.2/1



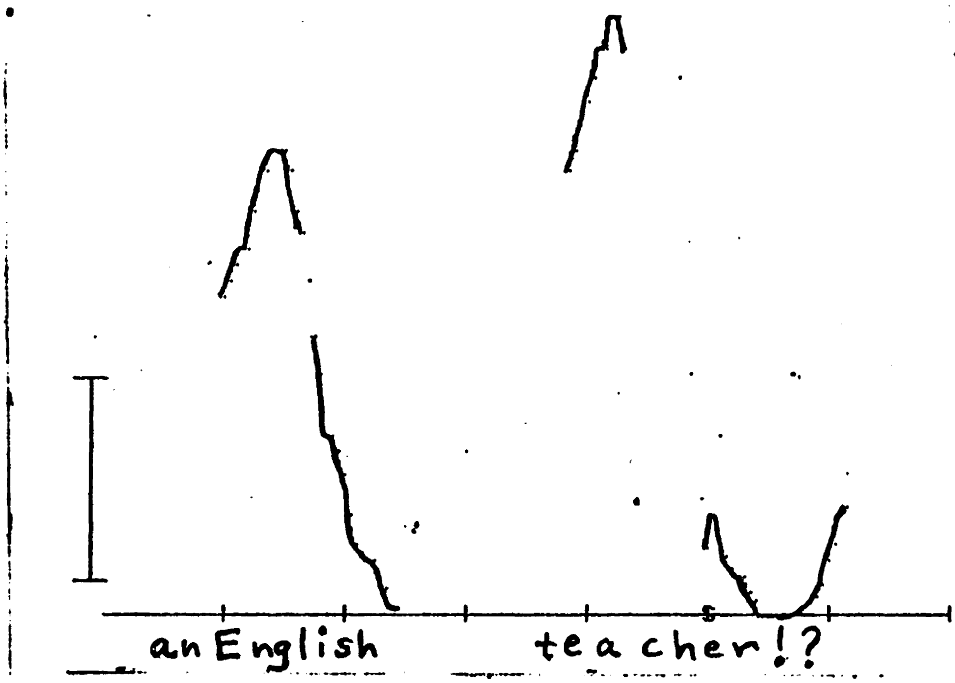
Example 1 above is a fairly neutral version of what may be called "declarative" intonation.

1.2.2/2



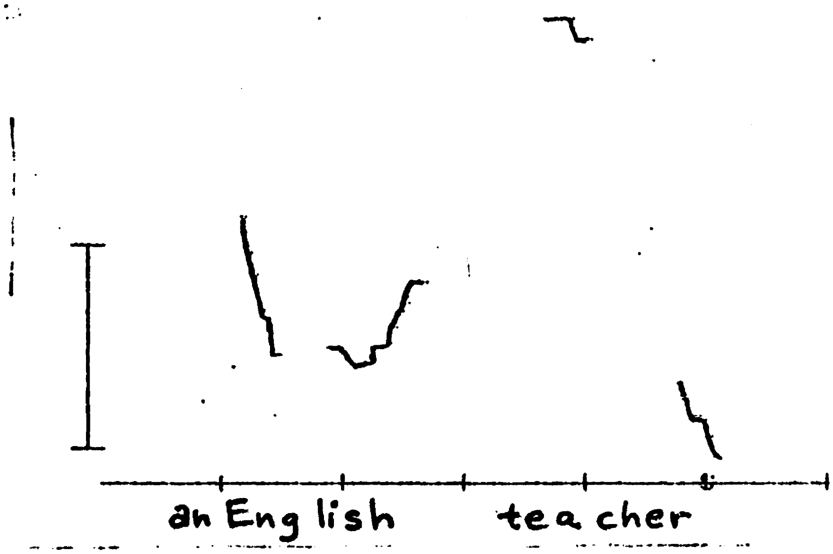
Example 2 represents the intonation which is commonly used in asking yes/no questions.

1.2.2/3



Example 3 shows a tune which is often used in expressions of incredulity -- this instance of it might, for example, be employed by an Anglophobic anti-intellectual who has just been informed that his daughter is to marry a teacher of British origin.

1.2.2/4



In example 4 we have a tune whose meaning in this case might be something like "what I'm saying is perfectly obvious -- what else could the answer be?" This contour is studied in some detail in <u>Sag and Liberman</u> 1975.

In the investigation of such "tune differences" we are hampered by the lack of any generally accepted notation in terms of which to describe them. Since the essence of any notation is a theory (expressed or implied) about the nature of the thing notated, this lack of an agreed-on notation reflects the lack of a theory of "tunes" successful enough to merit general acceptance.

There are, of course, a number of proposals in the literature intended to fill this need -- some of the more prominent ones will be reviewed in a later chapter. However, none of them allows us to answer the questions raised in 1.1 in an entirely satisfactory way.

Even without a theory which would specify the underlying form, relation to the rest of the grammar, and surface form of "tunes" such as those in 1.2.2/1-4, it is possible to study such "tunes" in a fruitful way. Simply claiming that certain cases are systematically different from each other, as we did in reference to these four simple examples (a procedure which is carried out in an infinitely more ambitious and systematic way in the literature), is an advance, in that it helps to define the job that a theory of intonation must do. Examples, and classifications of examples, are easy enough to multiply; such investigations are often interesting in their own right, for the light that they are able to shed on issues in semantics, pragmatics and even syntax.

However, as anyone who has done much reading in the literature on intonation is well aware, it is possible to describe quite systematically a very large number of examples of "tunes" without making much progress towards answering the fundamental questions of what these "tunes" are, and how they should be integrated into linguistic descriptions. Furthermore, the description of even moderately complex cases becomes quite murky, and we often find very basic disagreements about how to describe what is going on, even in those cases where all parties have got the F_0 contour right. This descriptive murkiness is a direct result of the lack of theoretical agreement and/or understanding.

1.2.3 Phrasing Differences.

A third mode of intonational differentiation involves cases where "intonational phrasing" distinguishes one example from another. This phenomenon is often discussed in terms of the location of "commas," "pauses," "intonation breaks" etc., although not all examples lend themselves very easily to such treatment.

Differences in phrasing often reflect very basic changes in the nature of the sentence under consideration --

1.2.3/1a Sam struck out my friend.

b Sam struck out, my friend.

Thus without a "comma" or "intonation break" in front of it, the noun phrase "my friend" is the object of the verb "struck out"; when n the "comma" precedes it, the same noun phrase is a vocative.

In this case everyone would agree that the intonational difference corresponds to (and presumably results from) a difference in syntactic form. Our understanding of such cases is roughly comparable to our position with regard to stress patterns -- we can describe the phenomenon (e.g., by inserting or removing a comma) in a way which is adequate to allow us to study its distribution; in some cases we can explain that distribution (e.g., by hypothesizing that "intonation breaks" occur at certain sorts of syntactic boundaries); but we have no satisfactory way of describing the effect of intonation breaks on what Joshua Steele called "the Melody and Measure of Speech."

We also lack a general theory of what "intonation breaks" are, a theory which would answer questions like: is there more than one kind? are some stronger than others? what is the relation of IB's with notable syntactic and/or semantic effects to those cases in which the speaker merely pauses to think of a word? what factors can determine the position of IB's? what linguistic rules can refer to them? should IB's be viewed simply as a kind of marker inserted into phonological strings, as the "comma" notation implies, or are they instead the various constituent boundaries of some sort of intonational constituent structure?

Questions like these must clearly by faced in constructing a theory of intonational phrasing, but they also arise, at least implicitly, whenever the task of linguistic description requires us to deal in any detail with intonational facts.

One fairly straightforward example arises in connection with an issue in the theory of semantic interpretation. In his thesis[5], Howard Lasnik points out that the presence or absence of an optional intonation break in front of certain adjunctive adverbials correlates with the scope of a negative element associated with the main verb of the clause:

1.2.3/2a Senator Eastland didn't grow cotton, to make money.

2b Senator Eastland didn't grow cotton to make money.

In 2a above we state that Senator Eastland didn't grow cotton, and add that the reason for this was to make money (i.e., by receiving Government subsidies); in 2b we state that it is not the case that Senator Eastland grew cotton to make money. This difference can be described by saying that in 2a the adjunct "to make money" is outside the semantic scope of the negative element, while in 2b it is within the scope of the negative.

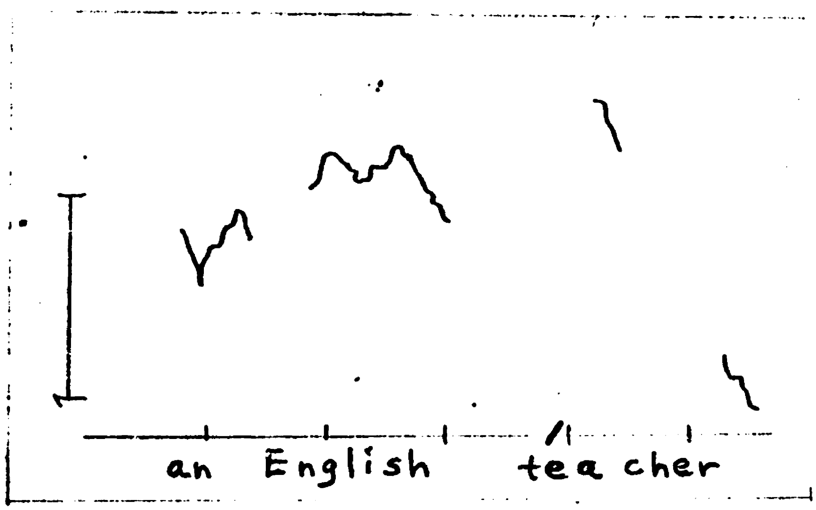
Lasnik concludes that this distinction should not be represented as a difference in syntactic structure, but rather as an optionality in the intonational phrasing of a single syntactic form, thus requiring intonation assignment (or at least the assignment of intonational phrasing) to precede the semantic interpretation rule that determines scope of negation. In a paper entitled "On Conditioning the Rule of Subject-Auxiliary Inversion"[6] I argued that in such cases an underlying difference in syntactic constituent structure is responsible for both the scope difference and the phrasing difference.

These two treatments of the same facts are primarily concerned with the nature of semantic interpretation rules, but they also suggest different views of intonational phrasing and its place in grammatical theory. Many other problems in syntax and semantics lead us into areas of data where our description must (at least implicitly) make claims about the theory of intonation.

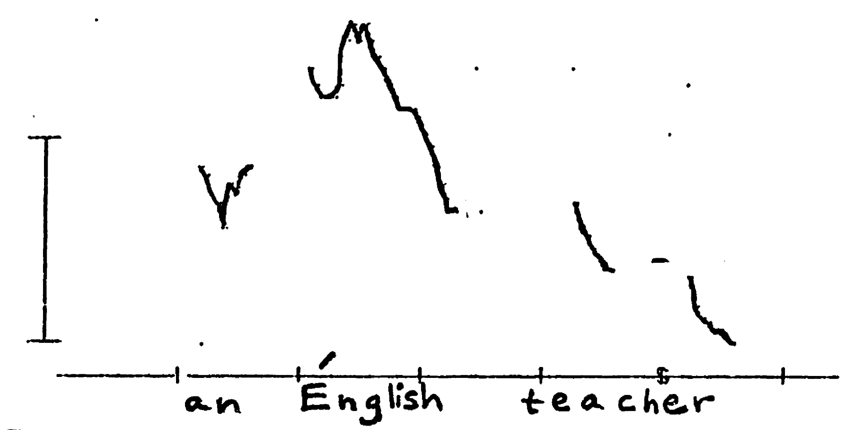
1.2.4 Mixed Cases

Although we have been treating stress, tune and phrasing as separate issues, in many cases a change in one has some effect on the others. Thus the stress difference in 1.2.1/1 will normally change the pitch contour --

1.2.4/1a



1.2.4/1b



Here the tune is in some sense the same (both are examples of the "declarative" contour previously mentioned in reference to 1.2.2/1a), but the association of the tune with the lexical phonology is clearly different. It is usually the case that a change in stress pattern results in this sort of change in pitch contour.

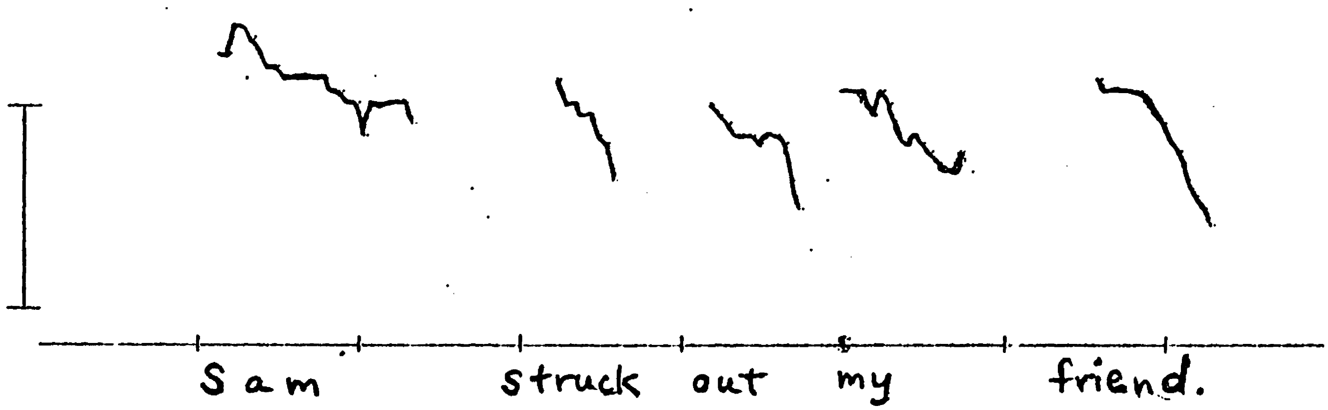
When we change intonational phrasing, as in example 1.2.3/1, both the stress pattern and the pitch contour are usually affected.

```
                        2         3      1
1.2.4/2a               Sam struck out my friend.
                        2         1        2
      b                Sam struck out, my friend.
```

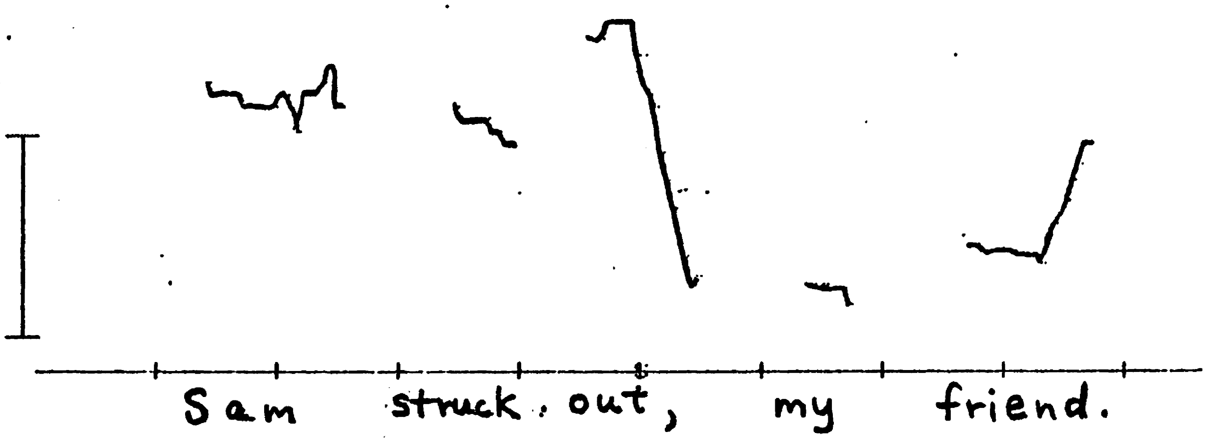
Whatever the precise numerology ought to be, in the example with my friend as the object, the highest stress is normally on friend, whereas in the example in which my friend is a vocative, the highest stress is on out.

Likewise, the two cases will have rather different pitch contours -- here is one possible version of the distinction:

1.2.4/3a

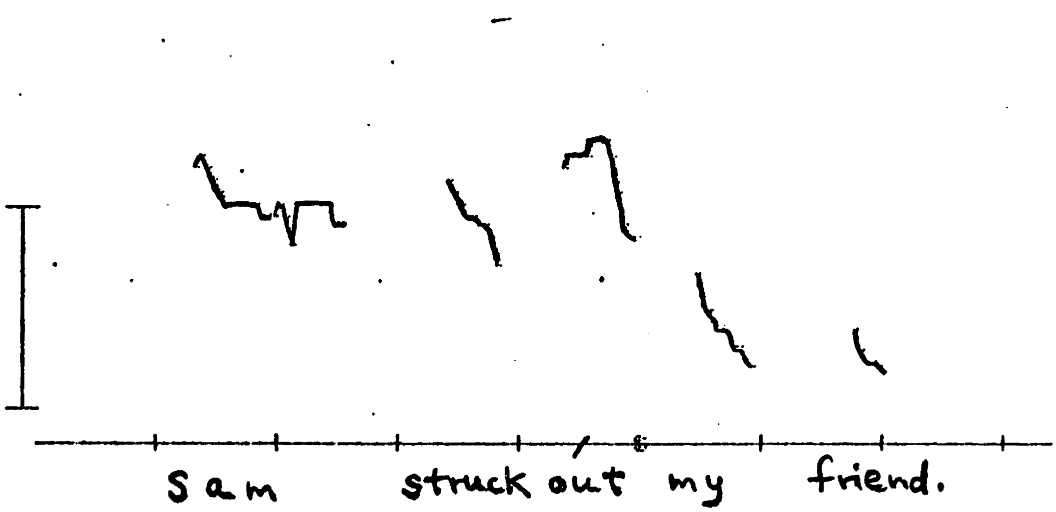


1.2.4/3b



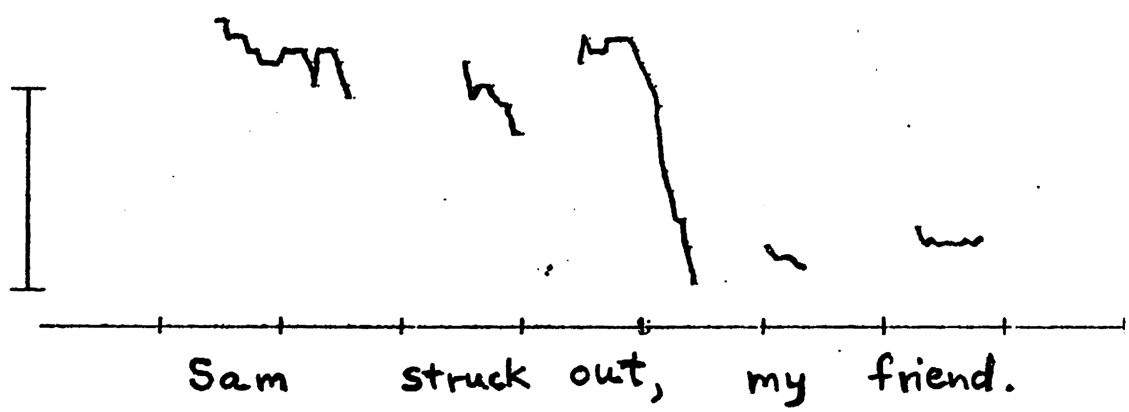
Another aspect of the interaction between stress, tune and phrasing can be seen in the fact that 1.2.4/3b is rather different from what we would normally find if my friend were object rather than vocative, but struck out nevertheless received the main stress (because of being contrastive, and/or because my friend was redundant).

1.2.4/4



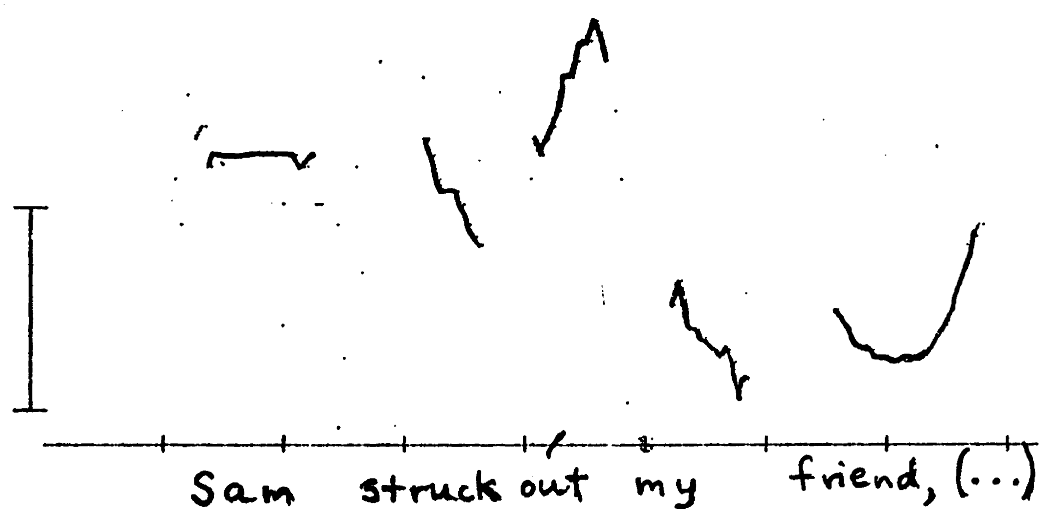
It is possible for the vocative to lack the terminal rise:

1.2.4/5



It is also possible for the case in which <u>my friend</u> is direct object, to have such a terminal rise, for example in the environment "Sam struck out my friend, but it took seventeen pitches":

1.2.4/6



However, in all of these cases the vocative/object distinction remains perceptually clear, at least in part because of the characteristic pitch contours of the word out -- when the "comma" follows, out falls to a relatively low pitch, whereas in the cases in which an object noun phrase follows, out remains relatively high, with a subsequent low pitch on the object. Thus the effect of intonational phrasing on pitch contour cannot entirely be considered a secondary consequence of its effect on the stress pattern.

We have seen that different tunes can be associated with a given string of words without affecting the stress pattern or intonational phrasing of that string; when we change the stress pattern, we may keep the tune constant in some systematic sense, but the particular way in which the tune is associated with the lexical phonology, the observed pitch contour of the utterance, will usually be changed; and when we change the intonational phrasing, both the stress pattern and the pitch contour will in general be affected.

Most of the cases which come up "in the field," so to speak, in the course of investigating some intonational phenomenon, involve interactions of the kind we've just discussed. Giving some account of these interactions is therefore an important goal for intonational theory.

1.3 Some Preliminary Postulates.

In order to sharpen up our statement of the problems which we'll attempt to solve in the chapters to come, it will be helpful to make explicit the assumptions on which the inquiry will be based. Some of these assumptions are quite uncontroversial; others will be motivated briefly in this section; a few will simply be assumed. In general, the only persuasive argument for such principles is that a theory based on them works; therefore, our goal at this point is not to persuade, but simply to explain.

Postulate #1: Lexical entries in English are not specified for tonal features, but tone is a linguistically significant category in English.

By tonal features we mean the features which specify the underlying form of what we have been calling tunes. Thus postulate #1 says that English is not a tone language, but does have tonal features which are capable of being varied in a linguistically significant way.

Postulate #2: Tonal features are not inherent in the segments which make up the phonological representation of an utterance in English, but have independent identity.

We have observed that words in English do not come from the lexicon with any tonal specification. Postulate #2 says that however the tonal specification is achieved, it does not have the effect of simply adding tonal features to the feature specifications of the segments of the phonological string; the "tune" retains a separate identity.

This point of view has been defended quite persuasively for certain African tone languages in work by Leben, by Williams and by Goldsmith.[7] This aspect of their theories carries over quite well for English. One of its primary benefits is that it allows for a single tonal entity to be associated with several phonological segments, or for a single phonological segment to correspond to several tonal entities. This idea is implicit in many of the traditional theories of English intonation (e.g., that of the British school).

Postulate #3: Tonal representations consist of well-ordered strings of segments.

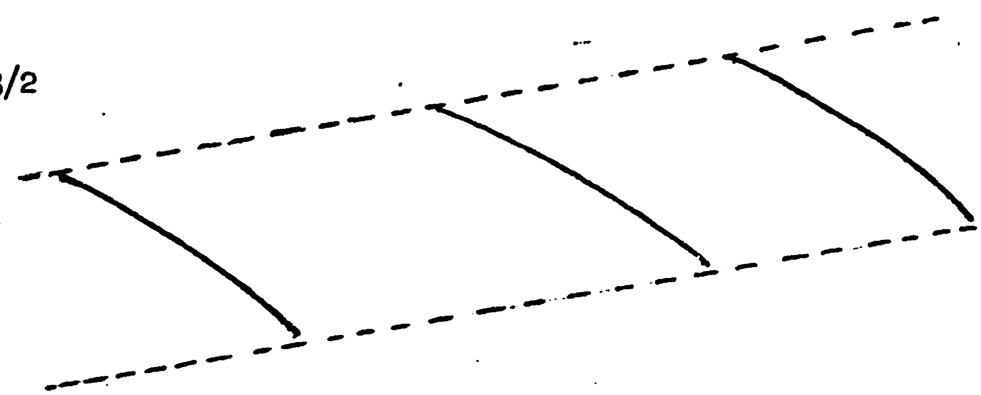
If one makes the segments large enough, this could hardly fail to be true; however, it is certainly conceivable that tonal features could overlap in various ways even at the most abstract level. For example, where T_1 - T_4 are tonal features of some sort, we might have (schematically) some organization like:

1.3/1

$$\underbrace{T_1 + T_2 + T_3}_{T_4}$$

Analyses of this type have been proposed, e.g., by Bailey and by Crystal.[8] Their proposals might be exemplified in a case in which T_1 - T_3 is a feature such as falling, while T_4 is rising (this is not quite how they put things, and should be viewed as a sort of free translation into the viewpoint being developed here):

1.3/2



Phenomena of the sort they describe certainly do occur in English. However, I am not convinced that the classification they arrive at corresponds to any set of linguistically significant distinctions. Later on, we will consider some cases which may lend themselves to an analysis such as that in 1.3/1, and may therefore give us reason to withdraw or modify postulate #3. It will nevertheless stand us in good stead in the analysis of quite a wide range of material.

Postulate #4: The underlying segments of tonal representations are static tones such as Low and High. Kinetic tones are always to be analyzed as sequences of static tones.

This principle is quite controversial, and will not be defended in any detail in this section. There are three kinds of arguments which might bear on the issue. The first kind might be called the argument from the nature of features -- those who believe that phonological features in general express articulatory targets or configurations, rather than articulatory gestures or processes, will be inclined to believe that tonal features share this property.

The second kind of argument may be called the tonemic inventory argument -- here we claim that the observed set of linguistically distinctive tonal entities has characteristics which are predicted by a theory which analyzes them as being made up of static primitives. We might cite in this connection some observations made by Crystal[9] on the restricted shape of that he calls "complex" and "compound" "nuclear tones." To discuss his views here would take us too far off the track; when we return to the matter we will argue that he has captured a truth, one which constitutes a tonemic inventory argument for static tonal segments.

Thirdly, we might argue that a particular theory based on the hypothesis of static tonal segments is able to predict a wide range of observed intonational outputs on the basis of simple and consistent underlying representations, acted on by a set of well-motivated rules. This would be what we might call a <u>tonological derivation argument</u> -- chapter 3 may be seen as such an argument, which, in my opinion, is the only kind that is really persuasive.

The ideas which we have been developing lead us to a mode of representation which can be graphically portrayed in this fashion:

1.3/3

$$\begin{array}{c} S_1 \ . \ . \ . \ S_n \\ \\ T_1 \ . \ . \ . \ T_m \end{array}$$

On one level we have a conventional phonological representation, consisting of strings of segments S_1 through S_n (with the associated constituent structure). On a second, independent level, we have a tonological representation, consisting of a string of tonal segments T_1 through T_m. As yet, we have no basis on which to decide whether or not the tonal string has any further structure.

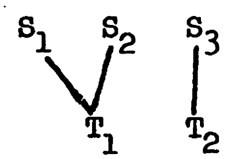
<u>Postulate #5</u>: The association between tonal and lexical (non-tonal) representations is established by linguistic rule.

It's of course conceivable that the two levels would represent temporally independent channels, proceeding in parallel on the model of simultaneously walking and chewing gum. This model is pretty clearly empirically false.

The <u>possibility</u> of association does not, of course, entail the <u>necessity</u> of association -- some tonal entities, for example, might float free, not being related to any particular elements of the "lexical" string, but simply presenting their ordering relative to the fixed points of those tones which do have lexical association. A demonstration of the <u>necessity of association</u> for tonal segments would, of course, be quite difficult, since it would require consideration of every systematically different case in the language. However, on the principle that we won't find rules that we aren't looking for, we will interpret postulate #5 to mean that all associations are rule-governed.

It remains to be seen what it means to "associate" the two types of representation. The view put forward in the "autosegmental" theory due to Goldsmith <u>et</u>. <u>al</u>.[10] is that the association should be viewed as a matching of segment to segment, which might be exemplified in a schematic way as follows:

1.3/4



We will suggest a somewhat different way of viewing the matter in chapter 2; however, the notation exemplified in 1.3/4 remains useful.

<u>Postulate #6</u>: Stress patterns exist independent of tonal representations, as a property of the <u>text</u> (the non-tonal phonological representation).

This principle is clearly implied by the observations which we made in section 1.2.2. We make it explicit here to avoid the possibility of misunderstanding.

1.4 The Problem Restated.

We are now in a position to state more precisely the questions which a theory of intonation must answer.

First, what are the underlying forms of stress patterns, tonal patterns (tunes) and patterns of intonational phrasing? This question is particularly important in the case of tonal patterns, since workable descriptive apparatus exists for the other two categories.

Second, by what rules is a given "tune" associated with its "text"? Why do changes in stress pattern or intonational phrasing affect this association in the way that they do?

Third, what is the nature of the phonetic representation of intonation, and by what rules is it related to underlying intonational form? In addition to tonal characteristics, we might in this connection want to say something about prosodically-conditioned duration and timing.

Fourth, how is the intonational system integrated into the theory of language as a whole?

2. The Association of Tune and Text.

Obviously, the investigation of how tonal patterns are associated with the words that bear them presupposes some theory about the nature of the things being associated. However, as we observed earlier, the only really convincing argument in favor of a particular theory about the underlying representation of tonal patterns would be a demonstration that such a theory, in conjunction with a theory of association rules, would work empirically.

Thus we are in an expositional bind -- we cannot expect to convince the reader of a particular "spelling" for tone patterns without a theory of association rules, and we cannot convincingly argue for a theory of association unless the reader is persuaded that the entities being associated have some reference to reality.

We can get around this difficulty by beginning with an investigation, not of normal speech, but of what might be called "chants." Since the tone patterns of these chants are clearly identifiable as a series of distinct notes, each clearly associated with a distinct portion of the text, we can arrive at a systematically interesting representation whose details are pretty much undebatable.

Of course, our argument may be faulted for dealing with examples that are perhaps as close to music as to speech. Our only defense, at this stage, is that the task performed in using these chants, the association of a given melody with a string of words, is exactly analogous to the task we have defined for intonational association rules; and that these chants can apparently be used, without training, by all native speakers of English, including those who are otherwise considered hopelessly unmusical. Once we have established the credibility of our approach through an examination of a pair of common chants, we will extend it to the intonations of normal, unchanted speech.

2.1 The Vocative Chant

There is a particular kind of chanted intonation which is used to call to people with whom the speaker is not in eye contact. It is discussed in Leben 1974, under the name of "vocative intonation." (Leben attributes the initial observations to R. Oehrle). Some of the examples below are taken from this source, as are the generalizations given in 2.1/2, although these last are given in a modified form.

The "tune" of the vocative chant consists of three pitches, of which the first is optional, while the second and third are obligatory. The third pitch is fixed a minor third below the second.

This interval is prominent in English chants, and apparently in those of other languages as well. It seems to be a very natural interval for people to sing, despite being mathematically a quite complex proportion, and consequently not representing any salient property of the overtone series. I've been told by several music teachers that it's the one interval that everyone knows how to sing without working at it.

The relation of the first, optional pitch to the others is not so precisely fixed, although it is always lower than either of those that follow. I think that intervals of either a fourth or a fifth below the following tone are fairly natural.

The second, highest tone is in fact always rising in its initial portion, a fact which we will return to later. For now, we will follow Leben in describing the tune of the vocative chant as (L) H M, using capitalized initial letters to refer to the tonal segments low, high, and mid, and parenthesis to indicate optionality.

The vocative chant also has a determinate rhythmic pattern, which, for the moment, we will ignore.

Some examples:

2.1/1a Alonzo L H M	1b Aloysius L H M	1c Sandy H M	1d John H M
1e Pamela H M	1f Tippecanoe L HM	1g Abernathy H M	

Certain generalizations clearly hold for tune-to-text associations in this chant.

2.1/2a The high tone is associated with the main stress of the text, and with any syllables which intervene between the main stress and the point at which the mid tone is associated.

2b If there are any syllables preceding the main stress, the low tone is associated with them; if no such syllables exist, the low tone does not occur.

2c If there is a secondary stress in the portion of the text following the main stress, the mid tone is associated with it, as well as with any following syllables.

2.1/2d If the syllables following the main stress are all unstressed, the mid tone is associated with the last of them.

2e If nothing follows the main stress, then that syllable is "broken" into two distinct parts, the second of which receives the mid tone.

Violations of these generalizations produce ungrammatical results, or at least results which can be understood only on the basis that the stress pattern has been arbitrarily altered in such a way that the generalizations will hold.

Some examples of ill-formed associations:

2.1/3a *Alonzo
H M

3b *Aloysius
L H M

3c *Pamela
H M

3d *Abernathy
H M

3e *Abernathy
H M

3f *Tippecanoe
L H M

2.2 <u>The Children's Chant</u>.

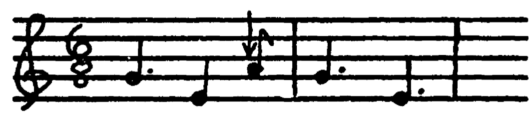
There is a ditty which is known to all English-speaking children, and therefore to most English-speaking adults. I don't know whether it is more general. It is used for taunting, exulting, singing certain nursery rhymes, and perhaps in other ways. Its most familiar instantiation is perhaps on the taunting nonsense string "nyah, nyah, nyah, nyah, nyah." It has both a fixed melody (the intervals being quite exactly defined) and a fixed rhythm.

Since the constraints on tune-text association (at least in their raw, descriptive form) are considerably more complex for this

Children's Chant than they were for the vocative chant, and since they are crucially related to the rhythmic pattern of the chant, we will give our initial examples in musical notation, so that the rhythm as well as the melody will be represented.

In its minimal form, the Children's Chant consists of five notes, arranged in the way given below:

2.2/1



In addition to this "lilting" compound-time version, a more foursquare rendition is also possible:

2.2/2

I find the version in 2.2/1 more natural, but others tell me that only 2.2/2 is possible for them; on the basis of a very limited sample, it seems possible that there is some geographical distribution of the variants. Different sorts of text may also influence the choice between "lilting" and "square" rhythms.

In both cases, we have two pairs of notes located a minor third apart, as in the vocative chant, with the higher note of each pair on the downbeat of its measure, while the lower note is on the second, weaker beat. Between these two pairs there is a sort of "grace note," which is always a submetrical interpolation into the rhythm established by the other four notes, whatever its actual time-value. The pitch of this grace note seems to be three-quarters of a tone above the higher of the other two notes -- the meaning of the downwards arrow over the <u>a</u> in the above example is that the actual pitch is somewhat flat of that note. Instrumental analysis of some native-speaker renditions of the chant suggest that in most cases this grace note actually has a pitch which moves through the range between <u>a</u> and <u>g</u>, or their transposed equivalents.

The nonsense-taunt version of this chant is performed exactly as notated in 2.2/1 (or 2.2/2), with each note being sung on the phonetic syllable [ñæ̃]. These five notes are obligatory -- optional positions may be added in two ways:

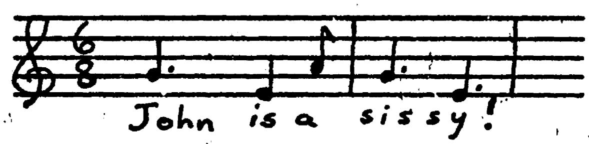
(1) An eighth-note upbeat may be added to the beginning of the chant. Its pitch is normally the same as that of the following downbeat, although in certain circumstances other pitches are possible.

(2) Any of the notes (including the optional upbeat) may optionally be subdivided. In the version most natural to me (given in 2.2/1) the subdivision of the dotted quarters produces the sequence ♩ ♪ if it is a binary subdivision, and the sequence ♪ ♪ ♪ if it is ternary. Subdivisions beyond this point are a little awkward, but ♪ ♪ ♫ is conceivable, and the optimal eighth note upbeat, as well as the obligatory eighth-note "grace note," can certainly be subdivided as ♫.

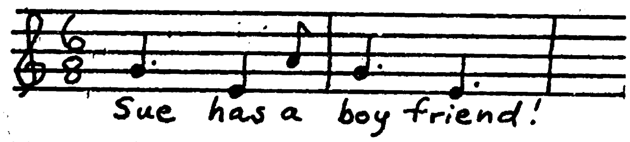
In all cases, subdivisions of the five obligatory notes that define the basic pitch sequence of the tune, retain the pitch of the note which they subdivide.

It is interesting to consider how various texts can be associated with this tune (and its rhythmic pattern). Some sequences of five syllables can be associated with the five obligatory notes of the tune on a one-to-one basis:

2.2/3a



3b



However, in other cases a five-syllable text must have two notes assigned to one of its syllables, while another note must be subdivided in order to carry the overflow:

2.2/4a

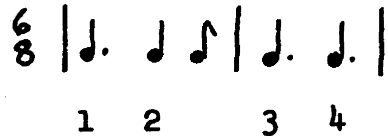
*


4b

The association of tune and text in 4a above is impossible, while that in 4b is fine. One obvious problem with 4a is that a metrically strong position, the downbeat of the second measure, is associated with a stressless syllable, while the following position, which is metrically weaker, is occupied by the main stress of the text.

To facilitate discussion, let's number the "beats" of the two measures of the Children's Chant as follows:

2.2/5



Now our observation about the difficulty with 2.2/4a can be expressed by saying that position 3 is metrically stronger than position 4, but *-sius*, occupying position 3, is unstressed, while *stinks*, occupying position 4, is stressed and indeed strongly stressed.

However, it is certainly not required that position 4 be unstressed -- in 2.2/3b it was occupied by *friend*, which bears stress. Nor does it seem to be required that position 3 be stressed -- examples like the one below seem quite possible:

2.2/6

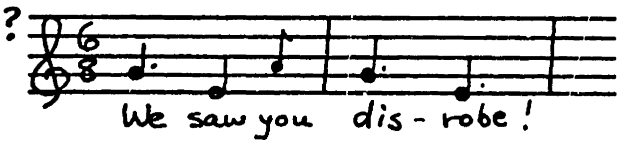
Nor does it seem to be correct to say that the syllable occupying position 3 must be (linguistically) more stressed than the syllable occupying position 4, since they can be the same syllable, as in 2.2/4b or 2.2/6.

The correct generalization seems to be that the syllable occupying position 4 may not be linguistically higher stressed than the syllable occupying position 3. Schematically, $4 \not> 3$. When the situation is *position 4 stressed*/ *position 3 unstressed*, as in 2.2/4a, the result is absolutely impossible. When both are stressed, but the syllable occupying position 4 has a higher stress than the syllable occupying position 3, either as a matter of word-stress ($\underline{\overset{3}{\text{mac}}\text{a}\overset{1}{\text{roon}}}$, $\underline{\overset{3}{\text{dis}}\overset{1}{\text{robe}}}$) or as a matter of phrasal stress ($\underline{\overset{2}{\text{big}}}$ $\underline{\overset{1}{\text{nose}}}$, $\underline{\overset{2}{\text{bag}}\text{gy}}$ $\underline{\overset{1}{\text{pants}}}$) the result t seems sometimes well-formed, sometimes questionable:

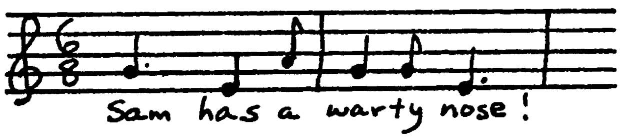
2.2/7a

7b

??



7c



7d

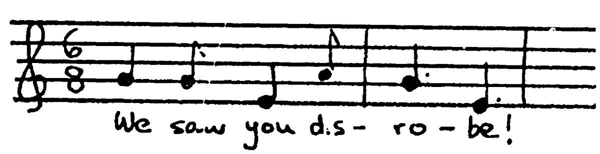
?

7e

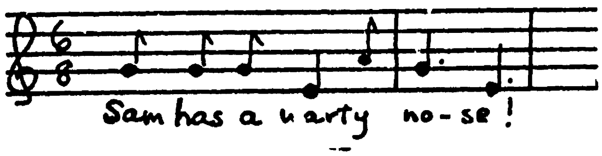
The tendency is certainly to prefer versions in which the highest stress falls on position 3:

2.2/8a

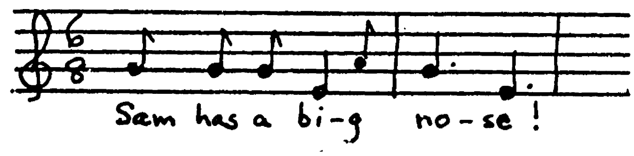
8b



2.2/8c



8d



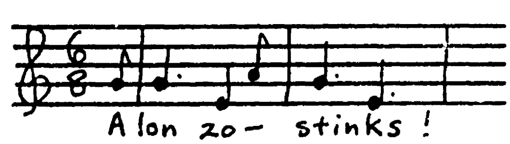
8e

We will leave the issue of relative stress aside for the moment, and interpret the formula $4 \not> 3$ to refer only to the distinction <u>+ stress</u>/ <u>- stress</u>. Since positions 1 and 2, the down-beat and second beat of the first measure, are metrically in the same relationship as positions 3 and 4, we would expect the same restrictions on textual association to hold there as well. In other words, we expect to find that $2 \not> 1$, that a stressed syllable cannot occupy position 2 if an unstressed syllable occupies position 1. This is indeed true:

2.2/9a

* 

9b



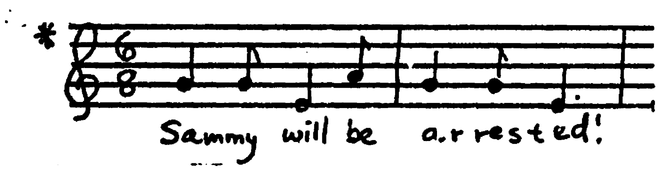
We find that 2.2/9a is impossible -- the text must be distributed as in 2.2/9b.

Since any additional expansions of the "beat," notes interpolated in between the notes occupying positions 1, 2, 3 and 4, are metrically weaker than the four "beat" positions, we expect to find a similar restriction holding for textual association with "beat" and "sub-beat" positions:

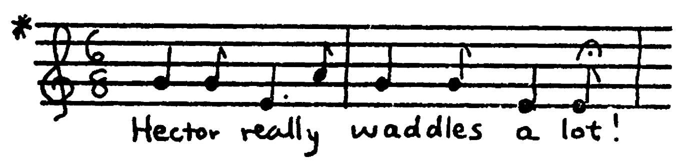
2.2/10a

10b

10c



10d



All of the specific rhythmic patterns in 2.2/10a-d are perfectly possible versions of this chant -- the reader should experience no difficulty in devising a text appropriate to each one. What is ill-formed in each case is the way in which the stressed and unstressed syllables of the text are matched with the metrically strong and weak positions. The first beat in 10a, the second beat in 10b, the third beat in 10c, and the fourth beat in 10d bear unstressed syllables, while one of the notes metrically subordinate to the beat bears a stressed syllable.

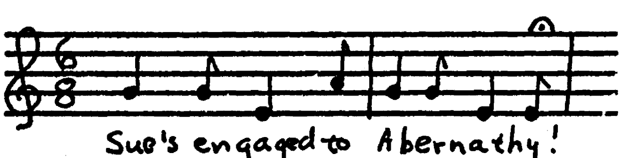
Thus, we can sum up all of the restrictions on text-tune associations we have so far observed for the Children's Chant in a single well-formedness condition:

2.2/11 For any two positions which are in the metrical relation *strong* to *weak* (e.g., *downbeat* to *offbeat*, *beat* to *expansion of the beat*), it is not permitted for an unstressed syllable to occupy the strong position if a stressed syllable occupies the weak position.

We are assuming, of course, that the metrical structure presupposed by this well-formedness condition is also associated in some way with the sequence of pitches characteristic of the Children's Chant. All of this will be expressed more precisely in the next section, when we have proposed a formalism for the representation of metrical patterns.

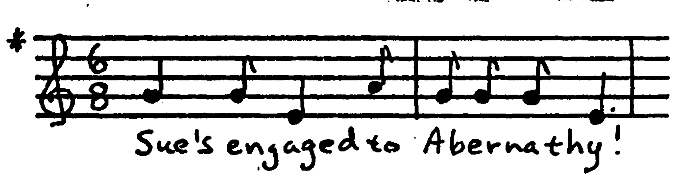
We may now observe that many of the restrictions on text-tune associations noted for the vocative chant are predicted by 2.2/11, and thus apply to the Children's Chant as well. For example:

2.2/12a



12b

12c



The reader may recall, incidentally, that misassociated examples of the vocative chant, as in 2.1/3a - f, required an inappropriate rhythm as well as an inappropriate text-tune association.

We pointed out, in the case of the vocative chant, that when the main stress of the text is followed by a series of unstressed syllables, the high tone goes (as usual) on the main stress, while the mid tone goes on the last of the unstressed syllables:

2.2/13a Pamela
H M

13b *Pamela
H M

If some unfortunate individual were named "Knowledgeable," the vocative chant version of his name would demonstrate that this remains true across three unstressed syllables:

2.2/14a Knowledgeable
H M

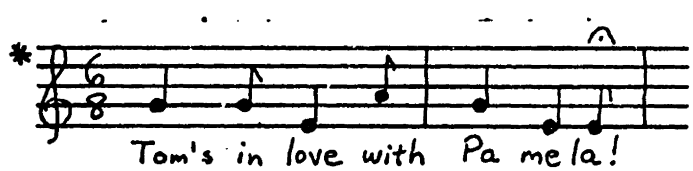
14b *Knowledgeable
H M

14c *Knowledgeable
H M

The well-formedness condition in 2.2/11 does not (as presently stated) predict this set of facts; analogous facts nevertheless may be observed in the Children's Chant:

2.2/15a

15b

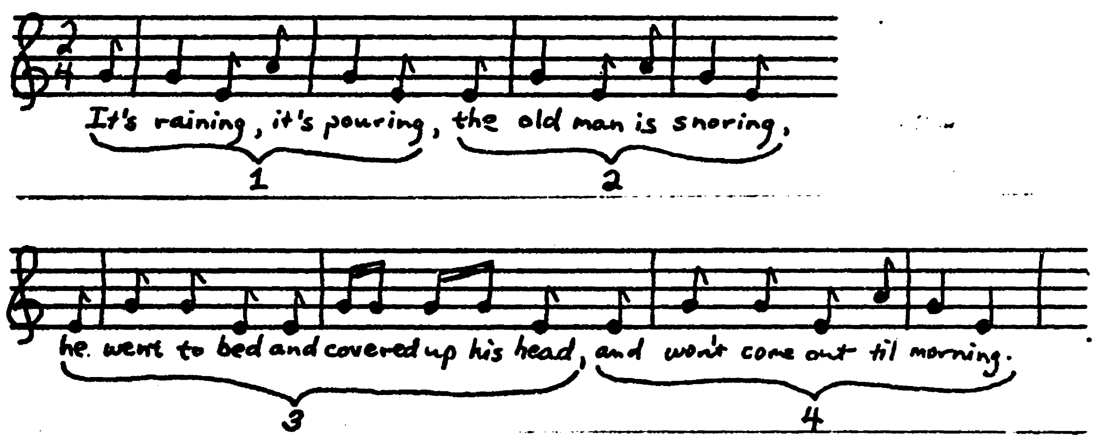


This pattern of association will be explained in due time. For now, we will be content to take note of it.

2.3 Some Extensions of the Children's Chant, and a Comment on it.

The two-measure Children's Chant has been expanded, in the traditional musical version of certain nursery rhymes, into an eight-bar ditty. For example (this particular case is in the "square" $\frac{2}{4}$ rhythm for me):

2.3'1



Of the four two-bar phrases which make up this song, the first, second and fourth are our old friend the archtypal Children's Chant. The single somewhat new element is that the pitch of certain upbeats "assimilates backwards" to the preceding note, rather than forwards to the following down beat (e.g., the pitch of "the" in "the old man is snoring").

The third two-bar phrase is a variant form, in which the "grace note" is lacking. This way of constructing a musical phrase (A A A' A) is not uncommon, although the limitation to three notes is a little more primitive than most.

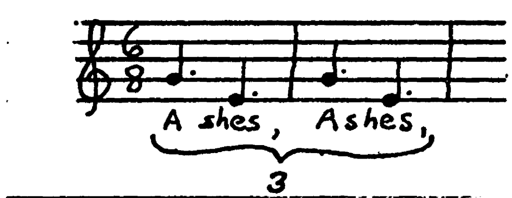
Another eight-bar nursery tune, quite similar in form to the one we've just examined, has a fascinating twist to it. The tune in question is "Ring around the Roses." It's employed by children in connection with a circle game, involving dancing around in a ring while selected individuals weave in and out; eventually everyone tumbles dramatically into a heap in the center of the ring. I must confess that I've forgotten the details -- the falling-down part is what I concentrated on, I think.

In any event, the falling-down part remains rather interesting, even to those of us whose tumbling days are largely behind us. Consider the song:

2.3/2

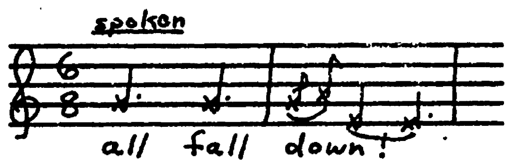
The first four-bar phrase is identical to the first four bars of "It's raining, it's pouring" -- two repetitions of the Children's Chant. The third section is the same (grace note-less) variation that we saw in the earlier song:

2.3/3



As the fourth two-bar unit, we confidently expect a repetition of the familiar Children's Chant. Instead, the child (that wise child who, it now occurs to us to remember, was able to learn to use this chant productively after hearing it perhaps once) stops singing and says:

2.3/4



It is hardly necessary to point out that the introduction of unchanted speech, in this way, into a primarily chanted song-structure means that the child is prepared to accept speech as the (so to speak) aesthetic equivalent of the chant. We will argue shortly that this equivalence is in some ways a very deep one.

We have examined the restrictions on Children's Chant tune-text associations in a format rather different from the one we used in the case of the vocative chant. We could have represented the underlying tune as

2.3/5 H M $\overline{\text{H}}$ H M

(where $\overline{\text{H}}$ is an extra-high pitch for the "grace note," while H and M are as they were before). This is undoubtedly in some sense a correct representation.

However, it is not of much help to us in establishing the principles of association for this chant -- these principles, we have discovered, crucially make reference to a metrical structure, to the fact that certain positions are "strong" while others are "weak." This metrical pattern has no necessary connection to the series of pitches which make up the Children's Chant tune -- the same pitches could be metrically organized in a different way, while the same metrical structure could be associated with a different string of pitches (as indeed it is in the familiar refrain "shave and a haircut").

In other words, we have assumed that the viewpoint mentioned in reference to 1.3/4 is, as promised, wrong. Instead of a picture in which a level of non-tonal segments and a level of tonal segments are associated with each other by linguistic rule, we have assumed a picture in which the tonal and non-tonal levels are each independently associated with a metrical pattern, and only through the intermediary of this third level achieve an association with each other.

We have yet to demonstrate that these two theories are empirically distinguishable, or that the metrical theory is preferable (although the reader may develop some opinions on the matter by trying to devise a non-metrical theory for the Children's Chant). Nor have we shown that any putative conclusions are of value in the analysis of non-chanted intonation. If we are to make any progress towards such goals, "prius de Arsmetrica arbitror exponendum. . . sine qua quicquid inclitum nequit pertractari," as Odington so aptly put it; "first we must, I think, explain the theory of metrics. . . without which nothing worthwhile can be successfully studied."

2.4 A Metrical Theory of Tune-Text Association.

2.4.1 A Note on Terminology.

The use of the term meter or metrical pattern to refer to an abstract structure related in some way to linguistic objects, is sanctioned by a well-established tradition. This usage is made quite explicit, for example, in the work of Halle and Keyser.[1]

The term meter originated, in classical times, in a theory which was intended to cover both poetry and music. This theory viewed rhythms as infinite series of repeated patterns, in which the repeated unit was drawn from a limited set of possible feet. In order for these infinite patterns to be used (in poetry or music) they had to be "measured" or "metered," that is, finite sections had to be excised. The laws governing how such cuts could be made, and what further modifications in the resulting pattern were possible, constituted the theory of metrics. Rules governing the association of such "measured" patterns with strings of words were also, in general, included for convenience under the same rubric.

The usage of Halle and Keyser differs from this classical view in that a metrical pattern, for them, is simply a given, and is not necessarily a "measuring" of any particular "rhythm" (although there is, in this tradition, considerable discussion of the derivation of "families" of metrical patterns.[2]).

In modern musical theory (e.g., Cooper and Meyer) the term meter is generally applied to the theory of time-signatures and bar-lines (or, more properly, what is represented by them). The term rhythm is used to describe certain features of the grouping and organization of accented and unaccented beats.

Rhythm is a term of ordinary language, as well, while meter, in any sense close to our present topic, is not. Webster's (college edition) gives the first sense of the word rhythm as "flow, movement, procedure, etc. characterized by basically regular recurrence of elements or features...in alternation with opposite or different elements or features."

We will appropriate the phrase metrical pattern to our own usage, which will be defined and explained in the sections that follow. Our metrical patterns will be abstract structures related to linguistic objects, as in the usage of Halle and Keyser, although the nature and functions of the structures we will propose are somewhat different from theirs. Of the original sense of meter, as a measuring of an infinite rhythmic pattern, little will remain in our theory. Our metrical patterns are perhaps somewhat closer to the rhythms than to the meters of modern musical theory, but are not the same as either of them.

The previous discussion has been intended to forestall misunderstanding on the part of those who have some experience with other uses of the term metrical pattern. We will cover the same ground somewhat more thoroughly in section 5, The Rhythms of Speech.

Our goal in what follows is to give some content to the suggestion made in 2.3, that the association of text and tune in the Children's Chant (and, we will claim, in English as a whole) is mediated by a metrical pattern. In order to turn this metaphor into a theory, we need a specific proposal about what a metrical pattern is, and how it could perform this function of linking tune and text.

2.4.2 The Nature of Metrical Patterns.

We can get an initial idea of what is required of a theory of metrical patterns by examining the assumptions that underlie the well-formedness condition, 2.2/11, which we argued to hold for the association of a given text with the Children's Chant. In applying that condition to the data which it covered, we assumed the existence of a hierarchical organization of strong and weak positions. For example, we assumed that in each of the "measures" of the chant, the first beat is strong in relation to the second beat, while within each "half-measure," the initial position (the "beat") is strong in relation to any subsequent positions. We might represent this, with reference to a particular example, in the following way:

2.4.2/1a

1b

```
s       w
s w     s w
Joey    Davis
```

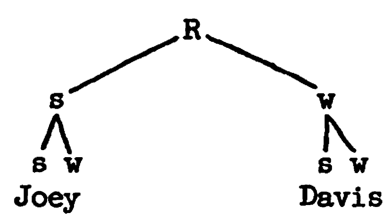
In 2.4.2/1b we have represented this hierarchical structure, for the first measure of the example in 1a, by two levels of alternating strong (s) and weak (w) positions. Thus the first syllable of Davis is strong on one level, but weak on another.[3]

The idea of hierarchical structures of this sort is quite common in the analysis of music. *Cooper and Meyer*, for example, observe that "rhythmic structure is perceived not as a series of discrete independent units strung together in a mechanical, additive way like beads, but as an organic process in which smaller rhythmic motives, while possessing a shape and structure of their own, also function as integral parts of a larger rhythmic organization."

We could say of the string "Joey Davis", in 2.4.2/1b, that its subparts *Joey* and *Davis* possess a shape and structure of their own, in each case the structure [sw] , while they also function as parts of a larger organization, *Joey* as *s* and *Davis* as *w* in the higher-level [sw] structure.

This organization of the metrical hierarchy can be represented by giving it a constituent structure:

2.4.2/2



The symbol *R*, for "root," in 2.4.2/2 is used because the nature of the constituent it labels is indeterminate, except by virtue of the role it might play in some larger metrical structure.

On this basis, we may suggest that *metrical patterns* are trees, with node labels *s* (strong) and *w* (weak). It is clear that these trees should be "oriented" (that is, constituents are ordered) and we will assume that they are "rooted" (that is, any such tree is always a single constituent at the highest level).

It should be noted, in passing, that the existence of hierarchical structure in metrical patterns does not prove that they are <u>trees</u> in the usual sense. Cooper and Meyer give many examples (of what they call "rhythmic structures") with improper bracketing, that is, where a given element may be shared between adjacent constituents. A simple instance is their analysis of the beginning of "Twinkle, twinkle, little star" (in one version):

2.4.2/3

Their — and ⌣ correspond to our <u>s</u> and <u>w</u>, respectively, and their horizontal brackets to our constituent structure. According to their analysis, the last note of the first measure is simultaneously part of a trochaic constituent with what precedes, and part of an iambic constituent with what follows.

We will assume that such circumstances either do not arise in language, or (more realistically) are always to be analyzed as a structural ambiguity, representing the existence of two equally possible metrical constituent structures for a given example. My reason for choosing this approach is partly that the formal properties of improperly bracketed trees are something of a mystery to me; more importantly, the use of trees of the (linguistically) normal sort represents a more restrictive hypothesis about what metrical patterns are, and thus deserves to be maintained until it can be shown to be wrong.

The notions "strong" and "weak" which serve as node labels in metrical patterns are essentially <u>relational</u>. A <u>strong</u> position is strong only by virtue of being associated with a corresponding <u>weak</u> position, not because of any inherent property of strongness; a <u>weak</u> position can be considered weak only because, in some metrical pattern, it plays the role of <u>weak</u> in relation to some <u>strong</u>. By this argument, it should be impossible to have metrical constituents such as [ww] or [ss] , and we will assume that this is so.

The opposition <u>strong/weak</u> is also essentially a binary one. We will extend this property to metrical constituent structure, and assume initially that only the binary constituents [sw] and [ws] are permitted. By this assumption, patterns like [sww] , [wws] and [wsw] must have some further structure. This assumption, and its consequences, will be examined in more detail as we proceed; for now, we will present in its favor only the argument that it is the most restrictive possible hypothesis.

We now have an embryonic idea of what metrical patterns are -- rooted, oriented trees with (at most) binary branching and node labels <u>s</u> and <u>w</u>, in which every binary constituent must be either [sw] or [ws] . Our next task is to specify how such structures can serve to associate tunes and texts; this task will be attempted in the context of a re-analysis of the vocative chant.

2.4.3 Introduction to a Metrical Theory of Tune-Text Association.

As we pointed out in section 2.1, the vocative chant has a determinate rhythmic structure as well as a determinate sequence of pitches. Some examples in musical notation are given below:

2.4.3/1a

1b

1c

1d

1e

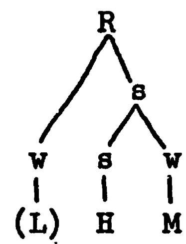
We represented the "tune" of the vocative chant as (L) H M. In the examples just given, the high tone always falls on the downbeat, a strong position, while the other two tones fall on weaker beats. Thus the lowest level of the metrical pattern corresponding to this tune must be:

2.4.3/2

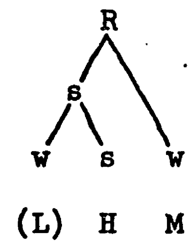
w	s	w
|	|	|
(L)	H	M

By our hypotheses that metrical patterns are rooted trees, and that branching is maximally binary, some further structure is required -- two possibilities seem to exist:

2.4.3/3a

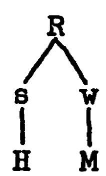


3b



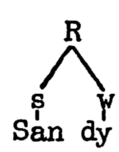
The fact that it is the initial low which is optional suggests that perhaps 3a is the preferable structure, but we have no sure way to choose between the two options, nor any evidence that we must definitively choose one way or the other. In any case, it is clear that without the optional low tone, the structure must be:

2.4.3/4



Now, let's consider the metrical structure of some possible tunes. The name Sandy has primary stress on its initial syllable; the second syllable is unstressed. Thus its metrical pattern is unambiguously this:

2.4.3/5



This rather simple structure is identical (except, of course, for its terminal elements) to the structure given to the vocative chant (*sans* initial low) in 2.4.3/4. So we have here exactly the sort of case we postulated at the beginning of the inquiry -- a *text* ("Sandy") and a *tune* ("H M") associated with *the same metrical pattern*. By our hypothesis, they ought thereby to be associated with each other. Since the vocative chant version of this text is in fact

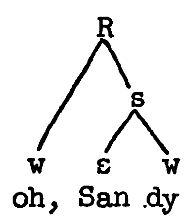
2.4.3/6

```
Sandy
| |
H M
```

the experiment may be considered a success; the empirically correct tune-to-text association has been achieved.

Used vocatively, the name "Sandy" could be preceded by the particle "oh." The main stress of the name remains the main stress of the phrase, in this case; and given the intuitively correct bracketing (which follows the position of word-boundaries) the metrical pattern for the phrase will be:

2.4.3/7



Since this is identical to one of the two possible metrical patterns for the full L H M vocative tune, the one given in 2.4.3/3a, we again have a text and a tune associated with a single metrical pattern, and we predict the text-tune association

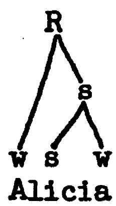
2.4.3/8

```
oh,  Sandy
|    | |
L    H M
```

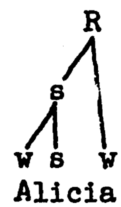
which is again empirically correct.

The stress pattern of the name "Alicia" tells us that the lowest level of its metrical pattern will be w s w. The higher-level structure is not determined by anything we now know -- either of the possibilities given below is permitted:

2.4.3/9a



9b



Intuition favors 9a, I think, but for our present purposes it doesn't matter, since both of the possible structures have counterparts in the scansion of the full L H M vocative chant, and the two versions yield the identical (and correct) association:

2.4.3/10

```
Alicia
| |  |
L H  M
```

In the three rather simple examples just considered, the number of elements in the tune and the number of syllables in the text have been the same, and their association has been on a one-to-one basis. Obviously this is not always the case. We have seen examples in which a single syllable of the text is associated with two notes of the chant, as well as examples in which several textual syllables are associated with a single tonal element.

In such cases, the metrical patterns initially assigned to tune and text will not be strictly identical, but they will be "congruent" in a sense that should become clear shortly. Inspection of some examples will suggest what this "congruence" of metrical patterns is, and how it should be used to define a tune-text association.

Suppose that our text for the vocative chant is a monosyllabic name like "John." As we have observed, the minimal form of the vocative chant tune is H M, with the metrical structure [sw] . "John," being a single syllable, has no internal metrical structure at all (at least in the sense of that term that we have been using). What happens in such a case, as we know from earlier examples, is that the monosyllable is "broken" into two parts, one for each of the tonal elements in the tune:

2.4.3/11

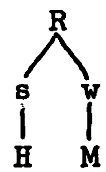
```
Jo - ohn
|     |
H     M
```

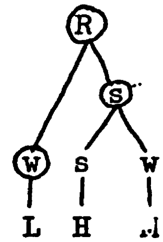
If the text has final stress, as in the name "McDuff", a similar process occurs:

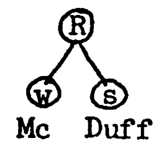
2.4.3/12

```
McDu - uff
/  |    |
L  H    M
```

The procedure of breaking a single syllable into a number of metrical positions (receiving the various notes of a tune) is called <u>melisma</u> in musical parlance. The melisma which occurs in the two examples just considered might be treated in the following way. The metrical "identity" of tune and text in examples like "Alicia" was a kind of <u>congruence</u> of metrical patterns -- one metrical pattern could be laid on top of the other, matching root for root, node for node, and (in those cases) terminal element for terminal element. Now consider what happens when we try to achieve such a congruence of metrical patterns for the vocative chant tune and the text "McDuff."

2.4.3/13a 

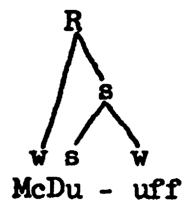
13b 

13c 

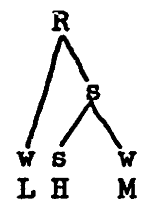
If we begin with the minimal version of the vocative chant, in 13a, and compare it to the metrical pattern of "McDuff," in 13c, root will match root alright, but the initial expansion is [sw] in one case, and [ws] in the other. The attempt must be aborted -- 13a is clearly not congruent to 13c. If we transfer our attention to 13b, we may again match root to root, and this time the initial expansion is [ws] in both cases, so that down to the circled nodes, the two metrical patterns are perfectly congruent. At this point we run out of text, so to speak, so that the superimposition is incomplete, but we have found no actual incongruence. We might then complete the match by "breaking" the stressed syllable <u>Duff</u>; alternatively, we might decide that our work is done, and that by achieving congruence down to the circled nodes in 2.4.3/13b-13c, we have simply linked a single syllable of the text to a tonal constituent with two terminal elements.

Since in the vocative chant, such syllables are clearly broken into two metrical units, we might adopt the former alternative for now, yielding the identical (in theory, superimposed) trees below:

2.4.3/14a



14b

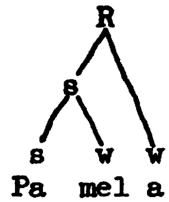


On this view, the "breaking" of a stressed syllable into two units (or perhaps more) would be an optional trick of scansion, performed at the point of assigning the text a metrical pattern. In the alternative view, according to which the superimposition of metrical patterns is considered successful if no actual metrical incongruence is found, and a terminal element of the text may thus be superimposed on a non-terminal node of the tune (or vice-versa), the observed "breaking" would have to be a secondary phenomenon, performed at some later stage of the intonational derivation. We will argue later that this second position is actually the correct one, that "melisma" is not in general a property of the underlying assignment of metrical patterns, but only a property of derived structures.

Now that we have given a preliminary account of some cases in which there is more than one tone per syllable, let's turn our attention to some cases where there is more than one syllable per tone.

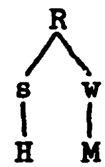
Consider the text "Pamela." Its first syllable is stressed, while the second and third syllables are unstressed; this means that the lowest level of the metrical pattern must be sww. Our hypotheses about the nature of metrical patterns will impose on this sequence the constituent structure given below:

2.4.3/15

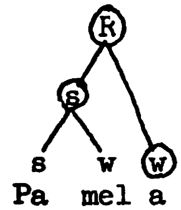


No other parsing is possible. In matching this metrical pattern against that of the vocative-chant tune, we are forced to take the "short form" H M, since the initial expansion of the metrical pattern of the text is [sw] .

2.4.3/16a



16b

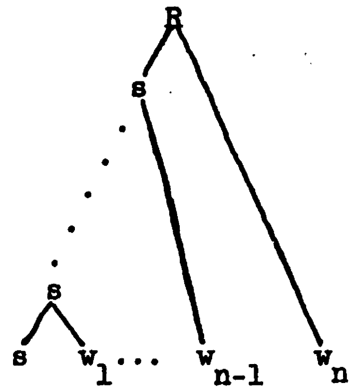


Thus we achieve congruence down to the circled nodes in 16b; at this point we run out of tune to work with.

Before proceeding any further, we should take note of an extremely important fact. The congruence established in the example presently under consideration implies that the mid tone of the vocative tune will be associated with the last syllable of "Pamela," and indeed with the last of any series of unstressed syllables following the main stress.

This is exactly the regularity that we noted in 2.1/2d -- "if the syllables following the main stress are all unstressed, the mid tone is associated with the last of them." Our proposal about the nature of metrical patterns entails that any sequence of the form $[sw_1 \ldots w_n]$ will always be parsed as indicated below:

2.4.3/17



Our proposal that text-tune associations are achieved by establishing congruence between metrical patterns entails that if the vocative tune is to be associated with a structure like that in 2.4.3/17, the terminal mid tone *will always be associated with* $\underline{w}_n$, (that is, with the last weak position).

Thus this previously noted generalization about text-tune associations (which also, we pointed out, holds for the Children's Chant) is no longer an arbitrary principle, which might as well have picked out the penultimate or the initial of the series of unstressed syllables, but an inescapable consequence of the most basic assumptions of our approach. This is striking confirmation that we are on the right track.

Returning to our consideration of example 2.4.5/16, in which we were matching the metrical pattern of "Pamela" to the metrical pattern of the vocative tune "H M," we find a situation rather similar to the one which we observed in the case of "McDuff." We have been able to superimpose the two patterns without finding any incongruence, but a single tonal element (the initial H) is superimposed on a textual constituent which contains two syllables ("Pame-"). The observed pattern of association is just what this implies:

2.4.3/18 Pamela

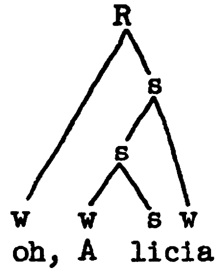
H M

The association of a single tonal element with several syllables may thus be seen as a kind of <u>anti-melisma</u>, in which the tonal element is "broken" into a number of metrical positions to match the metrical structure of the corresponding portion of the text. Again, we have a choice between two views. We could require the congruence of metrical patterns to be identity down to terminal elements, necessitating "anti-melisma" as an option to be exercised in the initial assignment of metrical patterns to tunes. Alternatively, we could consider a superimposition of metrical patterns successful (i.e. consider the patterns in question to be congruent) as long as no metrical incongruence exists, even if the metrical trees do not completely "cover" each other; on this view, "anti-melisma" would arise in the further derivation of a case in which a terminal tone is superimposed on a textual constituent which contains more than one syllable.

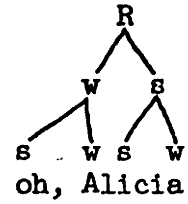
These positions are not at all equivalent; in particular, the first one is wrong. It seems elegant and appealing in the abstract, but would involve us in some nasty complications quite quickly.

An example of these complications is easy to find. Suppose our text is "oh, Alicia." Then, on the basis of the stress pattern and position of word-boundaries, we have a metrical pattern something like the one in 20a below:

2.4.3/20a



20b

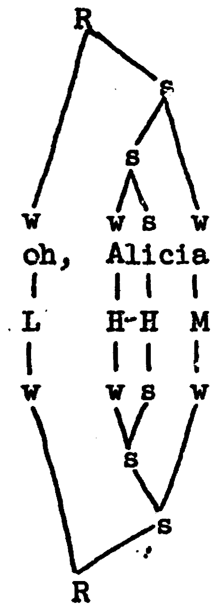


20b is another possible metrical pattern for the phrase, and one which might be preferred on the basis of its simple "alternating" structure; however, 20b has the definite drawback that it violates the natural constituent structure of the phrase, indicated by the word-boundary. The fact that we are inclined to write a comma between these words suggests that this boundary is all the more real.

We might imagine that some sort of cliticization process could take place here, involving an erasure of the word boundary and thus _permitting_ the structure in 20b. But it seems improbable that we could _rule out_ the structure in 20a as a possible metrical pattern for this text.

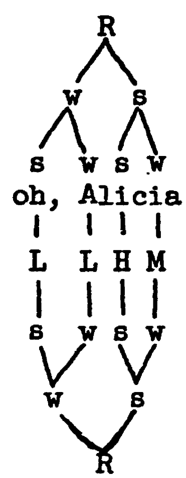
Given that 20a is at least a possible "scansion" of the text, it is easy to see that the portion of the tree in which the nodes are circled will match the metrical pattern of the vocative tune "L H M." But if we follow the position suggested above, in which "anti-melisma" will be invoked when such a match relates one tonal element to several syllables of the text, we get the following result:

2.4.3/21



The tune-to-text association thus produced is clearly wrong. Now, if we had chosen the option in 20b, we would have come up with the right match-up:

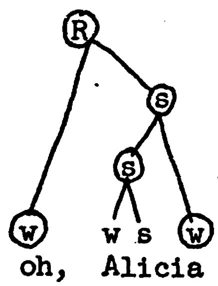
2.4.3/22



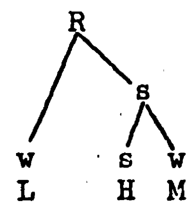
So we might try to find some way to insist, by main force or gentle persuasion, on the scansion in 20b. The prospects in this direction are not, I think, very bright.

A much more promising alternative emerges if we allow the scansion represented by 20a, but abandon the idea that "anti-melisma" is a property of underlying metrical patterns, and instead embrace the alternative treatment. In this alternative treatment, remember, we permitted the superimposition of metrical trees to be a congruence even if terminal elements of one tree were superimposed on non-terminals of the other. For the example under consideration, this will give us the following match-up:[4]

2.4.3/23a



23b



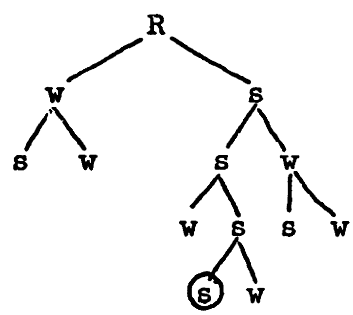
When we superimpose the metrical patterns of tune and text, the tonal tree "covers" the portion of textual tree in which the nodes are circled. The initial L tone is superimposed on the syllable <u>oh</u>, the terminal M tone is superimposed on the syllable <u>cia</u>, and the medial H tone (or more precisely, the <u>s</u> which immediately dominates it) is superimposed on the non-terminal textual constituent $\begin{bmatrix}\text{w s}\\ \text{Ali}\end{bmatrix}$. We now have the option of giving this superimposition an interpretation which will be consistent with the empirically observed pattern of tone-assignment.

We will easily accomplish this feat by taking note of the fact that all changes of pitch in the examples we have been considering (except for the grace note in the Children's Chant) take place on <u>beats</u>. The notion of "beat" was one which we employed in our earlier, intuitive discussion of tone assignment in the Children's Chant, but which we have up to now left out of consideration in the more formal treatment; the time has come to reintroduce it.

By the nature of things, beats are strong in relation to positions which are not beats. Thus in the metrical constituent [sw] , if a "beat" is to be defined, it must coincide with the strong element. Likewise, in the constituent [ws] , it must again be the strong element which is defined as the beat.

This much will be enough for "oh, Alicia," but later on we will be faced with the problem of defining the position of the "beat" in metrical constituents of arbitrarily great complexity, e.g.

2.4.3/24



The general definition (adopting the technical term "designated terminal element" in place of ordinary-language "beat") is this:

2.4.3/25 The designated terminal element of a metrical node N is that terminal element dominated by N which is reached by a path starting from N that intersects no nodes labelled W.

For all the kinds of metrical structures we are allowing, the designated terminal element of any given node always exists, and is always unique.

Now, we can express the observation that pitch-changes tend to occur on the "beat" by saying that when a single tonal element is linked (by superimposition of metrical patterns) with a metrical constituent of the text, it is to be attached to the designated terminal element of that constituent. Then, the linkage established in 2.4.3/23 between the H tone of the vocative tune, and the metrical constituent $\begin{bmatrix} w\ s \\ \text{Ali} \end{bmatrix}$ of the text "oh, Alicia," will result in the H tone being attached to the syllable li. Combined with the other associations established in 2.4.3/23, this gives us the following text-to-tune matching:

2.4.3/26

```
oh,  Alicia
|     | |
L     H M
```

We will now need some sort of assimilation or "tone spreading" rule to ensure that the initial syllable of "Alicia" will be low in pitch; that is, we need to assign by rule the dotted line in the diagram below:

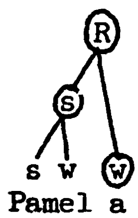
2.4.3/26

```
oh,  Alicia
| - -'  | |
L       H M
```

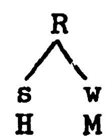
It will commonly occur that certain textual segments are "left out" in the text-tune association defined by congruence of metrical patterns, if that association is carried out in the way we have suggested. The issue of how to determine the pitch of these "free" syllables, in chants and in unchanted speech, will be discussed in detail in later sections. In most examples of chants, a left-to-right "tone spreading" of the kind schematized in 2.4.3/26 will give the correct output.

We can now complete the derivation of "Pamela," begun in reference to example 2.4.3/16a-16b. The superimposition of the metrical trees of text and tune, repeated below for clarity, results in the M tone of the tune being linked to the last syllable of the text, while the H tone will be superimposed on the constituent $\begin{bmatrix} \text{s w} \\ \text{Pame-} \end{bmatrix}$.

2.4.3/27a

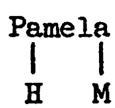


27b

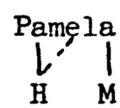


By the rule that we have proposed, this will result in the H tone being associated with the <u>designated</u> <u>terminal</u> <u>element</u> of $\begin{bmatrix} \text{s w} \\ \text{Pame-} \end{bmatrix}$, namely <u>Pa</u>. We thus have the underlying tune-text association given in 28a below; the left-to-right tone spreading rule will give the e derived association in 28b.

2.4.3/28a



28b



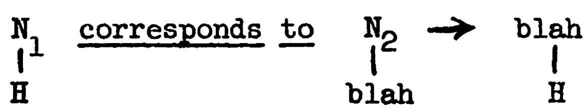
We ought now to consider whether this treatment of "anti-melisma" (the circumstance in which a single tone corresponds metrically to several syllables) can be generalized to cover melisma (the circumstance in which a single syllable corresponds metrically to several tones) as well. Suppose we adopt the following general principle :

2.4.3/29 **Metrical Association Rule:**

If two metrical trees T_1 and T_2 are congruent, and N_1 is a metrical node, $\in T_1$, which immediately dominates a terminal element Δ, and which corresponds to a metrical node $N_2 \in T_2$, then Δ will be associated with the designated terminal element of N_2.

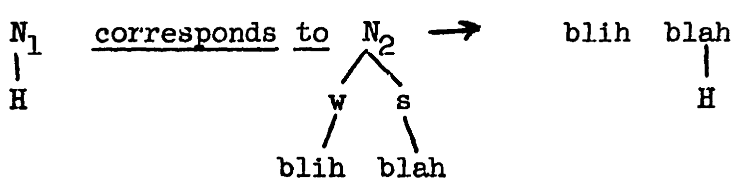
If N_1 and N_2 both immediately dominate terminal material, then their daughters will be associated by this rule:

2.4.3/30



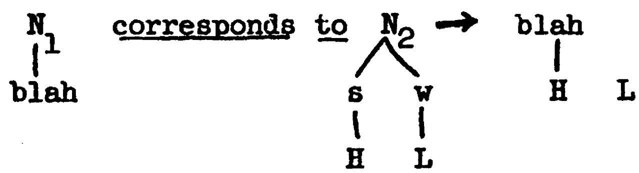
If N_1 is a single tone, while N_2 contains several syllables, then (as in the cases previously analyzed) the single tune will be associated with the designated terminal of N_2, while the rest of the syllables in N_2 will be free, their tonal association being defined by further rule if at all:

2.4.3/31



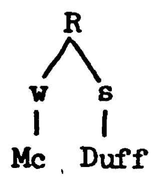
If N_1 is a single syllable, while N_2 contains several tones, then rule 2.4.3/29 will associate the single syllable with the designated terminal of N_2, while the rest of the tones in N_2 will be free, their textual association being determined by further rule if at all:

2.4.3/32

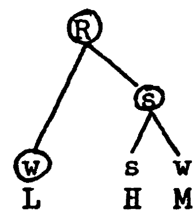


This last case is the one we are presently interested in. Let's consider how this mode of derivation would apply to the example we used earlier, "McDuff":

2.4.3/33a



33b

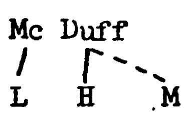


The metrical pattern of "McDuff" corresponds to the circled portion of the metrical pattern of "L H M." Two simple applications of the Metrical Association Rule will give us the underlying tune-text association

2.4.3/34

in which the terminal mid is a "free tone." To get the desired output, we will only need a rule or principle which will add another association line, the dotted line in the diagram below:

2.4.3/35



This may seem like an unnecessarily complex way to get a simple result. However, in more elaborate examples, especially those involving natural, unchanted speech, the resulting patterns of association will be quite complicated, and the system we are discussing, remaining (in essence) valid, will prove in the end to be a simple way to get a complex result.

Our system will be both conceptually simpler and empirically more adequate if we generalize the idea behind the Metrical Association Rule, that text and tune are associated on the basis of designated terminal elements, by introducing the concept of a Metrical Grid.

2.4.4 Metrical Grids.

We pointed out earlier that the vocative chant and the Children's Chant, in addition to a determinate sequence of pitches and a determinate association of these pitches with their accompanying texts, have a determinate "rhythm." We used this term in something like its ordinary-language sense, meaning both an abstract hierarchical organization of strong and weak elements, and also a temporal patterning of the terminal elements of this hierarchy. The fact that these two aspects of the "rhythm" are distinct can be seen in the difference between the "lilting" $\frac{6}{8}$ version of the Children's Chant and its "square" $\frac{2}{4}$ version -- these versions are identical in their hierarchical organization of strong and weak elements, but differ systematically in the temporal patterns that result.

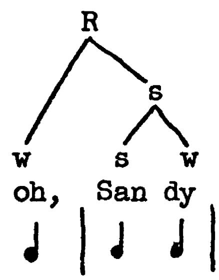
Our theory of metrical patterns has been intended to cover the first aspect of the "rhythms" which we perceive in these chants, the abstract hierarchical organization of strong and weak elements. We have said nothing about the temporal patterning of the terminal elements of our metrical patterns.

One rather obvious fact about the temporal realization of metrical patterns is that the ordering of their terminal elements is preserved as an ordering in time. Although obvious, this property of "sequentiality" is neither necessary nor trivial, and we will return to it later.

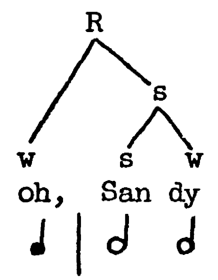
In addition to sequentiality, there are clearly other ways in which the temporal realization of the chants should be specified. Let's consider some examples:

2.4.4/1a

1b



1c

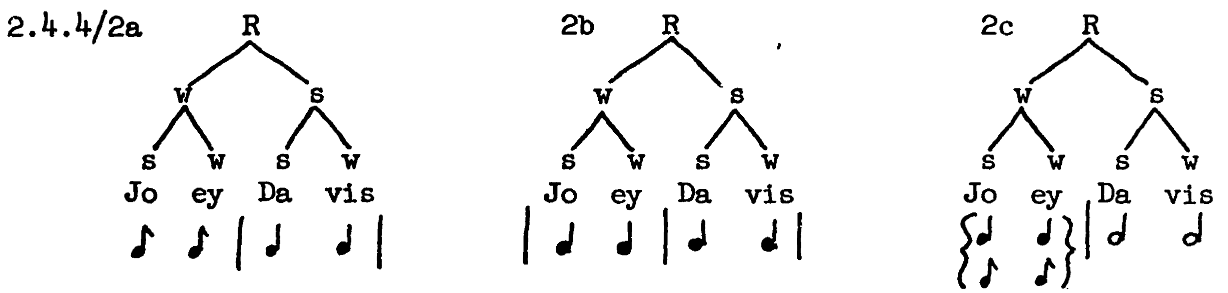


N.B. For ease of comparison in these examples, one should set a fixed tempo, say ♩ = 120.

In the simple case of the vocative chant with "Sandy" as text, the two syllables of the chant are realized on notes which are perceptually equal in length. They are not necessarily objectively equal in length, or even very close to it, but we may refer this problem to the domain of psychology, and proceed.

If we expand our text to "oh, Sandy," we find that we have several options -- two easily available ones are given above. We can perform the three syllables of the chant on three perceptually equal notes, as in 1b, or we can draw out the second and third notes so that each is twice the length of the first note, as in 1c. Of course, the downbeat remains on the designated terminal element "San-" in all cases.

In this case, the preferred option seems to be 1b, which gives all syllables equal length. However, it is easy to find a case in which we prefer to assign dissimilar lengths to the syllables of the text:



When the vocative-chant text is "Joey Davis," we would tend to assign equal length to the last two syllables, as before, but to make the first two syllables shorter, in such a way that they add up together to equal (perceptually) <u>one</u> of the longer syllables. This option is schematized in 2a above. In 2b we see the result of making all the syllables equal, while in 2c "Davis" is drawn out, preceded by either of the previous versions of "Joey." These versions are all possible, although the easiest or most natural one seems to be 2a.

The effect of these preferences is to maintain a pattern ♩ | ♩ ♩ | for the vocative chant -- apparently this pattern is for some reason the standard or desired form of the chant, at least in my usage, although deviation from this standard is easily tolerated.

Examples could be multiplied, but we have enough evidence now to be able to outline the system which assigns durations to the various terminal elements of a given chant. This system establishes a correspondence between *metrical patterns*, of the kind we have become familiar with, and what we might call a *metrical grid*. A metrical grid is a pattern which subdivides intervals of time, in the way that is familiar from musical notation; we can represent it with a table of the sort that is often used to explain this notation:

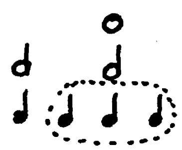
2.4.4/3a

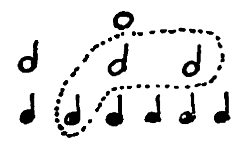
The different note-values are simply a convenient way to refer to the different levels at which we divide, subdivide, sub-subdivide etc. an interval of time. Non-binary (especially ternary) divisions are commonly permitted -- we leave them out, at the moment, purely for convenience of exposition.

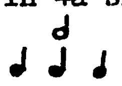
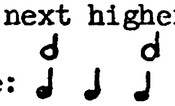
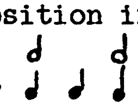
The Grid, like time, is in principle infinite -- we can pick any place in it to start, and have available sequences forward and backward, or up and down in the hierarchy, as we choose.

In a system of temporal patterns based on this premise, any sequence of temporal intervals must be assigned some structure as a fragment of the Grid. (This way of looking at things is neither logically nor humanly necessary -- various alternative ways of thinking about cutting up time, especially as they apply to language, will be discussed in the section on the Rhythms of Speech).

For example, the two sequences of temporal intervals represented for the vocative chant "oh, Sandy" in 2.4.4/1b-1c, represent the following fragments of the Grid:

2.4.4/4a 

4b 

Now, one might well ask how we know that the sequence ♩ ♩ ♩ in 4a should be aligned with the next higher level in this fashion:  , rather than in this one:  . The reason is simple: a position in the Grid which coincides with a position on a higher level is stronger than one which does not, and the second of the three quarter-notes corresponds to the strongest position in the metrical pattern of the chant. The rejected alignment  would cause the "main stress" of "oh, Sandy" to occupy a grid position weaker than the position occupied by the other two syllables in the chant.

Thus the Grid has its own intrinsic "stress pattern," which is not to be violated in the process of finding note-values (and their associated structure in the Grid) for the terminal elements of a metrical pattern. The reason for the existence of this "stress pattern" is quite simple. Imagine a single temporal interval, which we will designate as a half-note, 𝅗𝅥 . This temporal interval is to be performed (as a hum, if you like) by person A. We imagine dividing this interval into two parts, symbolized as ♩ ♩ . This sequence of temporal intervals is to be hummed by person B. Now, if we adhere to our original idea that the sequence ♩ ♩ is a "division" of the original 𝅗𝅥 (a state of affairs which is symbolized $\overset{\text{𝅗𝅥}}{\text{♩ ♩}}$), then the two performers will begin and end their humming together. In order to coordinate this, they must arrange to start at the same time. At the prearranged signal, person A begins his 𝅗𝅥 , and person B begins his first ♩ .

Some time later, person B must break off his first note and start his second one; however, person A has no particular duty at that time. Now, if person B had decided not to perform his second note, he still would have had to coordinate the beginning of his first note with his partner, but the same is not true of his second half-interval. Thus the note which coincides with the beginning of a higher-level note is, in a sense, equivalent to that higher-level note. This equivalence is even more marked if the "performance" is in terms of some medium without temporal staying power, such as hand-claps. If our two performers had clapped in a pattern like that below,

2.4.4/5	<u>person A</u>:	(clap)		(clap)		etc.
	<u>person B</u>:	(clap)	(clap)	(clap)	(clap)	

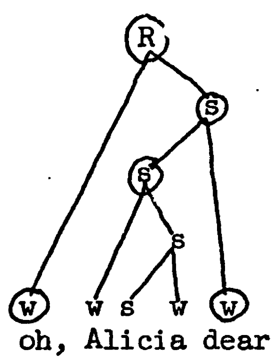
this meaning of the vertical organization of our grid diagrams would be even clearer.

Fundamentally, our explanation for the various possibilities to be found in the temporal realization of the vocative chant will be simply that some alignment with a metrical grid must be found in which the metrical pattern of the chant is not in conflict with the intrinsic stress pattern of the grid. This point in our exposition is not the place to go into detail about this process, however.

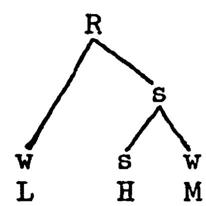
The idea of the metrical grid must be introduced eventually, in order to describe the facts of the temporal organization of chants, and also, we will argue, of speech. But one reason for introducing it at this stage in the proceedings is somewhat different -- we wish to point out its relationship to the Metrical Association Rule introduced in the preceding section, 2.4.3.

Let's re-examine one of the examples which led us to that rule:

2.4.4/6a



6b



As the reader will recall, we began our derivation by establishing a congruence bewteen the metrical patterns of text and tune -- 6b matches the portion of 6a in which the nodes are circled. The Metrical Association Rule then told us that the resulting pattern of association would be:

2.4.4/7

```
oh, Alicia dear
|     |     |
L     H     M
```

But the necessity of aligning the chant with a metrical grid tell us the same thing. There are various possibilities for assignment of note-values to this chant -- a few of them are given below, in musical notation and as grid fragments:

2.4.4/8a 8b 8c

Note: in these examples, we do not insert a time-signature for the same reason that it is not done in recitative-- a time-signature implies constraints on the structure of the metrical grid which don't exist in the context of natural speech, and the vocative chant partakes of this freedom. The Children's Chant, on the other hand, is constrained in the way implied by a time-signature.

These various versions all have in common that the strongest position in the chant, the syllable li (as well as the high tone) is always aligned with a grid position which is the strongest in the fragment. That this should be true is not surprising -- in any other state of affairs, by assumption, the metrical pattern of the chant would be in conflict with the inherent stress pattern of the grid, which would be antithetical to the fundamental nature of a metrical pattern. Suppose, for example, the strongest grid position were aligned with the first syllable of Alicia, as symbolized below (we can imagine the chant performed in a monotone, to eliminate the question of where the tune goes):

2.4.4/9

Then the main stress of the chant would be aligned with a much weaker grid position than the totally unstressed first syllable of the name "Alicia." It is clearly at odds with the nature of the relationship which corresponds to the node labels <u>strong</u> and <u>weak</u>, for the strong element to fall on a weaker grid position than the weak one.

A similar argument holds for the alignment of the tune of the chant -- the H tone can go nowhere but on the strongest position in the metrical grid. Thus, the fact that the H tone and the main stress of the text are associated with each other, follows from the fact that each of them individually must be associated with the same position in whatever fragment of the metric grid is utilized.

In other words, the fact that the chant, text and tune together, is aligned with a metric grid, ensures that the association of text and tune will proceed in the way specified in Rule 2.4.3/29, our "Metrical Association Rule." Therefore we are entitled to expect that when we develop a more precise theory of "grid alignment," rule 2.4.3/29 will be cast aside as an automatic consequence of an independently necessary process.

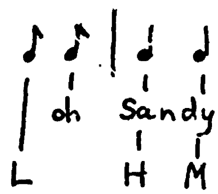
We might be tempted to go even farther, and to suppose that a theory of the alignment of texts and tunes with metrical grids is all we need, enabling us to dispense with the whole apparatus of metrical patterns, congruence, etc. To yield to this temptation would be a mistake.

The initial argument against the position which says "the Grid is All" is that a theory of grid alignment presupposes the existence of metrical patterns associated with the things aligned -- thus some independent theory specifying metrical patterns for texts and tunes is required.

But one might grant this point, allowing some way of indicating relative stress for texts and tunes (an SPE-type notation, for example) which would govern their independent alignment with a metrical grid, and thus with each other. This position would most crucially differ from the one we have been constructing in this chapter in that no notion of "congruence" between the metrical patterns of text and tune would be involved.

There are three arguments against this "congruenceless" position, and all of them are fairly strong ones.

1) The independent alignment of text and tune with a metrical grid will not always cause them to be aligned with each other. The "main stresses" will come out together, but beyond this point, there will in general be various feasible grid alignments in a given case, and it is not possible for tune and text to make their choices independently. For example, suppose that in "oh, Sandy," the text took the option in 2.4.4/1b, while the tune took the option in 2.4.4/1c, thus yielding:

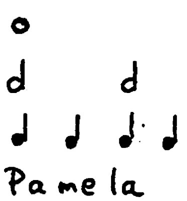
2.4.4/9 

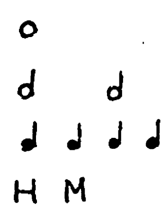
One might be able to give an interpretation to such an outcome, but the prospect is a messy one. Still worse, the independent alignment of tune and text will often result in wrong associations -- there is nothing to prevent the association "Pamela," for instance,

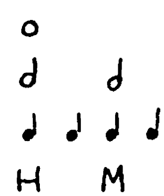
H M

since the stress pattern of the text,

100, can easily be metrically aligned as in 9a, while the tune can be given the alignment of 9b as easily as that of 9c:

2.4.4/9a 

9b 

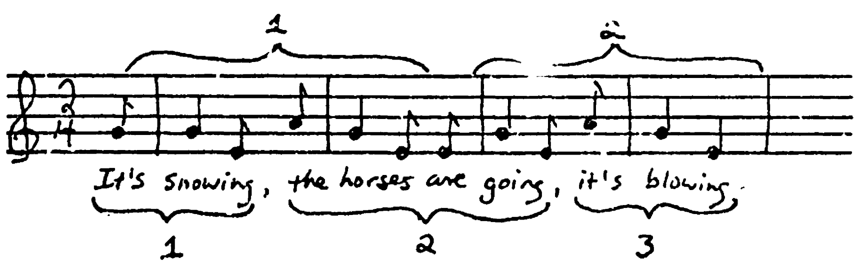
9c 

These difficulties for the "congruence-less" theory would be very hard to patch up. Out theory, as it stands, relies on the process of alignment with a metrical grid primarily to define temporal patterning of already associated texts and tunes (which is its proper function). The only role of grid alignment in tune-text association, for our theory, is to ensure that a single terminal element of text or tune is finally associated with the designated terminal element of a metrical constituent with which <u>congruence</u> has given it an initial association. This last task will always be accurately and unambiguously performed by the alignment process, as we have pointed out.

2) It is intuitively a fact that the "phrasing" of text and tune coincide. It is hard to imagine a quasi-polyphonic situation, in which the text could be phrased, for example, (W S) W , while the tune was phrased W (S W). This intuitively necessary coincidence in phrasing between tune and text is expressed in the requirement of congruence between their metrical patterns. Since it is clear that texts do have phrasing (based primarily on their syntactic structure), and since a tune such as that which underlies the chants we have looked at can certainly have a phrasing of its own, performed simply as a melody with no words attached, without the requirement of congruence there would seem to be no bar to divergent phrasing of tune and text.

3) In cases in which complex intonational patterns are created by stringing together simpler ones, the congruence theory will ensure that the association of these simpler units with the text does not conflict with the text's own structure. We will give many examples of this sort in our discussion of unchanted speech -- for purposes of exemplification at this point in the discussion, we might concoct an ill-formed nursery song, in which the textual phrasing is at odds with the structure of the tune:

2.4.4/10

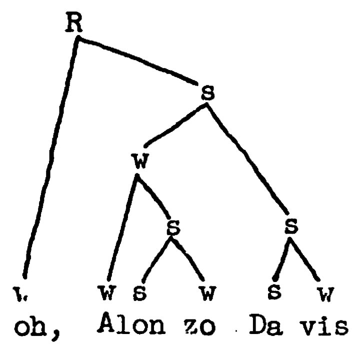


While this type of incongruence is musically possible, it bespeaks a sophistication beyond the reach of the nursery, and also, we will argue, beyond the reach of the intonational patterning of ordinary language.

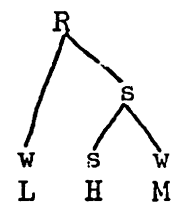
These arguments seem to me to be strong enough to motivate a congruence theory of tune-text association, at least as a working hypothesis. However, if the hypothesis is indeed to work, we need to take note of one further wrinkle.

As we have defined congruence, there are certain cases in which tune-text association will simply be impossible. Here is a simple example:

2.4.4/11a



11b



If our text for the vocative chant is "oh, Alonzo Davis," we face a problem. The expansion of the root is [w s] in both text and tune; however, the subsequent expansion of the s position in the text is [w s] , while in the tune it is [s w] .

The empirical fact is that it is possible to set this text to the vocative tune. Indeed, it would be unfortunate if it turned out that certain kinds of texts simply were "unmetrical," and couldn't be chanted (or, as we will see, said).

There are two ways in which we could arrive at a congruence between the trees in 11a and 11b (that is, superimpose them without metrical conflict): first, we could rearrange the constituent structure of 11a, and second, we could ignore or "skip over" a portion of it. Certain evidence, which we will now consider, suggests that both options are possible.

This evidence concerns some of the possible grid alignments of this chant. In particular, we observe that there are two types of such alignments -- those in which <u>oh</u> occupies a stronger position than the main stress of <u>Alonzo</u>, and those in which its position is weaker. Of course, in all cases the strongest position of all is occupied by the main stress of the phrase, "Da-".

2.4.4/12a

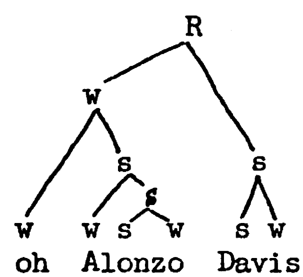
12b

In both of these cases, there are three quarter-note "beats," of which the second is the downbeat -- however, these three beats are arranged in a different pattern with respect to the text. In one case the pattern is "óh, Alonzo Dávís," while in the other case the pattern is "oh Alónzo Dávís."

We have seen other examples of optionality in the grid alignment of a metrical pattern (e.g., in 2.4.4/8). However, in these other cases, the placement and kind of stresses in the text (i.e., its metrical pattern) always remained constant -- in this case, the relative stressing of "oh" and "Alonzo" has changed.

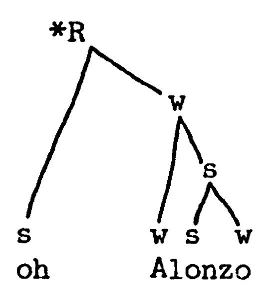
The metrical pattern for "oh, Alonzo Davis" given in 2.4.4/11a entails that "oh" is stronger than "Alonzo," since it is the weak half of the highest constituent in the tree, while "Alonzo" is a weak position expanded at a lower level. The word "oh," as is the case with most non-lexical words, is by its nature weak with respect to the element it modifies. Therefore, if we were to reparse it with "Alonzo" (a process which could be seen as a kind of "cliticization," although in this case it does not necessarily involve the deletion of the word boundary), in the resulting structure "Alonzo" would become the strong position:

2.4.4/13



Because of the nature of metrical patterns, a constituent [w w] is impossible; because of the nature of "oh," the pattern

2.4.4/14



is impossible; thus, given the decision to rearrange constituent structure, the structure in 2.4.4/13 is unavoidable.

Given that the constituent structure is rearranged, the grid alignment

2.4.4/15

is now compatible with the metrical pattern of the phrase. This grid alignment is, incidentally, in some ways preferable -- the relatively unimportant, nonlexical monosyllable "oh" no longer occupies such a strong position. In fact, this temporal organization of the chant (or some other one compatible with the metrical pattern in 2.4.4/13) does seem more natural than the one in 2.4.4/12a, in which "oh" occupied a stronger position than "Alonzo."

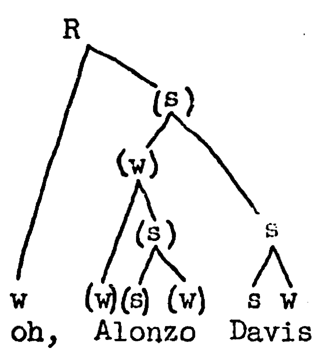
Thus the existence (and indeed preference) of this grid alignment suggest that the metrical reanalysis in 2.4.4/13 is possible. The resulting metrical tree can easily be made congruent to the metrical tree of the vocative tune.

We will find that this type of rearrangement of metrical constituent structure commonly occurs with certain types of words, specifically nonlexical elements such as "a/an," "the," "is" etc. We will examine this phenomenon, and its relationship to the phonologically motivated "cliticization" of these words, later on.

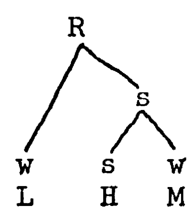
Let us now return to our discussion of how to deal with the inherent incongruence of the vocative tune and the metrical pattern of certain texts. We suggested that there were two ways to cope with this situation, either to reanalyze the metrical structure of the text, or (alternatively) to ignore or skip over a portion of that metrical pattern in establishing the congruence. We have just argued that reanalysis of the metrical pattern of the text sometimes does take place. However, many examples in which similar difficulties arise do not involve nonlexical words, and thus do not lead themselves to a reanalysis of the sort we outlined. Furthermore, there does exist a possible grid alignment for "oh, Alonzo Davis" which implies the original metrical constituent structure.

To cover those cases of tune-text incongruity in which reanalysis of the metrical pattern is not to be invoked, we need to be able to "skip over" certain incongruous sections of the tree. Thus in the metrical pattern of "oh, Alonzo Davis," we can achieve congruence with the tune if we disregard the portion of the tree in which the nodes are enclosed in parentheses:

2.4.4/16a

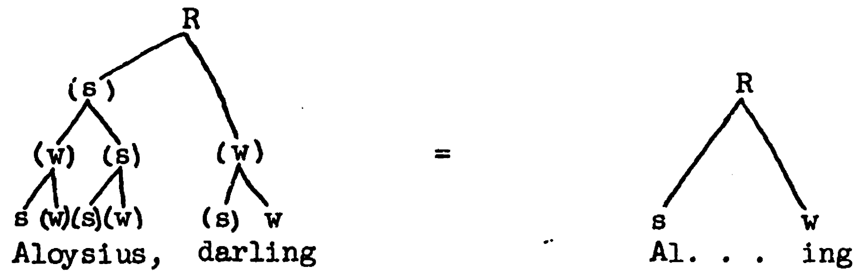


16b



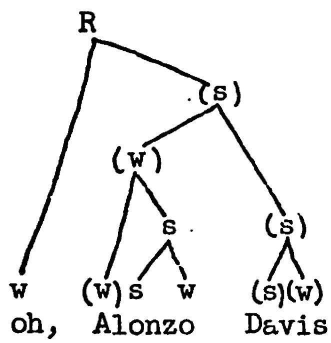
Now, obviously, we do not want to weaken our idea of congruence to the point that any subpart of any metrical tree can be ignored at our convenience. In particular: (1) we do not want to skip over material unless we are forced to, (2) we do not want to skip over discontinuous material, as in the case below,

2.4.4/17



and (3) we want to ensure that the designated terminals of the constituents in question line up with each other, so that we would not want to parenthesize 2.4.4/16a as follows:

2.4.4/18



since that would cause the main stress of the tune (the high tone) to be associated with a secondary stress of the text.

All three of these desiderata will follow if we adopt this rather simple reinterpretation of the notion of <u>congruence</u>:

2.4.4/19a If T_1 and T_2 are metrical trees, then their roots <u>correspond</u>.

19b If the nodes N_i and N_j <u>correspond</u>, and their daughters match in node labels, then the left daughter of N_i <u>corresponds</u> to the left daughter of N_j, and the right daughter of N_i <u>corresponds</u> to the right daughter of N_j.

19c If the nodes N_i and N_j <u>correspond</u>, and N_i is iambic (i.e. has the form $[w\ s]$) while N_j is trochaic (i.e.

2.4.4/19c has the form [sw]), then N_j corresponds to the right (i.e. strong) daughter of N_i.

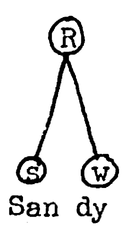
19d If every node of T_1 corresponds to some node of T_2, then T_1 and T_2 are congruent, and the set of correspondances between their nodes is a congruence.

The first, second and fourth clauses of the above definition are simply a formalization of the notion of congruence we have been assuming all along. The third clause (19c) is designed to handle the "skipping" or "ignoring" of certain constituents which we have been discussing. It should be clear that with the addition of 19c, a unique congruence is now defined between any two metrical trees whatsoever, independent of any possibilities of metrical reanalysis.

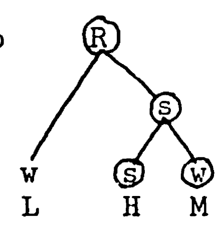
This definition of congruence will give us the right correspondances between the vocative tune and the original parsing of "oh, Alonzo Davis." As we will see in chapter 3, it (along with the Metrical Association Rule, or a theory of Grid alignment with the same effect) will give us the right text-tune correspondences in general.

It has another interesting application -- a reanalysis of the optionality of the initial low tone in the vocative tune. Remember the kind of case in which this optionality had to be invoked:

2.4.4/20a



20b



If we apply our new definition of congruence to 20a and 20b above, the initial low tone will be left out of consideration, and the congruence will be established among the circled nodes of the two trees. This will define the following text-tune association:

2.4.4/21

$$\begin{array}{ccc} & \text{San} & \text{dy} \\ L & H & M \end{array}$$

The L is now a "free tone", whose ultimate fate is in the hands of further rules. We have yet to discuss the operation of such rules, but it is easy to see that they might cause the "free" L to make a contribution to the phonetic rising tone which will occur on the main stress of such an example, and then disappear.

Thus, the definition of congruence in 2.4.4/19 offers a reason why it is the initial low tone of the vocative chant which is optional, and not some other of its elements. Since the pattern $\overset{w}{(T_1)}\ \overset{s}{T_2}\ \overset{w}{T_3}$ will be shown to be a common one for intonational melodies, this is not an insignificant result.

It is now time to move on to a consideration of the intonation of ordinary, unchanted speech. We have not by any means explained everything about the two chants we have discussed, but they have served their expositional purpose, as exemplification for our theory of tune-text association. We will return to them in later sections, using them to introduce some other ideas; some of the ends left loose in this chapter have been allowed to dangle with these discussions in mind.

Before beginning our consideration of the tunes of ordinary speech, we will attempt to tie at least the major threads of our discussion into a neat knot of summation.

2.5 Summary of the Principles of Tune-Text Association.

We began with the assumption that intonation, in one of its aspects, involves the association of a "tune" and a "text." This way of looking at things is fairly common and well-accepted in the field; thus Bailey (1970) writes that "intonational patterns will be understood here to be suprasegmental abstractions...which are distensible over grammatical spans of differing (segmental) lengths..." Some of our other background assumptions differ somewhat from the generally accepted common ground, primarily in that we propose to view tunes as well-ordered strings of (static) tonal segments.[5]

Starting from these assumptions, we have proposed a specific theory of the principles which govern the association of tune and text, exemplified and justified through an examination of certain "chants." This theory may be summarized as follows:

1) Texts and tunes individually are associated with metrical patterns, which are hierarchical organizations of strong and weak elements, in the form of trees. These metrical patterns are in some ways similar to stress patterns, as these are commonly understood, but obviously differ as well. The relationship of these two concepts will be explored in greater detail in a later section.

The basis of these metrical patterns is undoubtedly what James Burnet, exactly 200 years ago,[5] called "a natural propensity in the human mind to apply number and measure to every thing we hear; and indeed to every thing, as it is a necessary operation of intellect, being that by which intellect creates to itself its proper objects."

Burnet continues: "And as this propensity of the mind is previous to any opinion or determination of the will, I think, you properly call it instinctive. This is undoubtedly the foundation of all rhythm..."

2) The condition on the association of a text and a tune is that their metrical patterns be congruent, in a sense defined earlier. Roughly, one tree must be superimposed on the other without any conflict of constituent structure or (metrical) node labels. In conjunction with the Metrical Association Rule given in 2.4.3/29, the (congruent) superimposition of the metrical patterns of tune and text defines the underlying associations of their terminal elements.

3) The notion of a Metrical Grid was informally introduced, and argued to be necessary as the basis for an account of the temporal patterning of the chants. In addition, the concept of alignment with a metrical grid, we argued, explains the Metrical Association Rule, and indeed supercedes it.

4) The underlying tune-text association, established by the process sketched in 2) above, may have "free tones" and/or "free syllables." In some cases, additional association principles may be necessary to deal with such unassociated elements. The nature of these additional principles will be discussed in a later section.

3. The Tunes of Ordinary Speech.

Appearances perhaps to the contrary, the ideas developed in the preceding chapter were not a theory about the improvisation of songs using certain fixed melodies (although they might stand as the beginnings of such a theory). They were instead intended to serve as the basis of an investigation of English intonation. We will now redeem that intent by applying the metrical theory of tune-text association to examples drawn from unchanted speech.

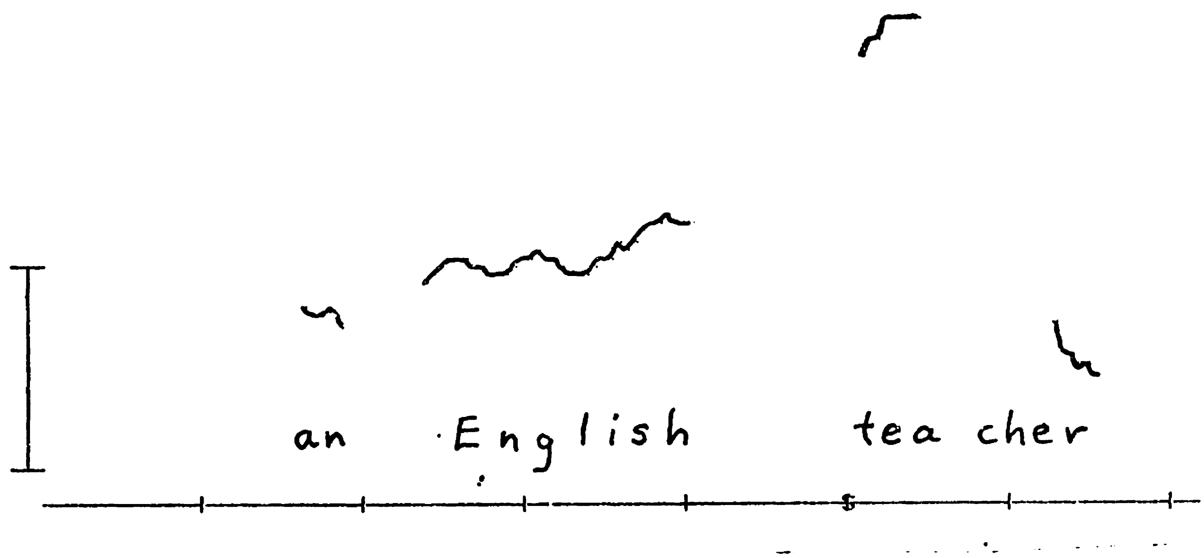
Throughout this section, our intent is primarily to develop answers to the questions raised in chapter 1, rather than to construct a listing of all the possible different "ways of saying" a given string of words. The development of an "intonational lexicon," especially one with an adequate theoretical basis, is an important task[1]; however, although we will discuss a wide range of different types of tunes, we will not attempt to write such a "dictionary." Since our goal is to construct a theory of English intonation, we will structure this section in a way which permits the orderly exposition of the key points in the theory, rather than in a way which would set out the various "tunes" of English in a systematic fashion.

We begin with an examination of certain simple tunes.

3.1 Tune-Text Association in Some Simple Cases.

Consider the tune of our example 1.2.2/4, which is reproduced below (in a slightly different version -- the difference will be discussed in section 3.2 , Boundary Tones.

3.1/1



We observed in chapter one that the meaning added by the intonation of this example might be something like "what I'm saying is perfectly obvious -- what else could the answer be?"

In our paper The Intonational Disambiguation of Indirect Speech Acts, Ivan Sag and I argued that the use of this tune is appropriate in two kinds of circumstances -- either "where the speaker is expressing surprise" or "where the speaker is suggesting that the utterance is redundant or unnecessary." These are some of the examples we gave:

3.1/2a Speaker walks into a room, is shocked by the paint job, , and says:

"My God -- the blackboard's painted orange!"

2b Speaker A: "What color is the blackboard?"

Speaker B: "I've told you a thousand times --

the blackboard's painted orange!

2c "Go open the door!" (understood: I shouldn't even have to tell you...)

2d "Well, butter my parsnips!" (understood: I never would've believed it...)

2e Speaker walks into friend's room, observes beautiful new oriental carpet, and says:

"Where'd you get the rug?"

2f Speaker and friend have been discussing friend's new rug. Speaker suggests the purchase of matching runners. Friend asks how they can find anything to match. Speaker says:

"Where'd you get the rug?" (understood: Obviously, they would have more of the same stuff...)

We also gave some examples in which the use of this tune would not be appropriate:

3.1/2 White House functionary mounts press-conference podium to announce:

"Ladies and gentlemen, the Marines have invaded Abyssinia."

This tune is included in the system of O'Connor and Arnold (whose book, *Intonation of Colloquial English*, is the nearest thing available to an adequate intonational lexicon) as "Tone Group 3." (N.B. Only those of their examples of this tone group in which the optional "Low Head" appears are relevant to this discussion). Their account of the meaning of this tune does not attempt to be as condensed as the surprise/redundancy hypothesis suggested by Sag and myself, but is quite compatible with it.

Commenting on various particular examples, they single out "senses" for this tune of "querulous or disgruntled protest," "reaction to something unexpected," "warmth-critical surprise that such an obvious course should not have occurred to the listener," "affronted surprise." The primary distinction is that the thread of negativity (querulous, disgruntled, critical, affronted...) that runs through their account is absent from ours -- we cited, in fact, examples like the one below in order to argue that the attitude expressed by this tune could be quite pleasant and positive:

3.1/3 "That's the most beautiful hat I've ever seen!" (as a warm and enthusiastic compliment)

Perhaps this is a difference between British and American usage, but some of the examples they give in the pattern drills that form the body of their text (which has a primarily pedagogical intent), are rather similar to the example we have just cited in 3.1/3. For example:

3.1/4 Verbal Context: What do you think of his latest painting?

Drill: I could hardly believe my eyes --

It's nothing less than a masterpiece.

N.B. This example is given with a sketched pitch contour, which is atheoretical but fairly easy to read, in place of O'Connor and Arnold's original notation, which is somewhat mystifying to the uninitiated. Along with the theory of intonational description which underlies it, this notation has a number of very interesting aspects, which we will discuss in a later section.

Thus it seems that the attitudinal negativity which often accompanies the use of this tune is not an essential part of its meaning, but simply a reflection of the nature of many of the circumstances in which a speaker would wish to project either surprise, or an implication that his utterance is unnecessary, or both.

The reason for giving so elaborate an examination of the meaning of this tune, is that it is necessary to demonstrate that there is some real linguistic entity here, whose properties are a fit object of study. We must show that among all the possible pitch contours which could have been given to tokens of the examples we have cited, the particular options represented, the particular versions we have chosen, form a natural class. If this is true, then the fact that our theory can predict the properties of the members of this class is an important result. If it is not true, then such a prediction is of no importance at all.

Thus the exemplification which we have given so far in this section should be seen as an argument that O'Connor and Arnold's Tone

Group 3, our surprise/redundancy contour, is a sort of intonational word, a unit of meaning.

Like any such argument, ours is essentially an appeal to intuition. It is well known that the meaning of a more conventional sort of word, e.g. "game," is difficult to state with theoretical precision, yet everyone will agree that there is a word "game," and that it does mean something. This agreement is based on our ability to recognize this word as an element of any utterance in which it may occur, as an abstract feature which is common to these otherwise quite different utterances, and which contributes something towards their final interpretation.

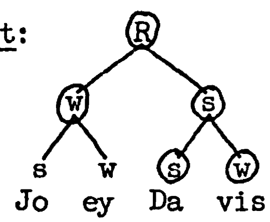
All we require from the discussion so far in this section is that the reader be convinced that there exists an intonational unit, a "tune," an abstract feature which is common to the otherwise rather different examples we have cited, and which contributes something to their communicative value.

If this point is granted, we can proceed to ask what the underlying form of this tune is, and how it is to be associated with various texts. On the surface, we see two crucial features: 1) a relatively low pitch at or near the beginning of the utterance, associated with a stressed syllable; and 2) a relatively high pitch associated with the main stress of the utterance, with a subsequent fall to low pitch again.

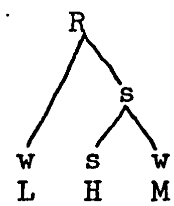
Given our hypothesis that tunes are underlyingly well-ordered sequences of static tonal elements, this surface description of the tune must be underlyingly represented as /L H L/. The fact that the

high pitch falls on the main stress of the text, tells us that the metrical pattern of this tune must be (some parsing of) $\overset{w}{/L}\ \overset{s}{H}\ \overset{w}{L/}$.
Thus our "surprise/redundancy" (henceforth S/R) tune has a structure which is quite similar to that of the vocative chant, which, the reader will recall, was given as $\overset{w}{/L}\ \overset{s}{H}\ \overset{w}{M/}$. Therefore, we would expect that their patterns of association with texts should match -- and they do:

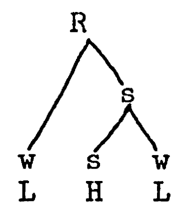
3.1/5a <u>text</u>:



5b <u>vocative tune</u>:



5c <u>S/R tune</u>:



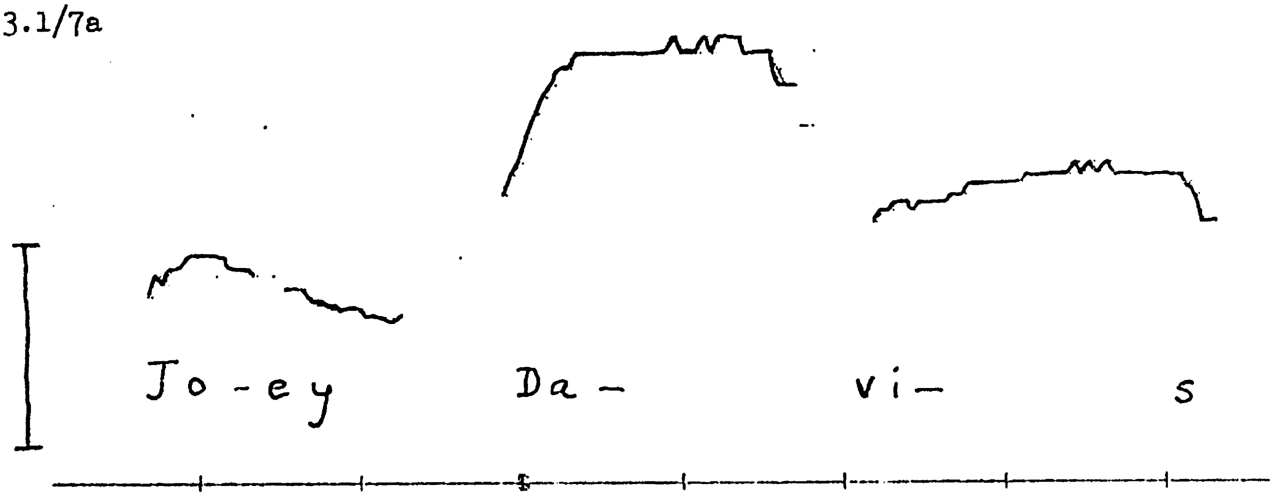
The metrical pattern of the text "Joey Davis" matches (in its circled-node portion) the metrical pattern common to the vocative tune and the surprise/redundancy tune. The resulting associations, by the Metrical Association Rule, are these:

3.1/6a <u>vocative tune</u>: Joey Davis
L H M

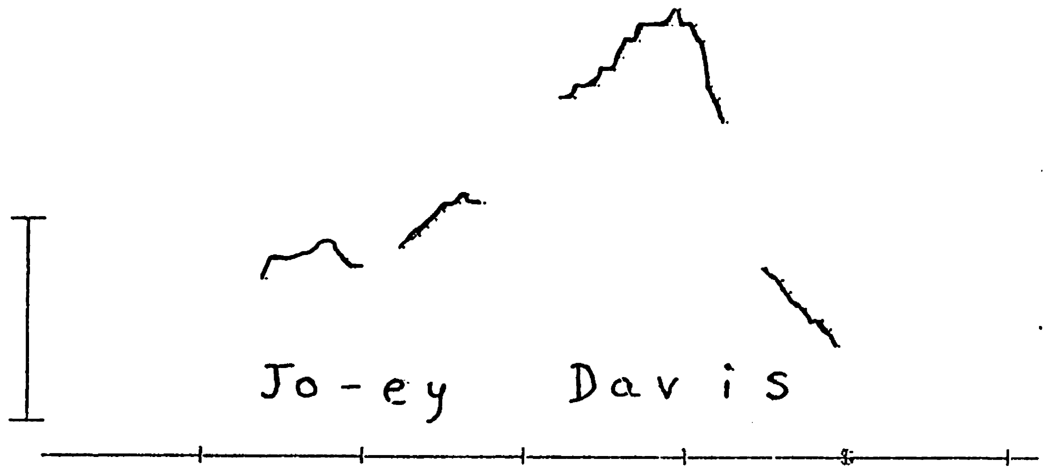
6b <u>S/R tune</u>: Joey Davis
L H L

F_0 contours for these two cases are given below:

3.1/7a



3.1/7b



Both of these results are consistent with the association-pattern predicted in 3.1/6a-6b, but the chanted and non-chanted examples differ in various ways that deserve comment:

1) In the chanted example, the tonal elements are realized as (approximately) level pitches, whereas this is not true in the spoken example. This was exactly the feature of the chants which initially attracted us, since the elements of their tunes, and the textual association of these elements, are thereby made clear.

2) In order for this use of phonetic level pitches in realizing the underlying tune to be possible, the syllables of the chanted example are considerably lengthened. Thus the chanted example (which happens to be about as short as a chant with this text is likely to be, in the natural course of events) is 1375 msec long, while the spoken example is only 810 msec long.

3) In the chanted example, we feel (at least) three "beats," at the positions represented by acute accents in "Jóey Dávís." It is also

possible to perform it in such a way that every syllable is felt to be a "beat," but three is the normal case. In the spoken example, on the other hand, we normally feel two beats, "Jóey Dávis." We will re-examine intuitions of this kind in the section on the Rhythms of Speech -- we mention them now simply because they distinguish chant from speech in such a noticeable way.

The above three observations are not intended to constitute a theory of the difference between chants and ordinary speech, but merely a set of facts which any such theory would have to account for. The next observation, however, will lead us to an important theoretical point.

4) In the case of the chanted example, the "free syllable" took on the pitch of the preceding syllable, a process which we represented earlier as the insertion of a dotted association line; Joey Davis. In the

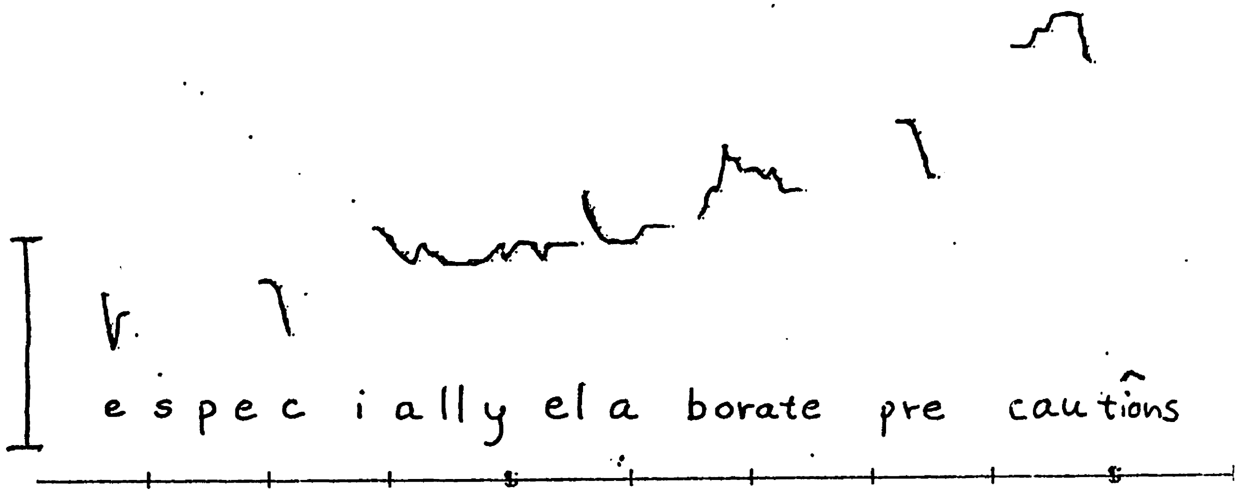
L H M

spoken example, on the other hand, the same "free syllable" was performed on a pitch midway between the preceding L and the following H. Actually, there was a reasonably smooth and unbroken rise from "Jo-" to "Da-". We might wish to represent this situation with some complex of dotted association lines, so: Joey Davis

L H L

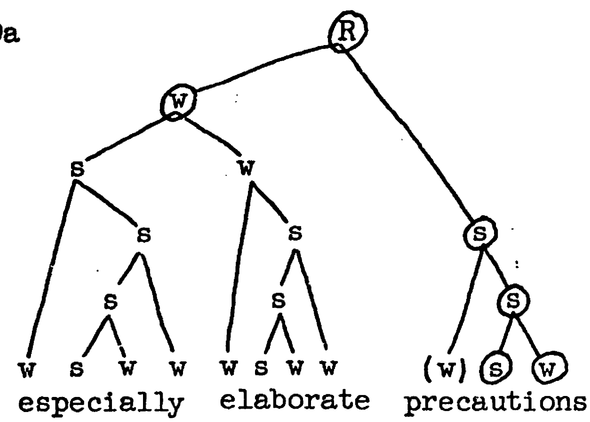
However, it is possible for a number of syllables to intervene between the syllable associated with the L and the syllable associated with the H, as in the S/R version of the NP "especially elaborate precautions," which might be performed as below:

3.1/8

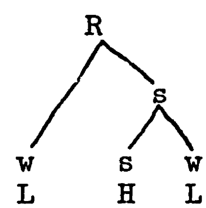


In this example, "especially" is assumed to have a higher stress than "elaborate;" with that proviso, its derivation is as follows:

3.1/9a



9b



Our definition of <u>congruence</u> will establish correspondances between the nodes of 9b and the portion of 9a in which the nodes are circled. The syllable <u>pre-</u> is skipped over by our principle for treating iamb/trochee mismatches. The effect of this congruence, plus the Metrical Association Rule, will be to define the pattern of tune-text Association given below:

3.1/10 especially elaborate precautions.

L H L

In this example, seven "free syllables" intervene between the site of association of the L and the site of association of the H. In actual performance of this example, these seven syllables simply fill in the interval between low and high with a rise. It seems neither feasible nor necessary to represent this fact as resulting from some complex process which adds additional association lines. It is reasonable to believe that if we specify point X as low, and point Y as high, then the observed treatment of the interval X - Y as rising may safely be left to physics, physiology and common sense.

If we were to perform this same text as a vocative chant, we would see the pattern of association below:

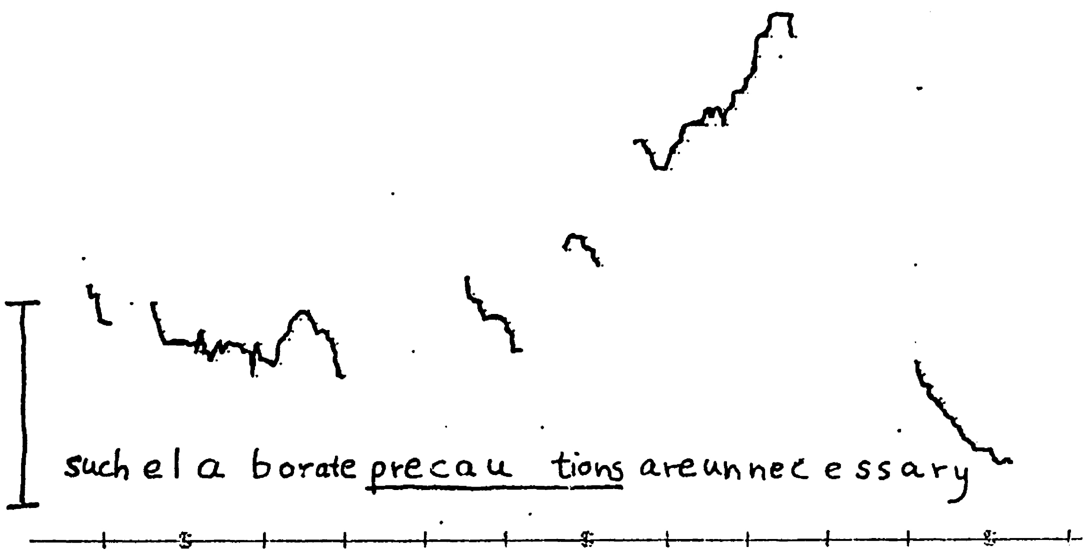
3.1/11 especially elaborate precautions

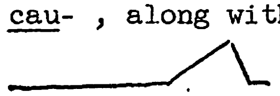
L H M

The fact that the L - H interval is not treated as uniformly rising may be attributed to an empirically observed property of chanting -- every syllable of a chant must be performed on one another of the set of pitches which make up the tune. The fact that the intervening free syllables take on the preceding low tone, rather than the following high tone, tells us that English prefers to spread tones left-to-right rather than right-to-left (really, forward in time rather than backward). Thus we may attribute the following property to chants, but not to unchanted intonation in English: *Every syllable must be associated with some tone*.

We have seen that free syllables intervening between two tonal association sites (at least in some cases) "fill in the interval" between the specified tones. How about free syllables which fall between a tonal association site and the boundary of the utterance? Let's consider some examples:

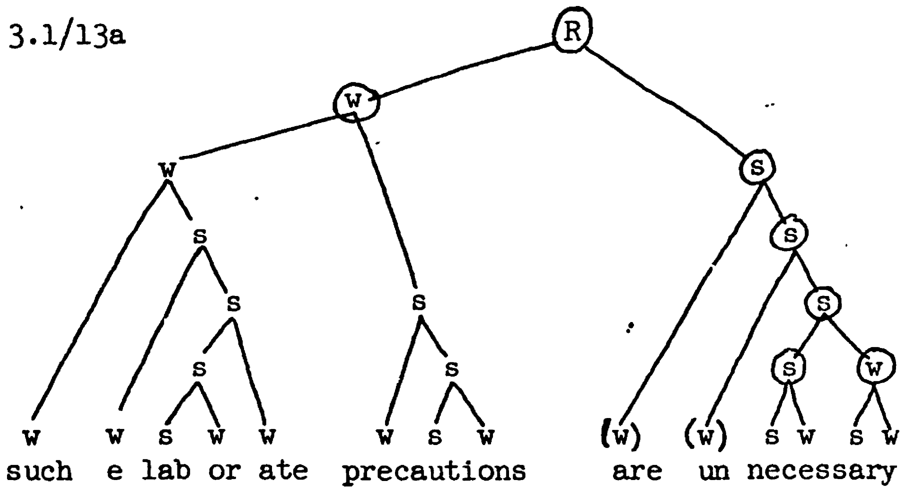
3.1/12



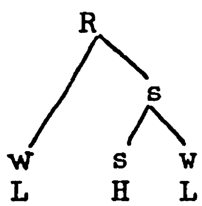
N.B. In this case the underlined syllables in "precautions" and "necessary" have been devoiced, or at least have had their voicing disrupted to the extent that the pitch extracter has not been able to extract a pitch. The falling pattern of the syllable cau- , along with other ups and downs within the overall  pattern, reflects the "micromelody" caused by F_0 effects of segmental phonology.

In the case represented by 12 above, the metrical congruence of text and tune suggested in 13a - 13b will yield the underlying association pattern given in 13c:

3.1/13a



13b



13c Such elaborate precautions are unnecessary.

L H L

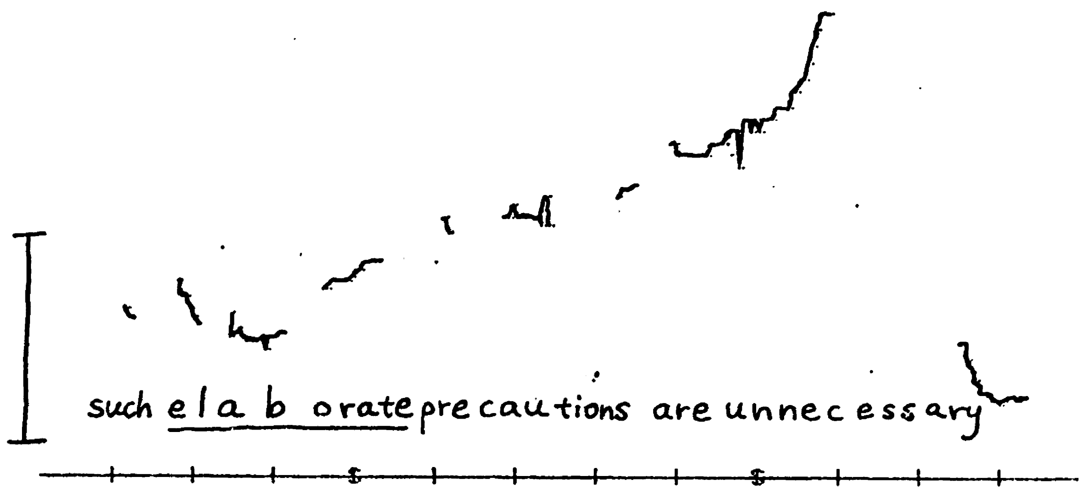
If we had chosen to give "elaborate" a higher stress than "precautions," as we are free to do, then the association pattern would have been:

3.1/14 Such elaborate precautions are unnecessary.

L H L

In this case, the F_0 would look like this:

3.1/15



Since our intuition (both in producing these sentences and in listening to them) is that the way of saying this sentence corresponding to the F_0 contour in 15 indeed reflects a higher stress on "elaborate," whereas the rendition in 12 reflects a higher stress on "precautions," these results confirm that our method of associating tunes and texts is working correctly.

In both example 12 and example 15, the free syllables between the locus (point of association) of the low tone and the locus of the high tone, simply fill in the interval between those fixed points. Thus in these cases, the point at which the contour "turns upward" corresponds to the highest stress in the portion of the utterance preceding the main stress.

We have explained this coincidence of "accent corner" and stress in terms of the association of the initial L tone of the intonational unit #L H L# with the designated terminal element of the textual constituent it corresponds to. (Actually, the notion of correspondance is defined on metrical nodes, not terminal elements like L, but when the node immediately dominating a tone corresponds to some textual constituent, we may loosely speak of a correspondance between the tone itself and that constituent).

The stretch of speech between the beginning of the utterance and the "accent corner," in both examples, is relatively level and low. This stretch comprises six free syllables in the version in 12, and two free syllables in the version represented by 15. In both cases, the treatment of these initial free syllables is consistent with the idea that their pitch is indeed "free," that is, not specified by linguistic rule. Thus we might regard the fact that they are all at a relatively level low pitch as another example of "just doing what comes naturally" in areas where the linguistic system provides no particular instructions at all.

However, the treatment of these initial free syllables is equally consistent with the hypothesis that they assimilate by rule to the tone to the right. So we are faced with a choice between a linguistic hypothesis and a nonlinguistic hypothesis in giving a theoretical account of these facts.

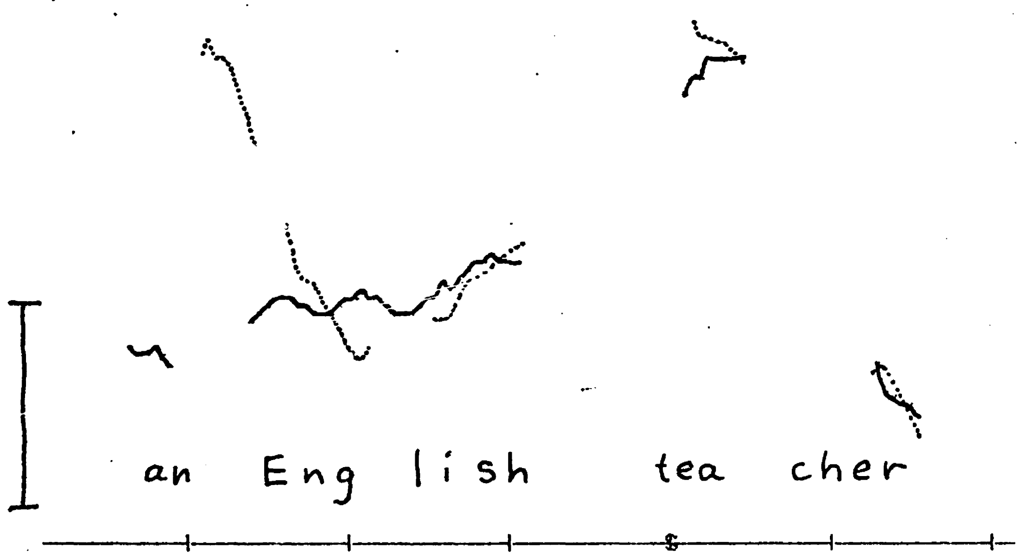
Certain additional data, which we will now examine, bears on this issue, and also will introduce us to some other aspects of the theory of intonation. The first set of facts that we will consider concerns an example of what we will call "boundary tones." We will then develop a more adequate inventory of underlying tonemic units, and proceed to examine a number of new tunes.

3.2 Some More Complex Cases.

3.2.1 The "High Prehead."

The version of "an English teacher" which we gave at the beginning of section 3.1 was different, as we noted at the time, from the version of the same text, set to the same intonational tune, which we gave in section 1.2.2. For ease of comparison, the F_0 contours of a similar pair of examples are superimposed below:

3.2.1/1



Both renditions of the text are examples of the surprise/redundancy contour we have been examining, and share its characteristic meaning. The dotted-line example, however, begins with a relatively high pitch on the initial syllable an, while the solid-line version has initial low pitch, like the other examples we have been considering.

This distinction is an instance of what is called, in the terminology of the British school, "high pre-head" vs. "low pre-head." In a group of tunes which includes tune 3, the contour we have analyzed as $\overset{w}{L}\ \overset{s}{H}\ \overset{w}{L}$, the choice between high and low pre-head is said to be free, although the version with low pre-head is generally given as the basic form of the tune. O'Connor and Arnold, in discussing the effect of this distinction on meaning, say that the effect of adding the high pre-head to a tune "is not to alter completely the impression made by the tune as a whole, but simply to add vivacity, liveliness, excitement and vehemence to whatever other attitudes the tune in question normally expresses."[1] This observation seems exactly right, and capable of being applied as it stands to the entry for this tonal entity in an intonational dictionary of English.

The tradition of intonational analysis to which O'Connor and Arnold belong analyzes the "tone group" into the sequence Pre-head + Head + Nucleus + Tail, of which all but the Nucleus are usually optional. The Pre-head is defined in slightly different ways by different writers -- O'Connor and Arnold call it "any syllables before the stressed syllable of the first prominent word,"[2] while Crystal calls it "any utterance which precedes the onset syllable within the same tone-unit," defining onset syllable as "the first stressed and usually pitch-prominent syllable."[3]

Crystal applies his definition in a way which leads to very small pre-heads -- he says that "in my data, the maximum number (of syllables) was five.[4]" O'Connor and Arnold, in their rather brief discussion, cite as an example of Pre-head a case in which it comprises four syllables. However, their discussion of the (for them, pragmatic) notion prominence,[5] which one would presumably apply in discovering the "first prominent word" in a given case, suggests that pre-heads with quite a large number of syllables ought to be possible.

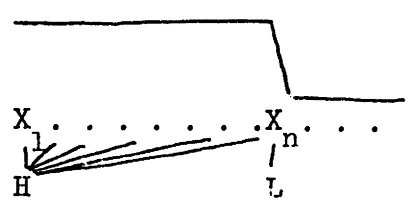
In terms of our theory, the "onset syllable" or "stressed syllable of the first prominent word" will in general (although not always) correspond to the locus of the first tone (excluding this "pre-head tone" we are discussing). We have seen that a fairly large number of syllables can precede this locus -- in example 3.1/12 there were six such initial free syllables.

However, the "high pre-head tone," whose nature we have been discussing, never seems to specify the pitch of more than one syllable, specifically the very first syllable of the intonational phrase. In other words, the state of affairs schematized in 3.2/2a below, in which a "high pre-head tone" is to be associated in some way with the string of syllables preceding the locus of the L tone, does not yield the result shown in 2b, but rather the one shown in 2c:

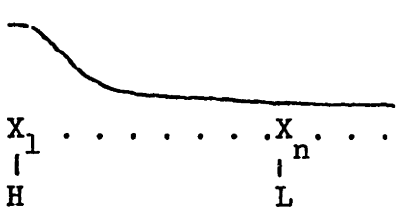
3.2.1/2a

$$\begin{array}{ccc} X_1 \;\ldots\ldots\ldots & X_n & \ldots \\ ? & | & \\ H & L & \end{array}$$

2b

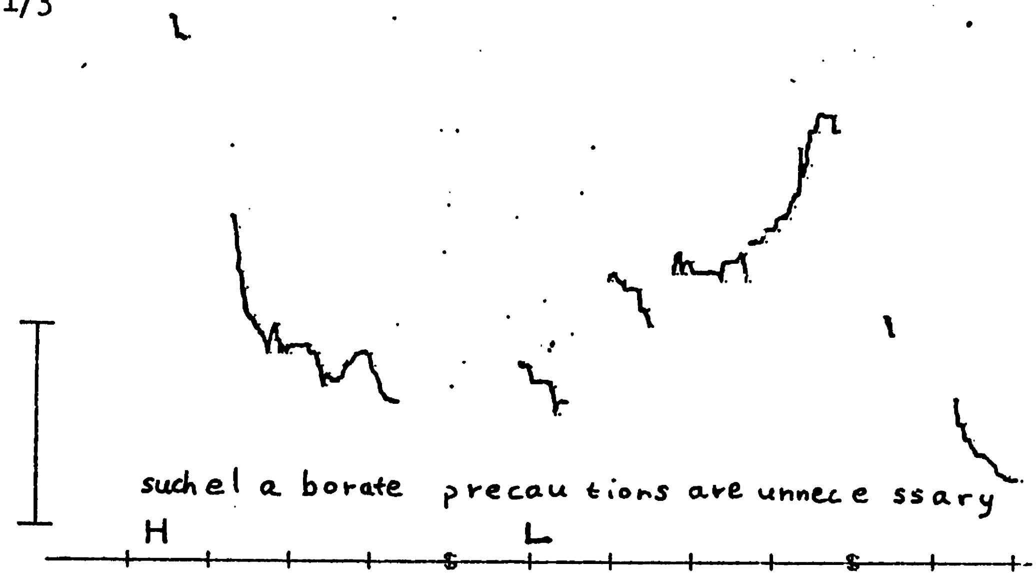


2c



An example is given below, "Such elaborate precautions are unnecessary" with "high pre-head:"

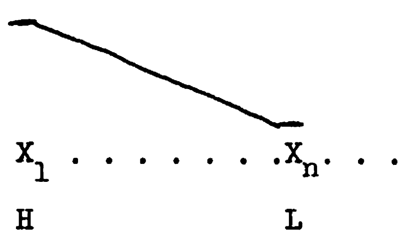
3.2.1/3



It should be emphasized that in this example the word such, although performed on a rather high pitch with a subsequent rapid drop, is not at all stressed, either in the speaker's intent or in the hearer's perception.

My intuition in performing such an example is of a contour like that schematized below:

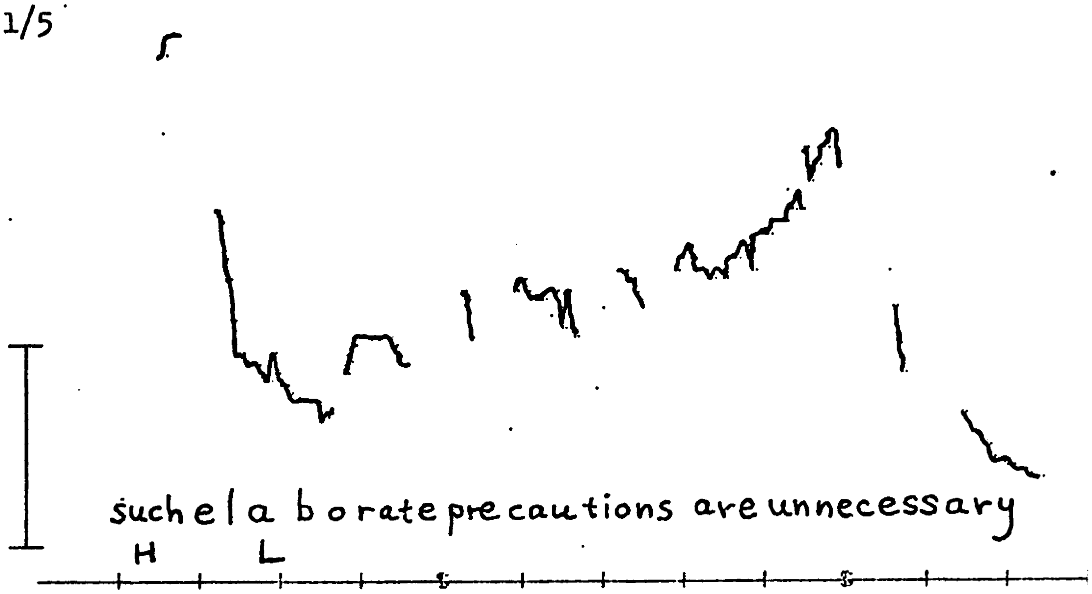
3.2.1/4



I have the feeling of descending smoothly from the initial syllable to the locus of the low tone, and it is always something of a surprise to see that instrumentally the greatest part of the descent occurs almost immediately after the beginning of phonation. I have attempted to monitor the laryngeal articulation of examples of this kind by the rough "field expedient" of placing fingertips on adam's apple, and find that I feel a prominent upward impulse on the initial syllable (<u>such</u>, in 3.2.1/3), and a prominent downward impulse on the locus of the low tone (the stressed syllable of <u>precautions</u>, in that same example). Whether I am performing naturally under such highly monitored conditions I cannot say.

The pitch-drop immediately following <u>such</u> would not look very much different, in this case, if <u>elaborate</u> had a higher stress than <u>precautions</u>, and therefore became the locus of the low pitch:

3.2.1/5



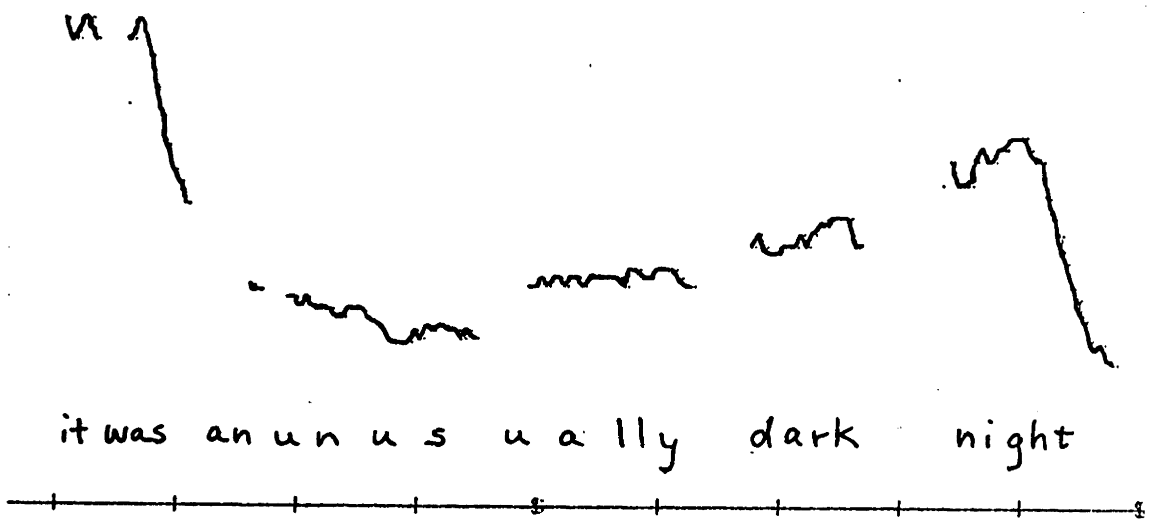
So we might want to argue that in the earlier case, in which the underlying locus of the L tone was in <u>precautions</u>, it has been associated by rule with the stressed syllable of <u>elaborate</u>:

3.2.1/6 such elaborate precautions

H L

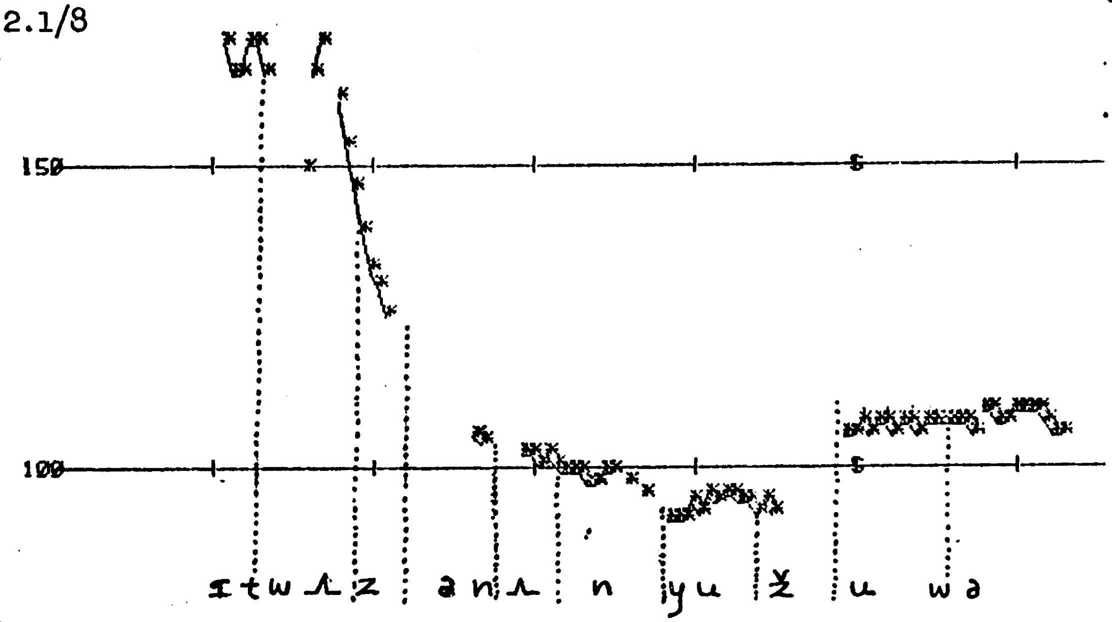
But this would be a mistake. In fact, it is pure luck that the leveling off of the initial abrupt fall, and the stressed syllable of <u>elaborate</u>, happen to coincide in this particular example. Suppose the same intonational pattern were used with the text "it was an unusually dark night:"

3.2.1/7



In this case, the point at which the abrupt fall levels off would coincide with the word <u>an</u>, which is one of the least stressed words in the whole sentence. This articulation of the pitch contour with the phonology of the text can be seen clearly in the expanded picture below:

3.2.1/8



N.B. The location of certain phonetic events (stop gaps, etc.) can be determined from the amplitude envelope which accompanied the original form of this figure. The locations of the phonetic segments in the figure, while of course approximate, were determined by this method.

What <u>an</u> (in 3.2.1/7) and -<u>lab</u>- (in 3.2.1/3) have in common is that both of them occur just about 200 msec. after the point at which the pitch begins to drop. I have not examined a large enough body of examples to say this with great confidence, but on the basis of the cases I have looked at, the segmental locus of the leveling-off of the F_0 contour (in cases of this nature) seems to be determined more by a blind temporal criterion than by any linguistically significant property of the text.

If this is so, then we would be justified in assuming that the "pre-head" H tone, and the L tone of the surprise/redundancy tune, have linguistically determined points of association, but the stretch in between them is governed by physics and physiology.

In general, it is well to be wary of facile physiological "explanations" in the area of intonation. A good example of an unsuccessful attribution of intonational processes to the account of physics and physiology is the idea that the pitch contours of utterances result from the characteristic subglottal pressure curve of the "breath group."

The danger of theories of this kind is that they remove whole areas of data from systematic linguistic investigation. Thus, if we decide that the tone of "free syllables" is not determined by linguistic rule, but represents a low-level interpolation between those syllables which do have tonal specification, then we will have no reason to pay much further attention to such syllables at the level of intonational theory, and we are unlikely to discover any linguistically significant regularities which may exist.

Nevertheless, I have not been able to discover any clear cases in which the systematic tonal specification of free syllables in English is necessary or even possible. Therefore, we will not attempt to devise a theory which would accomplish such specification. We will, however, continue to pay (purely descriptive) attention to the way in which such syllables behave.

3.2.2 The "High Pre-head" as a Boundary Tone.

The preceding discussion has glossed over a point which definitely is of systematic interest -- what is the underlying representation of the "pre-head high," and how is its textual association determined?

Two points are immediately clear: 1) the textual association of this tone is not, as the British treatment suggests, the whole of the "pre-head" as they define it; 2) neither is the textual association of this tone determined by congruence of metrical patterns, at least given that theory as we have developed it up to this point. In fact, the behavior of this tone is absurdly simple -- it associates with the first syllable of the text.

Metrically, this tone is clearly <u>weak</u> -- it is not perceived as being stressed in any way, and its "grid alignment" is intuitively always "off the beat." However, if we represent it as $\overset{w}{H}$, so that the surprise/redundancy tune with high "pre-head" becomes $\overset{w}{H}\ \overset{w}{L}\ \overset{s}{H}\ \overset{w}{L}$, we will get incorrect results -- given the text "such elaborate precautions are unnecessary," with "precautions" higher stressed than "elaborate," we would get the association pattern in 1a below, instead of the observed pattern in 1b:

```
3.2.2/1a      *Such elaborate precautions are unnecessary.
                    |           |               |   |
                    H           L               H   L

      1b       Such elaborate precautions are unnecessary.
               |                  |             |   |
               H                  L             H   L
```

N.B. The association in 1a is a perfectly well-formed intonation, whose generation will be discussed shortly. However, it is not a possible output for the case we are discussing -- in particular, it requires "elaborate" to be higher stressed than "precautions," contrary to assumption.

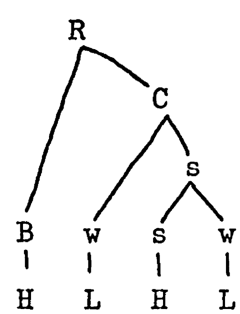
The view which is best in accord with the facts of the matter seems to be that the underlying association of this tone (and others like it) is in fact with a <u>boundary</u>. In all the examples considered so far, the boundary in question has been the beginning of the utterance; it is easy enough to find terminal boundary tones as well, and in more complex cases, these initial and terminal boundaries may be embedded in larger intonational structures.

Notationally, such boundary tones will be represented as $\begin{smallmatrix}B\\|\\T\end{smallmatrix}$. For the purposes of tune-text association, it will be assumed that corresponding boundaries exist in the text, and that <u>congruence</u> requires a matching of boundary to boundary. In the configuration $\begin{smallmatrix}B\\|\\T\end{smallmatrix}$, the tone T is a terminal element, while the symbol B is a non-terminal. For purposes of grid alignment, the B positions are (redundantly) weak. In order to include these boundary elements in our metrical trees, non-boundary material will be called "content," symbolized by the node label C.

Thus the surprise/redundancy tune with "high pre-head" would be represented as in 2a, with its most likely parsing being that in 2b:

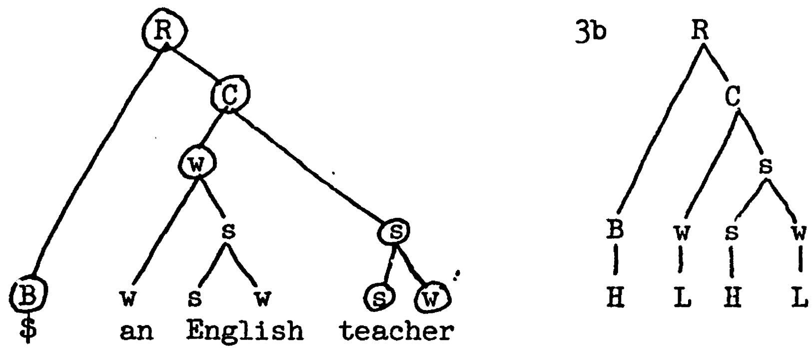
3.2.2/2a $\begin{matrix}B & w & s & w\\ | & | & | & |\\ H & L & H & L\end{matrix}$

2b



We will represent the corresponding boundary in texts with the terminal symbol $, which must always be immediately dominated by the nonterminal node B. Thus the derivation of "an English teacher," with the L H L tune and "high pre-head," would be as schematized below:

3.2.2/3a



The metrical tree of the tune will be made congruent to that of the text, in the familiar way, resulting in this underlying association pattern:

3.2.2/4

```
$   an  English  teacher
|       |          |  |
H       L          H  L
```

In order to get the observed output, we need to add the addition association symbolized by the dotted line:

3.2.2/5

```
$   an  English  teacher
|_ ·'   |          |  |
H       L          H  L
```

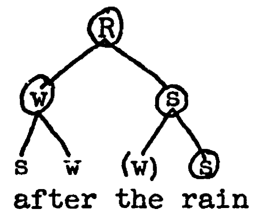
The reason for this additional association is fairly clear. There is a disproportion between the terminal element of B in the tune, and the terminal element of the corresponding node in the text. The tonal B is an entity with phonetic content (in this case, a high tone), while the textual B is an abstract boundary-symbol, with no phonetic content at all. It is a fact about intonation that tones never are produced by themselves, but only in association with textual material. Therefore, if a boundary tone is ever to be realized at all, it must find some toe-hold in the text. That this toe-hold should be the syllable immediately adjacent to the textual boundary, is rational enough.

We have recently rejected the hypothesis that all syllables must be associated with some tone, preferring to allow underlyingly free syllables to remain free. The treatment of boundary tones suggests that the converse hypothesis might well be true -- that *all tones must be associated with some syllable*. This would be a theoretical statement of the empirically valid observation made above, that tones are produced only in association with textual material. In fact, we are already in possession of evidence which argues that this proposition is true. In chapter 2, we tacitly invoked such a premise in deriving the vocative chant version of a text like "McDuff." We argued that the underlying association pattern in 6a would be modified by the addition of the dotted association line in 6b:

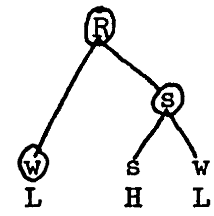
```
3.2.2/6a     McDuff                    6b     McDuff
              | |                               | |\
              L H    M                          L H M
```

Similar examples can easily be constructed for the surprise/ redundancy tune:

3.2.2/7a



7b



7c after the rain

L H L

7d after the rain

L H L

The metrical patterns of text and tune in 7a and 7b will give the underlying association pattern in 7c -- to get the correct output, in which "rain" falls from high to low, we need to add the dotted association line in 7d.

3.2.3 A Theory of Derived Associations.

John Goldsmith, in his theory of Autosegmental Phonology, has suggested that results somewhat similar to those that we have been representing by the addition of dotted association lines, should be derived not by a specific linguistic rule, but by reference to very gneeral well-formedness conditions like "all tones must be associated with some syllable." This notion of derivation by well-formedness condition is an extremely promising one a priori, since it severely restricts the class of possible "rules" that such a system could have. The cases in which we have found it necessary to invoke the addition of dotted (=derived) association lines to our underlying, metrically determined patterns of association, are exactly the sort of cases in which a Goldsmithian approach would work splendidly. We will therefore adopt such an approach, based on the following two premises:

3.2.3/1a All tones must be associated with some syllable.

1b Association lines may not cross.

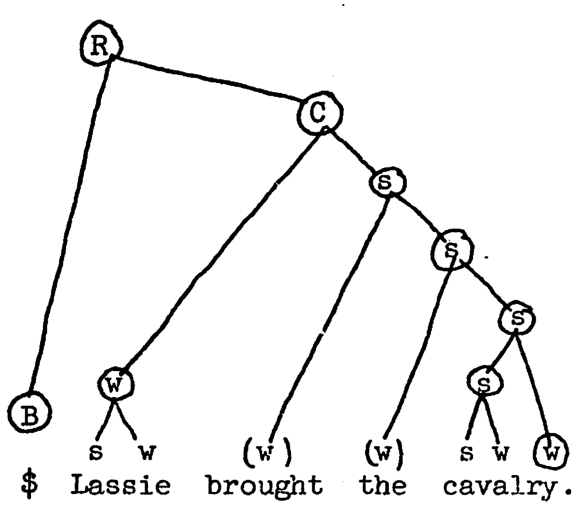
Both of these premises are borrowed from Goldsmith's work[6]; he utilizes some others (e.g. that all syllables must be associated with some tone) that we will not adopt. The idea of this theory is that if condition 1a is not met by some text-tune association pattern, then additional association lines must be drawn, subject to the condition in 1b.

In all of the cases we have encountered so far, principles 1a and 1b suffice to unambiguously determine the correct derived associations. Cases will arise later on in which 1a and 1b will have to be supplemented by certain additional principles, which we will add as they become necessary.

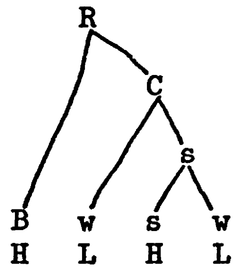
One such case can be adduced immediately. Consider the text "Lassie brought the cavalry," set to the tune

B w s w
H L H L :

3.2.3/2a



3.2.3/2b



2c $ Lassie brought the cavalry.

H L H L

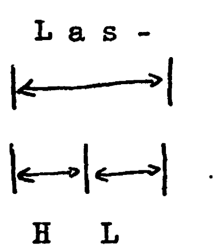
The trees in 2a and 2b will yield the underlying association pattern given in 2c. By hypothesis, the H tone underlyingly associated with the $ boundary needs to become associated with some syllable; the only candidate available is the one immediately to the right:

3.2.3/3 $ Lassie brought the cavalry.

H L H L

How would this be realized phonetically? By the principle of sequentiality, the ordering H L must be preserved. But by the principles of our metrical association theory, the L tone must be aligned with the same "beat" as the designated terminal element of $\begin{bmatrix} s \quad w \\ \text{Lassie} \end{bmatrix}$, namely its initial syllable. Since the underlying metrical correspondance is H to $\begin{bmatrix} s \quad w \\ \text{Lassie} \end{bmatrix}$, this process of grid alignment is the only means we have to get the tone on the proper syllable. If the phonetic realization of this situation were

3.2.3/4

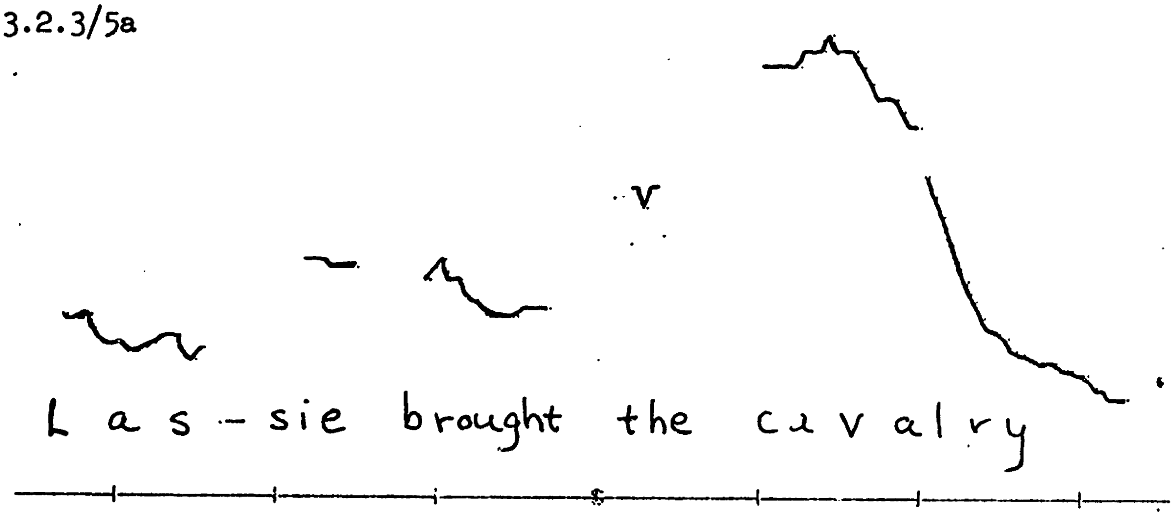


then the meter would be wrong, since it is normally the beginning of a note that is felt as the "beat." By this argument, the L tone must begin the syllable, and this leaves no room for the preceding boundary-linked H tone.

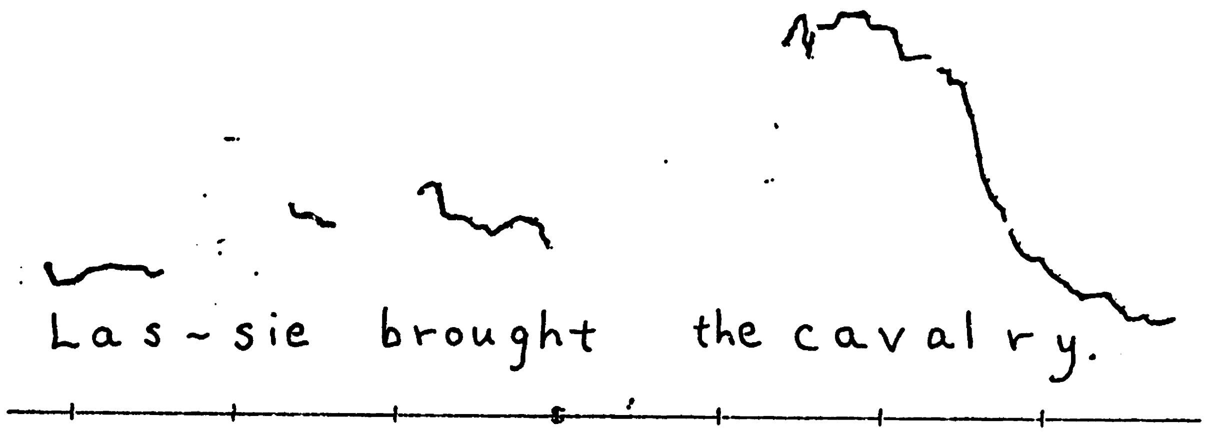
The force of this argument is constructively to delete the H tone -- we can draw in the dotted association line, but there's still no room for the tone to be performed. It is thus fortunate for our metrical association theory that the tone in fact cannot be performed.

Given a situation in which a nonboundary tone is associated with the initial syllable of the intonational phrase, the distinction between H and L initial boundary tones is neutralized. It is possible to "think" the difference, but there is little or no phonetic effect. The best I was able to do at performing such a distinction, on the example at hand, is given below:

3.2.3/5a



3.2.3/5b

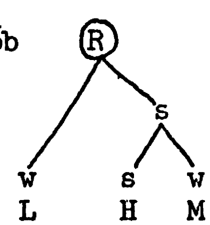


5a is the case in which the boundary tone was supposed to be high -- 5b is the case in which it was supposed to be low. It is possible to feel quite a striking difference between such examples while producing them, but the phonetic effect is minimal, and the resulting tokens, on playback, do not seem distinguishable perceptually.

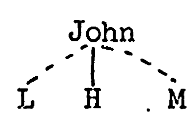
It does not seem to be necessary to invoke a special "tone deletion rule" to handle this case -- the principles of the theory of metrical association will guarantee the observed result. These same principles give the correct result in a number of analagous cases as well. Suppose we attempt to set the monosyllabic text "John" to the tune of the vocative chant:

3.2.3/6a

6b



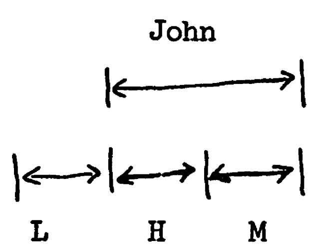
6c



Given the metrical patterns for text and tune in 6a and 6b, the definition of congruence given in chapter 2 will simply make root correspond to root, and then stop, having run out of text. The principles of grid alignment, discussed in reference to our Metrical Association Rule, will ensure that the designated terminal elements of the corresponding nodes will be aligned -- this may be seen as giving the solid association line in 21c. The well-formedness condition in 3.2.3/1 will add the dotted association lines.

Since, as we argued earlier, it is the initial portion of the realization of the syllable which is metrically strong, the metrically strong high tone will be aligned with the beginning of the syllable. This again leaves little or no room for the preceding low tone. There is nothing to prevent the alignment of the following mid tone with the second portion of the syllable. Thus the predicted alignment is schematically:

3.2.3/7

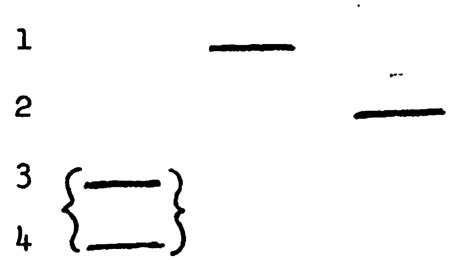


In this case the low tone is constructively deleted, although no special rule need be invoked. If it contrasted with another tone in the same position in the tune, we would expect that contrast to be neutralized. Indeed, there is a tonal alternative in this position --

there are actually three versions of the vocative chant, depending on the relation of the first tone to the second. In one version, the first tone is a minor third below the second tone, and thus has the same pitch as the third tone -- for some reason, I persist in feeling it to be in between the second and third tone when I perform this version, but on replay, as well as instrumentally, it is clearly the same as the third pitch. This version is fairly highly marked, and doubtless rarer than the other two. Its nature is rather unclear to me; it seems to occur only when the overall range of the chant is set fairly high.

In both of the other two versions, the first tone is lower than either the first or the second tone, and in both cases, the second and third tones form a descending minor third. The two versions differ, acoustically, in that in one of them the first tone forms a markedly larger interval with the second tone. They thus require four tonal levels for their description.

3.2.3/8



The version with the smaller initial interval seems to be appropriate in circumstances in which the intended hearer is closer to the speaker, either physically or metaphorically. For example, if the speaker and hearer have just parted company, and the speaker wishes to regain the hearer's attention for some reason, then the version with the smaller initial interval would be preferred. If the speaker is calling a great distance, or is not sure that the hearer is really there, then the version with the larger initial interval is more appropriate. To the extent that I have intuitions about the appropriateness of the version in which the first tone is on the same pitch as the third, it seems to pattern with the "distant" version rather than the "proximate" version.

The fact that these three versions of the vocative chant require four tones for their description raises an interesting point. The number _four_, as we will point out in the next section, recurs in descriptions of the systematically distinct pitch levels of English intonation. Let us suppose that there is a good reason for this, and that the underlying system really is tetratonal. What will be the properties of a musical version of such a system, a tetratonal _scale_?

In all the human scale systems that I know of, octaves are considered "the same note." In other words, any given pitch and the set of other pitches related to it by octave intervals are considered to be what we might call a _tonal equivalency class_. Thus a scale of _n_ steps must be fitted into the distance between the pitch taken as basic, and two times that pitch, which is one octave up.

For example, the familiar 12-tone chromatic scale of Western music divides the interval 2x into twelve equal steps. "Equal" in perception of intervals is defined by multiplication, not addition, so the chromatic "half-step" which is the characteristic interval of the twelve-tone scale is a ratio _R_ such that each scale-step is R times the pitch of its predecessor. Thus the base pitch is x, the first scale step is Rx, the second scale step is $R \cdot (Rx) = R^2x$, the third scale step is $R \cdot (R^2x) = R^3x$, and so forth. Since the whole gamut must be fitted into an octave, we know that $R^{12}x = 2x$, or, by simple algebra, $R = \sqrt[12]{2}$. This number is approximately 1.0595, and constitutes the characteristic ratio of half-steps in a tempered 12-tone scale.

By a similar argument, a tetratonic scale composed of equal intervals has $R = \sqrt[4]{2}$. This number is approximately 1.189, and corresponds to the musical interval of a minor third (the minor third defined by the equivalence <u>three perfect fourths</u> = <u>octave plus minor third</u> is slightly flat of it, but the correspondence is pretty close). Of course, it is possible to see by inspection of the piano keyboard that a minor third is the characteristic interval of a tetratonic scale proceeding by equal intervals, e.g., C-E♭-G♭-A-C.

Now, this choice of a scale is not a good one musically, primarily because the interval of a fifth (defined by the second harmonic of any musical tone, and therefore second only in importance to the octave, the first harmonic) plays no role in it. But it is the only way to define a tetratonic scale by equal intervals, independent of any considerations imposed by the overtone series. So the overtone series strongly militates against the choice of a tetratonic scale, and any reasonable aesthetic concern for sonority dictates that in choosing a scale for extensive musical use, one pay attention to the demands of the overtone series.

However, if one were not starting from any aesthetic considerations, but primarily from the desire to superficially transform into music (="chant") a linguistically given system with four tonal elements, it is reasonable that the intervals would be set so as to give a tetratonic scale of the kind we have been discussing. Most probably, therefore, the reason that the mathematically unnatural interval of a minor third is humanly so natural, is that the linguistic tonal system specifies four pitch-classes.

This (I think rather striking) argument to the side, we have observed that there are various possibilities for the initial tone of the vocative chant, in particular a distinction between a wide interval with the second tone (of about a major sixth) whose pragmatic effect we characterized as "distant," and a narrow interval (of about a tritone) whose pragmatic value we characterized as "proximate."

Whatever its exact meaning, this distinction is clearly neutralized in case the main stress of the text is on its initial syllable -- no difference of this sort can be effectively produced in the vocative-chant version of a name like "John" or "Sandy." This is as predicted by our metrical theory of tune-text association.

3.2.4 The Tonal Phonemes of English.

The combination of the minor-third interval of the second and third tones of the vocative tune, with the two variant forms of the initial tone, defines four tonal levels:

3.2.4/1

We will give these levels the following names:

3.2.4/2

1	high	(H)
2	high-mid	(H-M)
3	low-mid	(L-M)
4	low	(L)

Given a set of two binary distinctive features [± High], [± Low], the four levels will be specified as follows:

3.2.4/3

1	high	[+High, -Low]
2	high-mid	[+High, +Low]
3	low-mid	[-High, -Low]
4	low	[-High, +Low]

This assignment of features to the two mid tones is simply arbitrary, at the moment, but I believe it is correct, for reasons that will emerge as we go along. The four levels are defined abstractly, not phonetically, and their phonetic realization will be influenced by the expansion and contraction of the speaker's <u>tessitura</u> (the range over which the speaker is prepared to modulate pitch, at a given point in time). The nature of these levels, or <u>registers</u>, will be discussed somewhat further in a later section -- we will argue that the unavailability of hard-and-fast phonetic criteria for the identification of a particular <u>register</u> (e.g. Low = 100-110 Hz) should not be particularly disturbing, since such context-free criteria are rarely (if ever) available for any phonological features.

The idea that English has (basically) four phonemically distinct levels of pitch is hallowed by tradition -- American Structuralists have proclaimed it openly. In arguing for the validity of a four-way distinction, Pike[7] says: "a description in terms of three levels could not distinguish many of the contours. . .A description in terms of five or six levels would leave many theoretically possible contrastive combinations of pitches unused. The four levels are enough to provide for the writing and distinguishing of all of the contours which have differences of meaning so far discovered. . ." Agreeing entirely with this view of the matter, we stand squarely in the Wells-Pike-Smith-Trager tradition in terms of our view of the underlying intonational phonemes of English; we differ somewhat in our terminology for the four-fold distinction (theirs is "extra high, high, middle, low"), and in the distinctive-feature analysis we give it.

In terms of this analysis, the "distant" form of the vocative-tune is $\begin{bmatrix}\text{-High}\\ \text{+Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{-Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{+Low}\end{bmatrix}$ (L H H-M); the "proximate" form is $\begin{bmatrix}\text{-High}\\ \text{-Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{-Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{+Low}\end{bmatrix}$ (L-M H H-M). Thus we can abstract out the sequence $\begin{bmatrix}\text{-High}\\ \pm\text{Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{-Low}\end{bmatrix} \begin{bmatrix}\text{+High}\\ \text{+Low}\end{bmatrix}$ as common to both versions, with the +Low/-Low distinction in the initial toneme corresponding to the distant-proximate distinction in the meaning of the resulting tune.

The preceding discussion raises the issues of the internal structure of intonational tunes, and the relationship of this internal structure to their meaning. In the recent history of the literature on this question, we may single out two general positions, the structural position and the phonesthic position. According to the structural view, tunes are made up of sequences of distinctive elements (either static tonemes, or complex elements like "high falling nuclear tone") which are combined into units of meaning of greater or lesser complexity; these in turn combine to make up the tunes observed in ordinary speech. According to the phonesthetic view (primarily championed by Dwight Bolinger), there is not really any fixed inventory of structural elements, nor any possible division of a given pitch contour into well-defined and separate units of meaning; rather, there is some inventory of gestalt characteristics of contours, which may be local properties (e.g., the existence of a point of inflexion of the contour) or global ones (being generally falling, involving no sharp changes in slope, etc.). These properties affect the interpretation of the utterances bearing the contour, although the interpretation of a given property may vary somewhat from case to case, depending on the nature of the sentence, the surrounding circumstances, etc. A large part (perhaps all?) of the interpretation of a given property of a given contour arises from its participation in some sort of system of sound symbolism.

The point of view that we will adopt incorporates features from both of these positions. As we have made clear, we accept the view that tunes are to be analyzed as strings of static, phonemically distinct elements; however, this idea is not at all incompatible with the notion that intonation represents a (conventionalized) system of sound symbolism. We will try to demonstrate the reasonableness of such a synthesis in the next section.

3.2.5 Some Remarks on the Nature of the Intonational Lexicon.

If we are to make sense of the data which will come pouring in when we begin to consider a wider range of tunes, we need to know not only the nature of the phonemic units involved, but also a little about the kind of structurally significant sequences (and sequences of sequences) of these units that we can expect to find. The nature of the initial level of this further structure has been the source of enormous disagreement and confusion in the intonational literature. There are, I believe, two basic reasons for these difficulties -- first, the failure to make a distinction between morphemic analysis and idiophonic analysis of words, and second, the failure to recognize that words (of whatever type) are usually greater than the sum of their parts.

We presupposed some kind of decomposition of tunes into more basic elements when we analyzed the two vocative tunes into a portion which they held in common, and a portion (the feature $\pm$ Low in the initial tone) which distinguished them. The idea behind this move was that elements of the intonational lexicon might consist of incomplete tonal specifications of some sequence of segments, with a number of compatible entries "intersecting" to produce a completely specified string.

If we set aside for the moment the possible incompleteness of their feature matrices, these basic intonational elements, which may combine to form "tunes," are somewhat analagous to *morphemes*, which may combine to form *words*. *Morphemes* are properly structural units, units of form, but have some practical (if not theoretical) claim to stand as units of meaning as well. Even though the meanings of polymorphemic words are not (in general) compositionally derived from their component parts, we usually have some sense that these parts have meaning as well as form. Thus *pre-* in *precede* and *prepare*, or *-cede* in *precede* and *recede*, are felt to make a contribution of some kind to the meaning of the individual members of their family, even though a special lexical entry for each member is undoubtedly still necessary.

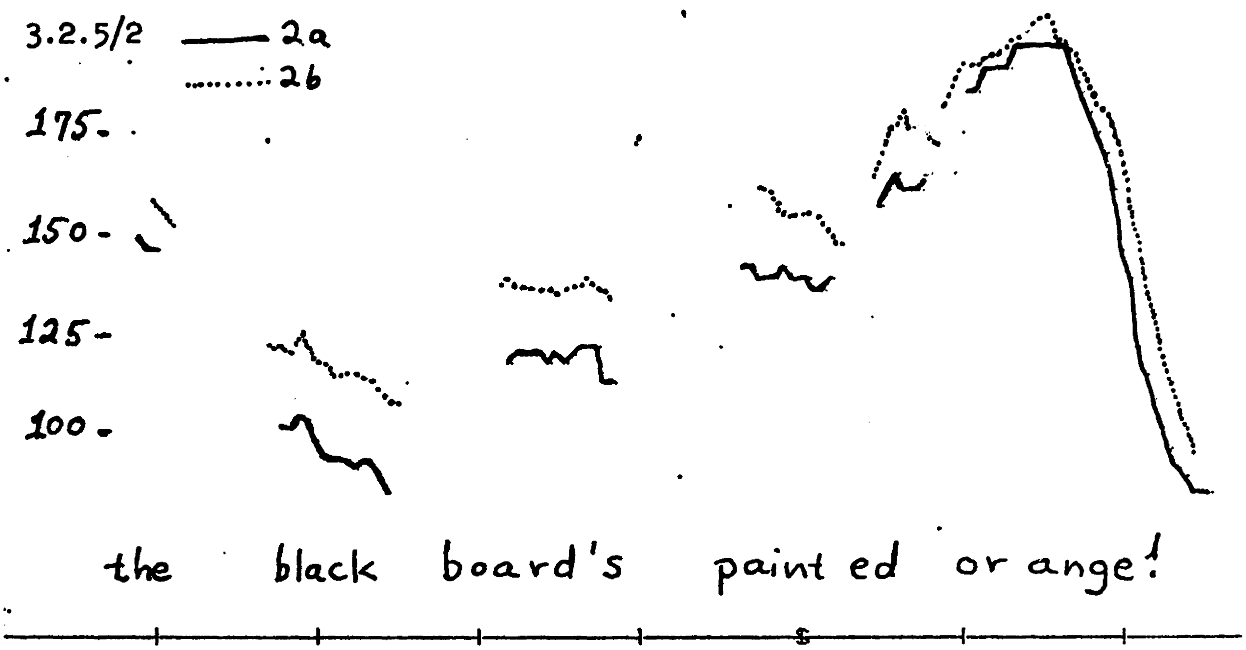
A similar situation undoubtedly obtains in the intonational lexicon. Our analysis of the two vocative tunes can be considered to divide them into a "stem" — H H-M with two contrasting prefixes L— and L-M—. In this case the "stem" appears to maintain a core of meaning between the two cases, comparable perhaps to *-pose* in *interpose* and *prepose*. We hypothesized that the two "prefixes" represented the meanings of *distant* vs. *proximate*, or something of the sort. Suppose we now compare our surprise/redundancy tune (L H L) with a tune in which the initial low is replaced with a low-mid:

```
3.2.5/1a     $ The blackboard's painted orange.
             H       L                    H L
```

3.2.5/1b $ The blackboard's painted orange.
H L-M H L

We maintain a high IBT in both cases; the derivations involve nothing we haven't seen before. F_0 contours for these cases would look like this

3.2.5/2



The intonation which results from substituting L-M for L in the "surprise-redundancy contour" probably falls (along with other cases which we would analyze differently) under the aegis of Tone Group 4, in the O'Connor-Arnold system -- they say that this Tone Group "avoids the disgruntled effect of Tone Group 3 (*our S/R tune*) whilst still retaining the lightness, the airiness and the effect of personal participation in the situation, characteristic of the High Falling nuclear tone." (Because of the nature of their treatment of "heads," argued by Crystal to be inadequate, there is not a very clear place in their system for our L-M H L tune).

We have already expressed disagreement with the notion that any meaning of "disgruntledness" is associated with the $\overset{w}{L}\ \overset{s}{H}\ \overset{w}{L}$ tune; our idea of the distinction between $\overset{w}{L}\ \overset{s}{H}\ \overset{w}{L}$ and $\overset{w}{\text{L-M}}\ \overset{s}{H}\ \overset{w}{L}$ will be discussed later on, although this sort of lexicography is not a primary goal of the present work. The point that we wish to make here is simply that there is no reason to expect the meaning of such tonal sequences to be a direct function of the meanings of their perceived parts.

The juxtaposition of the two intonations in 3.2.5/2 with the two versions of the vocative tune gives us two "prefixes" (L and L-M) and two "stems" (H H-M and H L). The resulting situation is somewhat like the one that arises if we take the two prefixes *pre*- and *inter*- , and the two stems -*pose* and -*dict*. The four words that result (*prepose*, *interpose*, *predict*, *interdict*) are perceived to have a certain internal structure, and the elements of this structure are not unrelated to its ultimate meaning, but the degree of semantic compositionality ranges from fairly great (interpose) to fairly small (interdict).

Although this situation is familiar enough to have become a matter of linguistic common sense in non-intonational morphological analysis, it seems to have caused a great deal of confusion in the analysis of intonation. Almost without exception, it has been assumed that an intonational element which is perceived to be a unit with some intuitively graspable "meaning," necessarily has a fixed semantic or pragmatic value which enters compositionally into the interpretation of every larger intonational unit of which it is a part.

An example of this "over-analysis" is commented on in Liberman and Sag 1974. We began from the observation that in certain cases, a particular choice of intonation affects the availability of otherwise optional differences in the scope of a negative. We explained this fact as a pragmatic result of the meaning of the contour in question, which we proposed to be that "the speaker is using the utterance which bears it to contradict. . .some assumption or implication." Pike, in analyzing an example of this same contour, suggested that it was composed of two units, one meaning "center of attention, pointing, contrasting," and the other meaning "incomplete deliberation, incomplete sequence." Commenting on this analysis, we observed that "the meanings he assigns to its (allegedly) constituent pitch morphemes are, like good astrological readings, not demonstrably inconsistent with the facts, but far too vague to be of much predictive value."

The difficulty is, as Pike points out, it is difficult to avoid the Scylla and Charybdis of "artificial complexity" and "oversimplification" -- intonational definitions tend to be either overparticular, and thus empirically wrong (e.g., O'Connor and Arnold's insistance on the "disgruntled" character of Tone Group 3, in the face of their own examples), or over-general, and thus empirically inadequate (like Pike's "incomplete pointing," which is incapable of accounting for the effect of the *contradiction contour* on scope of negation). We will try to show a little later why these problems are especially acute in the study of intonation. For now, we only wish to observe that much the same problems of vagueness and non-compositionality that we attributed to certain cases of intonational "morphemes" would arise in an attempt to derive the meaning of *interdict* from the meanings of its component morphemes. The assumption that the minimal structural units of intonational patterns are necessarily also their minimal units of meaning seems to result in the same difficulties that would arise from the parallel assumption in more ordinary lexicography. Thus, there is clearly some relationship between the "high falling nuclear tone" ($\overset{s}{H}\ \overset{w}{L}$) in 3.2.5/2a and the parallel $\overset{s}{H}\ \overset{w}{L}$ sequence in 2b. It may also be true that these two "morphemes" share some core meaning like "personal participation in the situation." Nevertheless, it is not to be expected that the meaning of (for example) the sequence $\overset{w}{L}\ \overset{s}{H}\ \overset{w}{L}$ will necessarily be compositionally derived by adding up some core meaning of the "stem" $-\overset{s}{H}\ \overset{w}{L}$ and some core meaning of the "prefix" $\overset{w}{L}-$, especially if we require that the contribution of $\overset{w}{L}-$ be the same as it is in the vocative tune (and other occurances).

An additional complication in the analysis of intonational units arises because of the extremely strong role played by phonetic symbolism in constraining intonational meanings. In no other aspect of language is "l'arbitraire du signe" less manifest than in intonation, and we have every reason to believe that a substantial portion of the content of the intonational lexicon of English is determined by the universal symbolic (better: metaphorical) value of tones and tone-sequences. However, there are also many clear examples of language-specific tunes, and meanings for tunes, so that some degree of arbitrariness or conventionalization must be built into the system.

In the non-intonational lexicon, phonetic symbolism clearly cross-cuts morphology and even phonology, with non-distinctive oppositions (e.g., short/long, for English) and non-morphological sequences (e.g., -ink in wink and blink) often playing a key role.

Some psychologically significant intonational oppositions (e.g., terminally falling/terminally rising) should be seen as being of this nature. This is an extremely important issue. If it is true that the overall features "rising" and "falling" are of systematic significance, then any theory with static tonal segments seems to face a problem, since there is no feature inherent to such a theory which distinguishes (for example) the sequences /low + mid/ and /mid + high/ on the one hand, from the sequences /high + low/ and /high + mid/ on the other hand.

Our use of two distinctive features to specify the four tonal levels alleviates this problem slightly, since it becomes possible to specify "Archi-sequences" such as the sequence / $\begin{bmatrix}\text{+High}\end{bmatrix}$ $\begin{bmatrix}\text{-High}\end{bmatrix}$ / (which characterizes any of the set of sequences $\{$/high low-mid/ /high low/ /high-mid low-mid/ /high-mid low/$\}$). However, it is still not possible to express the features "rising sequence" and "falling sequence" in any simple way.

The necessary conclusion, for such a theory, is that the <u>rising/falling</u> distinction is not a direct characteristic of the phonology or morphology of the intonational system, but rather an overlaid distinction, a complex property of the systematic representation of the tune, like those distinctions which would be required in defining phonetic symbolism in general.

I think that there is excellent evidence that this is true. We have a sense that "rising" gestures in general share some property by opposition to "falling" gestures. Weak and strong beats in music are conceived of as rising and falling respectively (<u>arsis</u>/<u>thesis</u>, <u>levatio</u>/<u>positio</u>, <u>upbeat</u>/<u>downbeat</u> etc.). In dance, rising up on the toes is generally an arsic gesture, while coming down flatfooted is generally thetic. Raising the eyebrows is an other-directed gesture (greetings, expression of skepticism etc.), while lowering the eyebrows is a more self-directed gesture (signaling concentration, etc.). In sign languages, questions, nonterminal pauses etc. are usually signaled with an upward motion of the hands, while more "final" terminations are signaled with a downward motion (superimposed on whatever signs are being employed in the "utterance"). Examples could be multiplied indefinitely; the point is simply that "rising" and "falling" have some general metaphorical value independent of any role they may play in intonation, and that the roles which can in general be attributed to these concepts in intonation (e.g., other-directed vs. self-directed, nonfinal vs. final) are exactly what would be expected on the theory that we have proposed, that they are essentially paralinguistic metaphors.

The fact that the normal metaphorical value of "rising" and "falling" is sometimes violated in the case of particular intonational tunes, shows that universal sound symbolism does not completely determine the meaning of intonational words, although it obviously has a strong influence.

If we are to understand this situation, we would do well to examine the properties of conventionalized systems of sound-symbolism in general. Following a usage originally established to cover such phenomena in Bantu languages, and since extended to other cases, we will call these aspects of language ideophonic systems.

Ideophonic systems have five properties which will be of interest to us; the first and last of these they share with more conventional linguistic systems, while the remaining three tend to differentiate them from other aspects of language. It is the fourth property which most clearly sets ideophonic systems apart, but the other cited properties have an interesting tendency to correlate with it.

1) Ideophones are words; that is, they are made up of sequences of systematically distinctive elements (=phonemes), in patterns whose structure is determined by a morphology.

2) In general, the meaningful units in an ideophonic system are not directly given by the morphological analysis of a particular ideophone, but rather by some set of (more or less complex) properties defined on it.

3) The meanings of these units are typically <u>metaphorical</u> rather than <u>referential</u>; that is, they refer to a class of analagous aspects of different cognitive structures, rather than to any particular aspect of any particular such structure.

4) Ideophonic signs are not arbitrary -- the meanings of particular elements of an ideophonic system are strongly influenced by universal considerations. However, in any particular case, these form-meaning correspondances may become a specific, characteristic system, which is (usually) consistent with the universal basis, but is not entirely predicted by it.

5) Within a given system, "lexicalization" is possible -- that is, specific ideophonic words may take on particular meanings which are not predicted either by the universal basis, or by the particular system they belong to.

The most familiar linguistic examples of ideophones are echoic words, like English <u>clang</u>, <u>clank</u> etc. Words which are not exclusively echoic may also have an ideophonic component -- for example, it is not completely accidental that "gong" refers to a large metallic disk that gives a loud, resonant tone when struck, while "flute" refers to a high-pitched wind instrument. However, there are cases in which ideophonic systems extend far beyond the metaphorical relationship of the sound of a word to a non-linguistic sound.

For example, in Bahnar (Guillemet 1959, cited in Diffloth 1972) the words /blɔ:l/ and /blo:l/ are glossed as follows:

3.2.5/3a /blɔ:l/ 1. when a small fish quickly jumps out of the water.
2. when a man who has debts comes to your door or appears at your window.

3b /blo:l/ 1. when a big fish quickly jumps out of the water.
2. when an important person comes to your door or appears at your window.
3. when a great effort is made to reach an object which is out of reach.
4. suddenly speaking louder when one cannot be heard well.

Several aspects of the above example deserve comment. 1) we are dealing with words, made up of sequences of phonemes, not just free expressive noises; 2) the aspect of the word which is changed to produce a difference in meaning is (in this case) a single feature of a single phoneme, not a substitution of phonemes or sequences of phonemes; 3) the meanings of these words are extremely abstract properties, which pick out classes of situations related in some intuitively reasonable, but highly metaphorical way; the general "meaning" seems hopelessly vague and difficult to pin down, yet the application to a particular usage is vivid, effective and often very exact; 4) the particular phonological opposition which differentiates the two words, /ɔ/ vs. /o/, has a non-arbitrary connection to the meaning difference, which is something like small (or metaphorically related qualities, like "having debts") vs. large (or metaphorically related qualities like "important" or "loud."

This last point deserves some amplification. Suppose we make a partial listing of certain pairs of adjectives with intuitively corresponding properties:

3.2.4	large	small
	strong	weak
	important	unimportant
	loud	soft
	dominant	submissive

Now, there is some phonological feature opposition (say tense/lax) which characterizes the difference between Bahnar /o/ and /ɔ/. In the system of ideophones of which example /blo:l/ and /blɔ:l/ are members, this feature opposition (if it occurs in the proper position) has semantic content. But it is not at all clear that we want to say that tense means big while lax means little, and that the other meanings are metaphorical extensions of these core meanings. Rather, what seems to happen is that a systematic analogy is made between the phonological opposition "tense/lax" itself and the class of semantic oppositions big/little, important/unimportant etc., so that the actual "meaning" of the choice tense depends entirely on the nature of the situation to which one decides to apply the ideophone in question. Thus it may be true, in an ideophonic system, that the only meaning of a given element lies in the ability to analogize (in a systematic way) from its phonetic or phonological character to a large number of different concepts. In one sense, then, an ideophonic element means itself, given the human ability to make a kind of free-ranging metaphor out of its particular phonological or phonetic properties.

On this view, the non-arbitrary character of ideophonic sound-meaning correspondances, and their referential indeterminancy, the apparent abstractness of their meanings, are closely connected. This connection between the non-arbitrariness of the meaningful element and the essentially metaphorical character of the meaning is clearly exemplified in these examples from Korean (cited in Diffloth 1972):

3.2.5/5a tɔllɔŋ tɔllɔŋ
1. sound of small bells.
2. swaying movement of something suspended.
3. feeling of being left alone when everyone has gone.
4. someone appears flippant.

5b ttɔllɔŋ ttɔllɔŋ
1. sound of narrow bells, bells hit hard.
2. swaying movement of short object, tightly suspended.
3. feeling of being left alone when everyone has gone; shock of solitude comes more suddenly.

N.B. Diffloth cites more than twenty Korean examples from a "paradigm" created by holding constant the formal property "repeated disyllabic word with medial -l-"; this property seems to represent some meaning like "back and forth movement, oscillation, suspension, etc." Apparently there are thousands of possible words in the ideophonic system of Korean, including the possibility of nonce formations.

The meaning of the distinction t/tt , in these examples, is again best considered as the ability to construct a metaphorical connection between the phonological opposition itself and any one of a class of rather different concepts.

Following what I believe is a fairly standard usage, we will call this mode of meaning, in which the <u>signifié</u> is a general metaphorical extension of some intrinsic property of the <u>signifiant</u>, by the term <u>iconic</u>. The previously noted fact that the meaningful elements of ideophonic systems tend not to be their morphological units, but rather various properties defined on these units (single features in specified positions, certain sequences of features in certain positions, etc.) also follows from the fact that meaning in ideophonic systems is fundamentally iconic.

Iconic meaning is also characteristic of non-linguistic expressive noises, gestures etc. However, an interesting aspect of ideophonic systems is that they are linguistic, made up of phonemic sequences which are often arranged according to fairly restrictive morpheme structure constraints. As a result, they are more prone to conventionalization than paralinguistic systems generally are -- the range of possible metaphors is often restricted, the meanings of the ideophonic elements become less iconic and more arbitrary, and the degree of compositionality of the resulting ideophonic words may decrease, in just the same way that polymorphemic words in general decrease in compositionality across time.

In many cases (e.g. English) there is not a clearly identifiable ideophonic section of the lexicon, as there is in Bahnar, Korean, etc., but rather scattered classes of examples which have ideophonic or partly ideophonic character, and which shade off into areas where meanings are iconically arbitrary. An example would be certain classes of words for noises, like clang, clank, clink, click, clop, cluck, clomp, clunk, etc. A restriction of the metaphor to shape and consistency, rather than sound, is seen in glop and glob, and modes of fastening give us clip, clasp, and clamp. The system is of course far from complete -- climp and clont don't exist at all, while Clint is neither a noise, a mode of fastening, nor yet a smaller or sharper counterpart to glint, but simply a name. The fact that "cl-" is used for "noises with abrupt onset," while "gl-" is used for shapes, and for "attention-attracting emissions of light" (glow, gleam, glisten, glint etc.) is an example of apparently arbitrary restriction of ideophonic iconism.

The case of English thus shows us clearly that ideophonic analysis can intermingle with non-ideophonic sorts of lexical structure, in greater or lesser proportion.

To sum up our view of the situation: We make a fundamental distinction between ideophonic and morphemic modes of lexical structure, that is, modes of analysis of <u>words</u>. Ideophonic systems are fundamentally iconic in their mode of meaning. Their other characteristic properties (referential vagueness, meaningful elements not partitioning the string but intersecting in complex ways, etc.) follow from this fact. Morphemic systems are fundamentally arbitrary in their mode of meaning. Their other characteristic properties (referential precision, typically clear partitioning of the string into separable formatives, etc.) follow from this arbitrariness (the referential precision for obvious reasons, and the normal status of morphemes as underlyingly non-intersecting fragments because the assignment of arbitrary meanings to arbitrary strings of partly-specified feature matrices would result in hopeless ambiguity). Both ideophonically and morphemically derived words are subject to lexicalization, that is, to greater or lesser degrees of non-compositionality. It is possible for a given word to have both ideophonic and morphemic analyses independently, and most lexical systems have this mixed character to some extent.

It should now be clear how we propose to effect a synthesis between the "structural" and "phonesthetic" views of intonation. The intonational lexicon, in our proposal, has a fundamentally ideophonic structure. This implies that intonational words are made up of strings of phonemically distinct segments -- we have specified that these segments are defined by the distinctive features $\pm$ High, $\pm$ Low. These strings of segments may be arranged in morphologically canonical ways -- thus one important class of tunes in English has the form

3.2.5/6 $\quad (\overset{B}{[\ \ ]})\ \overset{w}{[\ \ ]}\ \overset{s}{[\ \ ]}\ \overset{w}{[\ \ ]}\ (\overset{B}{[\ \ ]})$

A complete specification of the feature matrices in 6 above is an intonational word -- like other ideophones, such a word can be a nonce formation, whose meaning is given compositionally by the sum of its "parts," or it can be more or less "frozen." The meaning-bearing "parts" which go to make up the whole range may be the presence or absence of a given feature in a given position, certain sequences of features in certain positions, or certain more abstract properties of the whole array. Examples of all of these cases will be given in due time. The fact that intonational meanings often seem hopelessly vague and difficult to gloss, but can have vivid, effective and often quite precise effects in their application to a particular usage or class of usages, is parallel to the observations often made in the literature about ideophonic meanings in general.

Metaphor, in human experience as a whole, resists reduction to any formal system, but is a uniquely powerful and efficient means for certain kinds of understanding and communication.

3.2.6 A Family of Tunes.

The members of one important morphological group of tunes in English consist of a sequence of three tones, with the metrical pattern w s w, flanked by a pair of optional boundary tones. The existence of this structure has been recognized, explicitly or implicitly, by most of the careful investigators of English intonation. For example, Trager (Trager 1961) writes: "American English intonational patterns consist typically of three pitches and a terminal contour. . .The central pitch accompanies the primary stress of a phrase or clause. . . When a clause begins with the primary-stressed syllabic, there are only two pitches, the central and the final, the initial being absent."

The surprise/redundancy tune, which we examined earlier, is a representative of this group, having the feature specification:

3.2.6/1 $\begin{bmatrix}\text{-High}\\ \text{+Low}\end{bmatrix}$ $\begin{bmatrix}\text{+High}\\ \text{-Low}\end{bmatrix}$ $\begin{bmatrix}\text{-High}\\ \text{+Low}\end{bmatrix}$

T_1 T_2 T_3

Varying the feature $\pm$Low in T_2 does not change the basic pragmatic value of the tune, although it does introduce a difference, to be described shortly. The variation in question yields the following two outputs:

3.2.6/2a	Low, High, Low	2b	Low, High-Mid, Low
	+High -Low		+High +Low

At this point, we have to face the problem which is posed by the following two facts: 1) Expansion and contraction of the <u>tessitura</u> (or pitch-range) may produce effects phonetically similar to variation in feature composition of individual tones. Thus a High tone, given a more contracted tessitura, may represent the same pitch as a High-mid tone, given a more expanded tessitura. If it is true that the features $\pm$High, $\pm$Low have articulatory content (and I expect that this is so), with pitch-range differences being the result of differences in the force of the articulations involved, then the two situations (high tone/contracted pitch range, high-mid tone/expanded pitch range) are articulatorily different, and very possibly perceptually different as well. Something of the sort seems to be true for tone languages, in any event. 2) Nevertheless, to the extent that a given aspect of intonational meaning is <u>iconic</u>, the pragmatic effects of varying the tonal features, and varying the pitch-range features, may be similar.

We could make a comparison to another domain -- in a language with phonemically significant vowel length, it is possible that a token of a long vowel, in allegro speech, may be as short, phonetically, as a short vowel in lento speech (although a native hearer would not be fooled, of course). However, if the words in question are not ideophonic, then the effect on meaning of the short/long distinction will have no connection at all with the expressive value of the allegro/lento distinction. In an ideophonic system, on the other hand, these two distinctions might have similar (though probably not identical) iconic value.

In order to cope with this situation, we will make two simplifying assumptions: 1) Even in purely iconic modes of meaning, the values of tonal features and pitch-range features are distinct enough to be worth studying individually; if grammaticalized, they will be grammaticalized independently. 2) In a single utterance, a speaker will not change tessitura.

Assumption 1) is probably true. Assumption 2) is certainly false -- we know at least that there often is a progressive narrowing of tessitura, called <u>downstep</u> in African tone language, and <u>declination</u> in most studies of English-type intonational phenomena. Other effects on tessitura, both paralinguistic and grammaticalized, undoubtedly occur. We disregard them, by choice, in the discussion that follows.

Returning to the discussion of the opposition $\pm$Low in position T_2 of the surprise/redundancy contour, we may illustrate its effects in some particular cases. The nature of the iconic metaphor is this: the basic gesture is rising and falling, with the peak of the rise stressed. The phonetic (and probably articulatory) effect of the feature +Low in position T_2 is to check or restrain the rise, resulting in a lower peak; the effect of -Low is to allow the rising gesture to proceed without restraint. If the basic expressive value of the gesture is <u>surprise</u>, then the metaphoric equivalent of "restraint" <u>vs</u>. "lack of restraint" in the central portion of the gesture might be something like "surprise tempered by some other attitude" (worry, ratiocination etc.) vs. "simple surprise."

3.2.6/3a <u>Detective looks up from his examination of suspect's safe-deposit box, knits his brows, and says in a puzzled tone</u>:

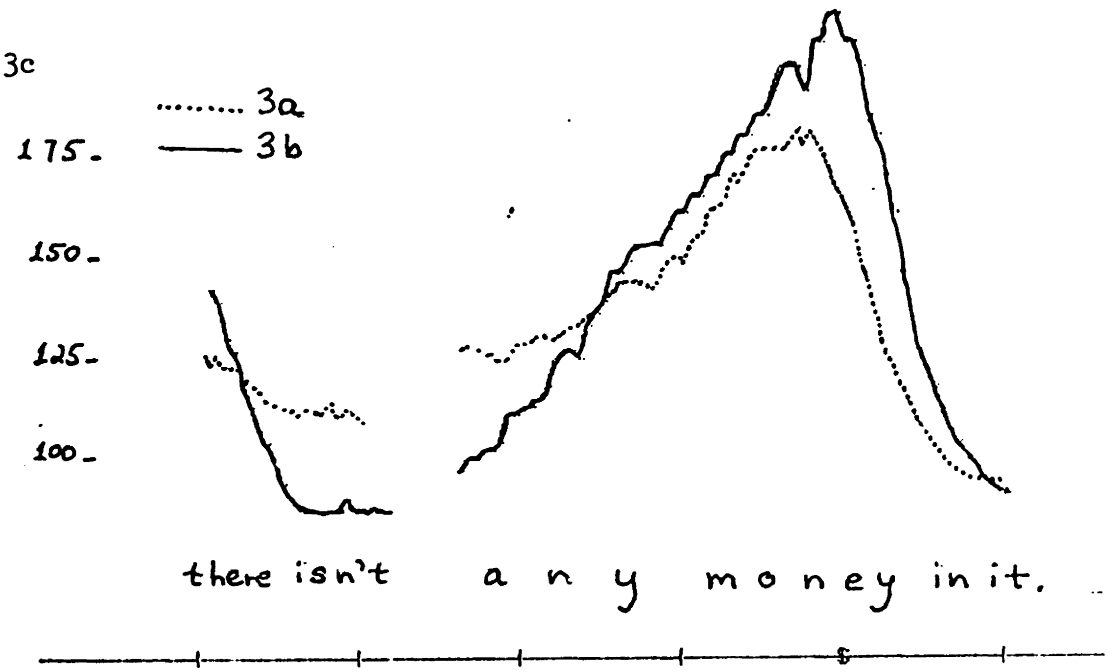
```
There isn't any money in it!     (expresses surprise,
          |          |   |        tempered by consideration
          L         H-M  L        of where else the loot
                                  might be stashed).
```

3b <u>Same detective, same situation; his jaw drops, and he says in witless amazement</u>:

```
There isn't any money in it!     (expresses surprise,
          |          |   |        simple and unalloyed).
          L          H   L
```

N.B. It is not claimed that this opposition is binary or discrete in expressive value, although it is based on binary and discrete features, since it is possible to "rein in" the rising laryngeal gesture to various extents by +Low articulations of varying degrees of strength, with obvious iconic results. This effect is parallel to the degrees of "terminal sonority" implied by varying degrees of expressive gemination in "clang," "clangng," "clangngng" etc., all of which are in opposition to "clank."

3c



N.B. Both of these examples show a H-M initial boundary tone, which serves only to increase the "vehemence" of the expression. The particular example of the intonation in 3a whose F_0 contour we see in the dotted line, also has L-M in place of the original L, a possibility we have alluded to before, and will discuss in its own right a little later.

Given the fertility of the human mind, other metaphorical modifications of <u>surprise</u> by the feature ±Low in central position (T_2) are doubtless possible. If the basic expressive value of this gesture were "the content of this utterance is redundant," then the analogues of <u>restrained</u> vs. <u>unrestrained</u> gestures might be "expression of redundancy restrained by consideration of mitigating circumstances, social taboo against imputing idiocy to others, etc." vs. "simple expression of redundancy." Again, other interpretations are possible, and the distinction may well be nuanced.

The facts just considered suggest that the basic form of the S/R contour should be

3.2.6/4
$$\begin{array}{ccc} \text{w} & \text{s} & \text{w} \\ \begin{bmatrix}-\text{High}\\+\text{Low}\end{bmatrix} & \begin{bmatrix}+\text{High}\end{bmatrix} & \begin{bmatrix}-\text{High}\\+\text{Low}\end{bmatrix} \\ T_1 & T_2 , & T_3 \end{array}$$

with the distinction $\pm$Low in position T_2 open to fairly free iconic interpretation. We might now ask what the effect is of making other changes in the feature specification for the three positions. Simple arithmetic tells us that there are 64 possible combinations -- we will not attempt to cover all possibilities, nor even to discover if all of them exist. We will, however, go through a few of the more interesting variants.

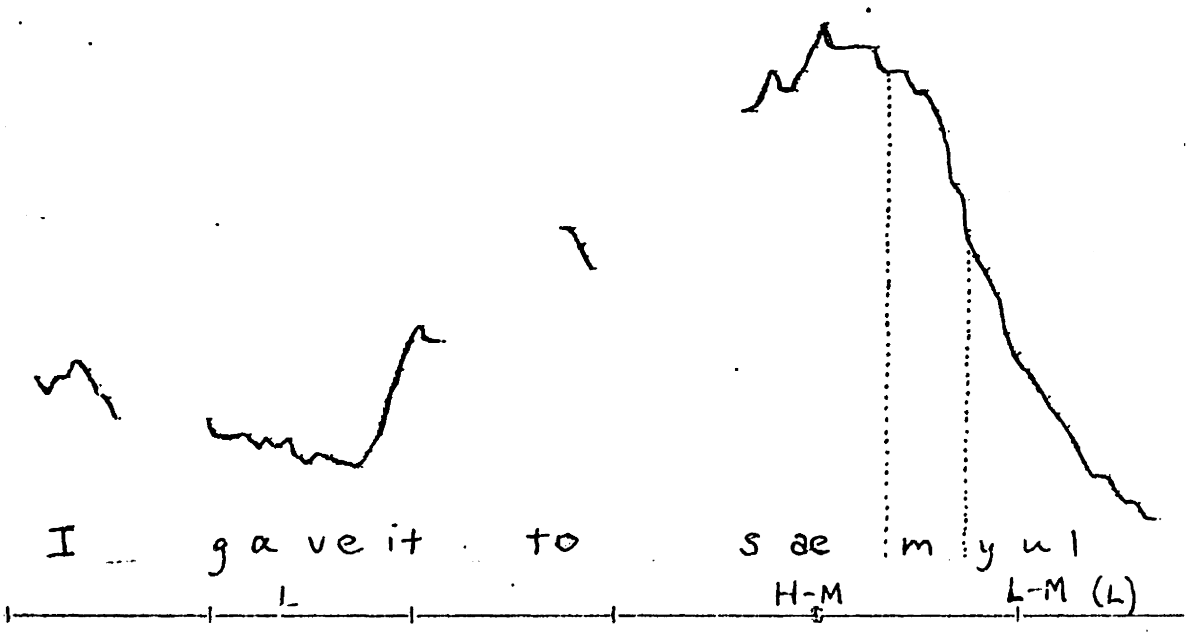
First, we have observed in regard to 3.2.5/1b and 3.2.6/3c that it is possible to substitute a low-mid tone for the initial low in the pattern given in 3.2.6/4, without enormous effect on the "meaning." The effect of the feature +Low in the initial position is to exaggerate the "upbeat" of an overall "low high low" pattern, and thus to exaggerate the pattern itself: we disagreed with O'Connor and Arnold's estimate that the +Low version expresses "disgruntlement" or some similar attitude, but it is clearly the marked option, and needs to have some reason for its occurrence. It is not clear that it is necessary to say anything more precise than that the version with +Low in initial position is <u>exaggerated</u>. It is also possible for the terminal fall to be arrested at a mid pitch. We would expect that the effect of the "neutral" low-mid tone in final position would be to give a relatively tentative or indefinite air to the ending, as opposed to the firmer and more definite termination of low tone.

This prediction is borne out by the facts, as the following pairs of examples suggests:

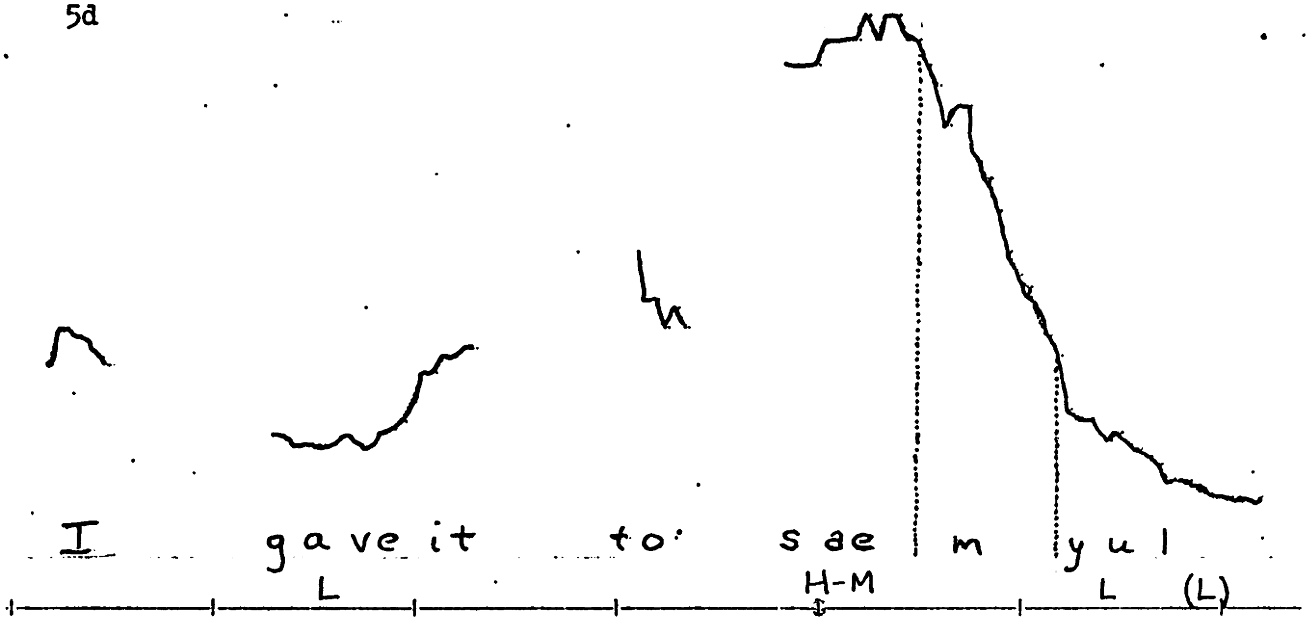
3.2.6/5a I gave it to Samuel.
L H-M L-M

5b I gave it to Samuel.
L H-M L

5c



5d



Interestingly, the terminal pitch drops to about the same level instrumentally in both cases, although perceptually 5d seems to end lower. 5c is more neutral or tentative, while 5d is more firm and positive, and may even seem abrupt or angry. The primary difference is that in 5d the pitch has already reached quite a low level at the release of the m (85 H_2 within 10 msec of the release) whereas in 5c, the pitch is by no means so low at the comparable point (about 135 H_2 10 msec after release). In both cases, the pitch trails off to about 75 H_2. We have proposed that the difference in pitch at the onset of the syllabic nucleus is due to a difference in the tone assigned to that syllable (low in one case, low-mid in the other). The reason for the trailing off is not so clear -- it might be attributed to the dying away of subglottal preserve (which would not be countered by any strong contrary action of the "neutral" low-mid tone), or it might be attributed to a L terminal boundary tone (represented in parentheses under the F_0 plots).

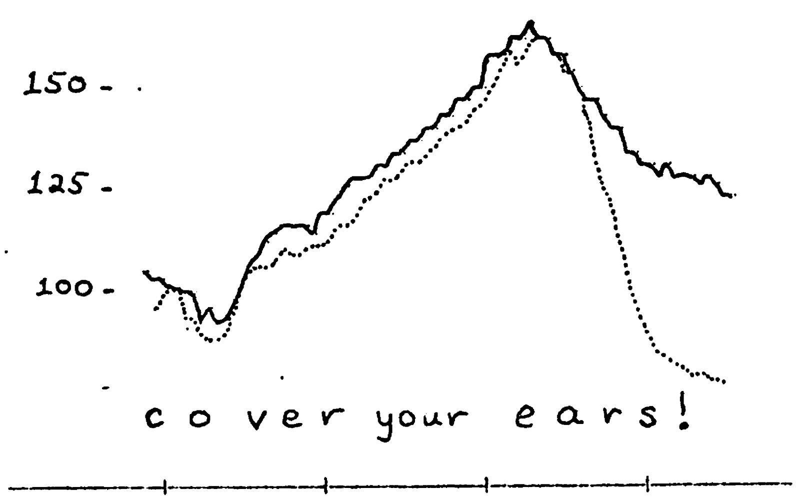
If we substitute a high-mid pitch in the terminal position, the effect is quite different. There is no question of the pitch trailing off to 75 H_2 as the breath runs out, or the necessity of a low terminal boundary tone, or whatever: the pitch stays up. When the terminal sequence is H H-M, we have the vocative tune, which also functions in unchanted speech in a way to be examined shortly; when the terminal sequence is H-M H-M, we seem to have another, rather different entity, which we will also take a look at.

The sequence $\left\{\begin{matrix}\text{L-M}\\ \text{L}\end{matrix}\right\}$ H H-M (what we called the vocative tune) occurs as well on (more or less) jocular admonitions:

3.2.6/6a Cover your ears!
L H H-M

6b Cover your ears!
L H L

6c



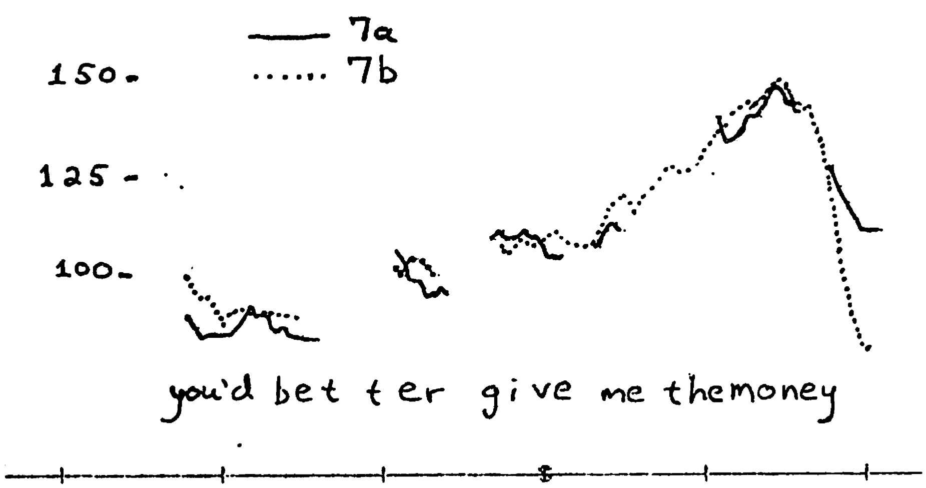
The version in 6a has a kind of playful, finger-wagging quality -- the version in 6b, in which the fall is all the way to low (or to low-mid, it wouldn't matter) has a very different implication; in this case, something like "I've told you a thousand times."

The /L H H-M/ tune is not always jocular -- there can be an undercurrent of menace, as in the example that follows. However, there is still a marked difference with the surprise/redundancy tune:

```
3.2.6/7a      You better give me the money.     ("or you'll be sorry...")
                  |                   / |
                  L                  H  H-M

      7b      You better give me the money.     ("obviously that's what
                  |                   / |         you should do...")
                  L                   H L
```

7c

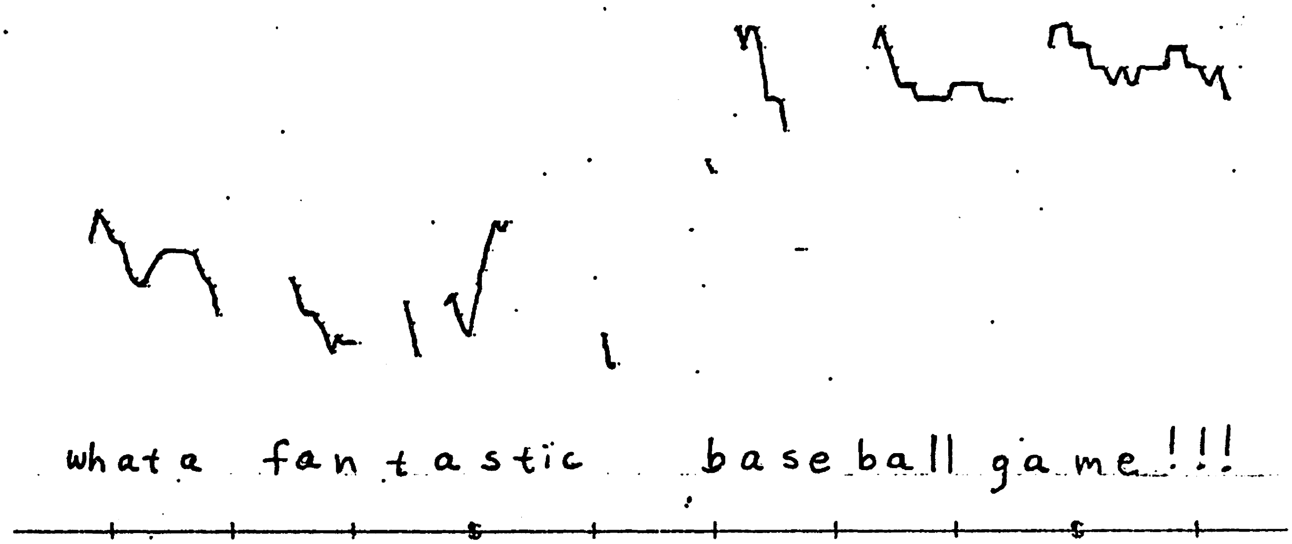


Thus the surprise/redundancy tune has the "skeleton" [-High] [+High] [-High], with the feature [±Low] in the various positions serving to "modulate" the effect -- roughly, +Low in initial position signifies <u>exaggerated</u>, +Low in medial position signifies <u>restrained</u>, and +Low in terminal position signifies <u>definite</u>, <u>final</u>. We have seen that changing the terminal position $[-\text{High}]$ to $[+\text{High}]$ produces a very different tune, which (given the terminal sequence H H-M) lends an admonitory air to the utterance, and can also be used for calling people (the "vocative tune") -- we will call this sequence $\begin{bmatrix}-\text{High}\\ \pm\text{Low}\end{bmatrix} \begin{bmatrix}+\text{High}\\ -\text{Low}\end{bmatrix} \begin{bmatrix}+\text{High}\\ +\text{Low}\end{bmatrix}$ the <u>warning</u>/<u>calling</u> tune.

Both the <u>surprise</u>/<u>redundancy</u> tune family, $\begin{Bmatrix}\text{L-M}\\ \text{L}\end{Bmatrix} \begin{Bmatrix}\text{H}\\ \text{H-M}\end{Bmatrix} \begin{Bmatrix}\text{L-M}\\ \text{L}\end{Bmatrix}$, and the <u>warning</u>/<u>calling</u> tune family, $\begin{Bmatrix}\text{L-M}\\ \text{L}\end{Bmatrix}$ H H-M, have in common that they are basically rising-falling gestures. If the terminal sequence becomes /H-M H-M/ or /H H/, we have a "level nuclear tone," and (perhaps because of this), its meaning is rather different from that of the <u>warning</u>/<u>calling</u> tune, from which it differs in only one tonal feature. When one of these "high level nuclear tones" occurs in nonfinal clauses, the effect is simply one of "continuation," but standing by itself, it is a highly marked, very emotionally charged tune --

```
3.2.6/8a      What a fantastic baseball game!!!
                     |        \        \
                     L       H-M      H-M
```

8b



Particularly if any syllables follow the main stress, the only way I can perform this tune naturally is to feign such excitement that the level terminal stretch is performed with <u>creaky</u> <u>voice</u> -- otherwise the contour sounds like singing instead of speaking, and seems unnatural. However, I have observed that some people do "sing" such exclamations.

We will consider one last class of modifications of the basic pattern we have been investigating, those which have "low falling nuclear tone" (which we will spell $\begin{array}{cc} s & w \\ \begin{bmatrix}\text{-High}\\ \text{-Low}\end{bmatrix} & \begin{bmatrix}\text{-High}\\ \text{+Low}\end{bmatrix}\end{array}$). "Low falling nuclear tone" is characteristic of O'Connor and Arnold's Tone Groups 1 and 2. Tone Group 1 consists of the string /(Low Pre-head+)(Low Head+) Low Fall/. Tone Group 2 consists of the string /(Low Pre-head+) Stepping Head+Low Fall/ <u>or</u> / High Pre-head+Low Fall/.

We have explained their use of the term "Pre-head" (everything preceding the stressed syllable of the first "prominent" word). The Nucleus (represented in these examples by "Low Fall") is the "stressed syllable of the last prominent word," recognized as "clearly. . . a landmark of the highest importance," and as the pivot on which "the whole tune centres." Any syllables following the nucleus are the "tail;" often the "Nuclear tone" is actually distributed over both nucleus and tail, obviously.

What they call "stepping head" is defined as follows: "the stressed syllable of the first important word is on a high, level pitch; that of the second word is a step lower; that of the third a step lower still, and so on until the nucleus is reached."

Crystal strongly criticizes the idea that O'Connor and Arnold's three-way classification of heads (low, sliding and stepping) is adequate, or that the particular types involved are of major importance. He observes that "the head of the tone unit is probably the most complex segment to describe," and proposes a four-way distinction into <u>falling heads</u> (with four main subtypes), <u>rising heads</u> (with two main subtypes), <u>falling-rising</u> (-<u>falling</u>) heads, and <u>rising-falling</u> (-<u>rising</u>) <u>heads</u>. He argues that "statistical analysis suggests that this classification is not too detailed: if anything, further sub-classification is needed, to anticipate the additional complexity. . . [of] longer heads. . ."

All of this suggests that the concept of "head" as a structural unit is simply not adequate to deal with other than the simplest cases. But we may assume that O'Connor and Arnold's distinction between Tone

Groups 1 and 2 has a real basis; so we will abstract away from the allegedly "stepping" nature of the Tone Group 2 head, and consider the distinction to be basically <u>low</u> vs. <u>high</u>. (Actually, a large percentage of their example of Tone Group 2 have only one "step," i.e. a high tone, anyway).

On this assumption, incidentally, it becomes highly interesting that Tone Group 2 normally has the (unmarked) low pre-head, but if there is a pre-head without a head, then the "pre-head" must be high. If the true "spelling" of Tone Group 2 is something like $/\overset{w}{\begin{bmatrix}+\text{High}\end{bmatrix}}\ \overset{s}{\begin{bmatrix}-\text{High}\\-\text{Low}\end{bmatrix}}\ \overset{w}{\begin{bmatrix}-\text{High}\\+\text{Low}\end{bmatrix}}/$, then the initial $\begin{bmatrix}+\text{High}\end{bmatrix}$ will go, equally predictably, on some secondary stress in the "head," or on some unstressed initial syllable (by British school definition, part of the "pre-head") if no secondary stresses precede the main stress. Facts like this strongly suggest that the notions <u>head</u> and <u>pre-head</u> have no systematic validity, although they are perhaps descriptively useful.

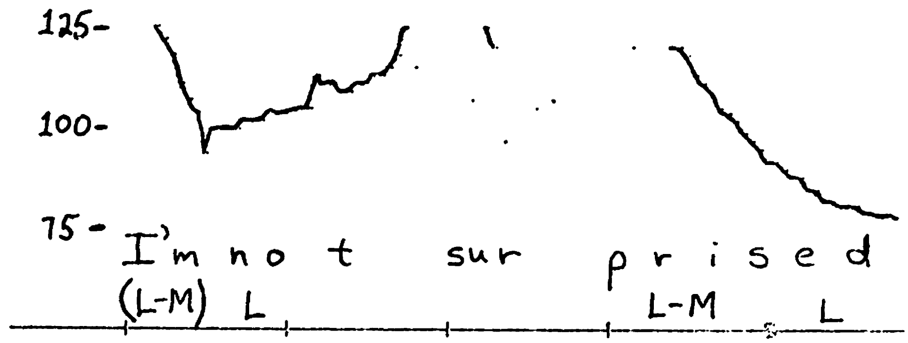
In any event, Tone Group 1 (as befits the "lowest" possible spelling of the rising-falling gesture, $/\begin{bmatrix}-\text{High}\\\pm\text{Low}\end{bmatrix}\ \begin{bmatrix}-\text{High}\\-\text{Low}\end{bmatrix}\ \begin{bmatrix}-\text{High}\\+\text{Low}\end{bmatrix}/$) is glossed as "characteristically used to convey a <u>cool</u>, <u>calm</u>, <u>phlegmatic</u>, <u>detached</u>, <u>reserved</u>, <u>dispassionate</u>, <u>dull</u>, possibly <u>grim</u> or <u>surly</u> attitude." The version with $\begin{bmatrix}-\text{High}\\-\text{Low}\end{bmatrix}$ in initial position (which O'Connor and Arnold do not differentiate) seems, if anything, more phlegmatic and duller.

Tone Group 2 (which we are transcribing as $/\begin{bmatrix}+\text{High}\\\pm\text{Low}\end{bmatrix}\ \begin{bmatrix}-\text{High}\\-\text{Low}\end{bmatrix}\ \begin{bmatrix}-\text{High}\\+\text{Low}\end{bmatrix}/$) is distinguished from the other cases we have examined so far by virtue of being entirely falling. It is glossed as "used to give a <u>categoric</u>, considered, weighty, judicial, dispassionate character to statements;" it is said that wh-questions are rendered "searching, serious, intense, responsible. . ." while commands are "firm, serious, considered, weighty, pressing, dispassionate." It is added that "a ring of impatience" or a sense of "irritability" are often conveyed.

O'Connor and Arnold's glosses for these tonal entities seem exactly right; we will add nothing and subtract nothing, except to point out the ideophonic appropriateness of the connections.

Here are some of their examples (in our notation, with F_0 contours as performed by the author) -- to enliven the presentation of the list, we will use individual cases to bring out certain points to which we will otherwise pay little attention:

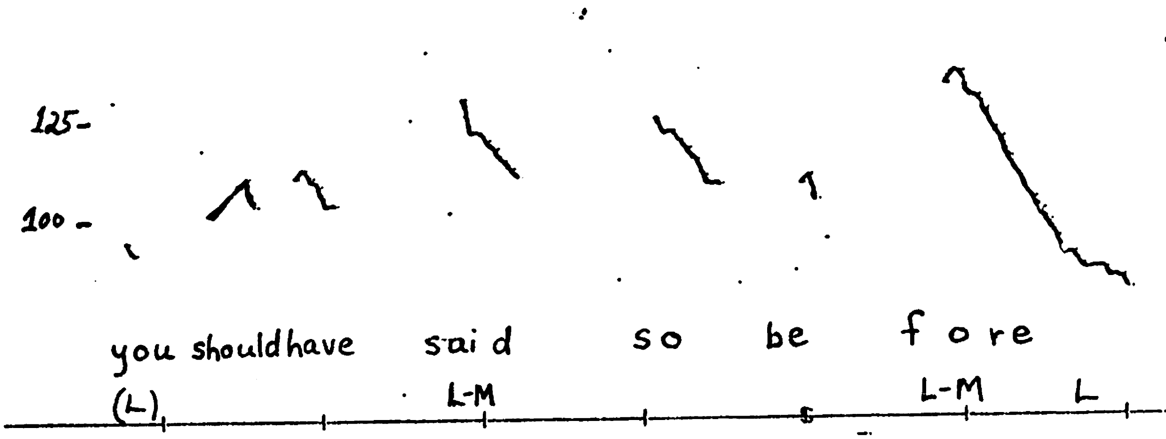
3.2.6/9



The above is an excellent example of the "phlegmatic, dull" Tone Group 1. I have analyzed it as consisting of a low-mid initial boundary tone (which is the most unmarked case), followed by the tonal

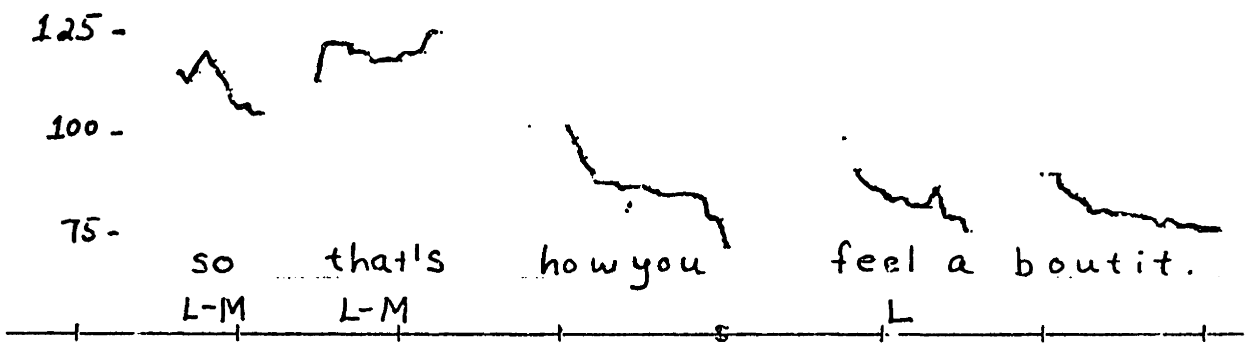
w s w
sequence L L-M L. In this (and the rest of the examples in this section) we will put boundary tones in parentheses, in the notation beneath the F_0 plots, in order to distinguish them from the rest of the tonal unit. The fact that the initial low is somewhat higher than the terminal low is fairly normal for my voice when the nonterminal low is not strongly stressed.

3.2.6/10



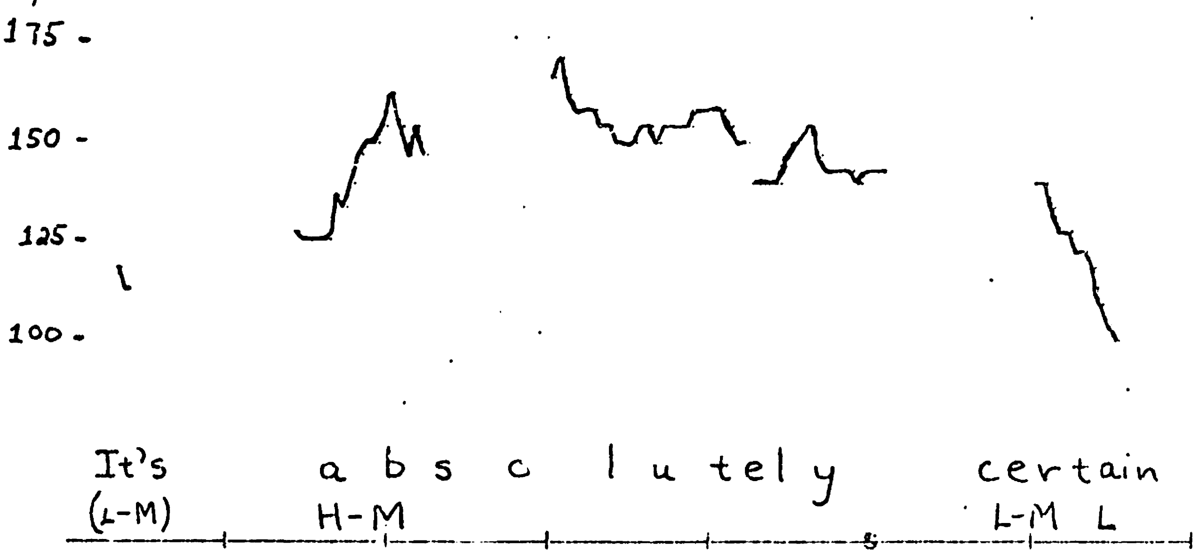
The fact that the syllables said, so etc. individually have a falling shape is due to their phonetic content -- the initial voiceless consonant tends to cause a rise in F_0 (noticeable only after its release, of course), while there is a lowering in F_0 before an obstruent, in general. Jacqueline Vaissiere has proposed the name micromelody for these phonetically-caused perturbations in F_0 superimposed on the overall intonation contour. The fact that the stressed syllable of before begins somewhat higher than the earlier L-M syllables do can be attributed to the fact that the heightening effect of initial voiceless consonants depends somewhat on stress, at least in my experience. (the actual values for the maxima, as recorded by the pitch detector, which is not necessarily reliable too near to consonantal articulations, are: said 130 H_2, so 125 H_2, -fore 135 H_2).

3.2.6/11



The above example of Tone Group 1 (the "phlegmatic tune") shows a rather long "tail" -- we will comment further in a later section on the fact that all syllables following the main stress of an intonational phrase lack independent tonal specification, with the single exception of the possibility of a terminal boundary tone.

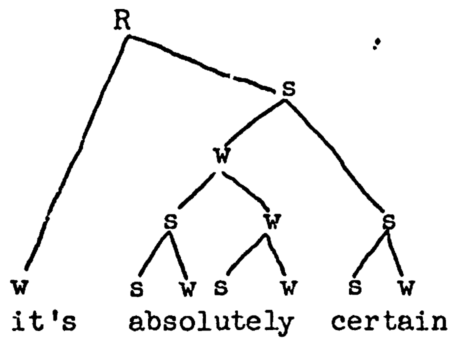
3.2.6/12



The preceding example is our first case of Tone Group 2 (the "judicious tune"). We can use it to make a point which has been implicit in much of the preceding discussion, and which was brought out in an embryonic form in chapter 2 -- the treatment of non-lexical words.

If the metrical pattern of the text "it's absolutely certain" were to follow its surface structure, it would be:

3.2.6/13



N.B. We are assuming the operation of a "rhythm rule" to shift the main stress of <u>absolutely</u> back to its initial syllable, in this particular case.

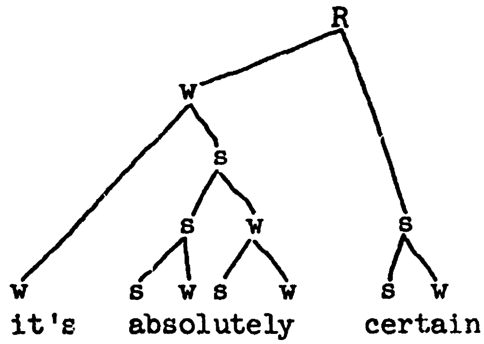
This structure would predict the tone assignment

"it's absolutely certain." While this represents a conceivable way
H-M L-M L

to say the sentence, it requires a rather strong stress on the word <u>it</u>, which does not normally take stress unless it is used deictically (i.e., in pointing to an entity new to the discourse, rather than anaphorically referring to something in the background of the discourse). This "stress" on <u>it's</u> results from the principles of grid alignment discussed in chapter 2, in reference to 2.4.4/11 and subsequent examples.

We have been assuming that in the normal case, the kind of structure in 3.2.3/13 is readjusted to give something like the following:

3.2.6/14



This metrical pattern will predict both the relative stresslessness of it's, and the correct tune-text association for 3.2.6/12. The readjustment involved in producing this metrical pattern is obviously not unrelated to the phenomenon of cliticization, in which certain non-lexical words become attached to lexical words in their vicinity. In general, the intonational treatment of these non-lexical words will precede normally, according to our theory of tune-text association, if this attachment is required unless the formative in question is being given special prominence (as in deictic or contrastive use of pronouns, etc.). Some nonlexical words (e.g. not) seem normally to have a certain prominence, so that in the usual case they do not undergo such attachment, although it is certainly possible for them to "glom on" intonationally.

I believe that the rationale we gave for this attachment process in chapter 2 is a valid one -- unless it happens, a formative whose value is only structural, or is otherwise redundant (e.g. it's) would occupy a more important metrical position than part of the real content of the utterance (e.g., absolutely).

Subsequent to this rebracketing (which is really a kind of enforced subordination) there is reason to believe that word boundaries may delete, causing unstressed nonlexical monosyllables to become a single word with one of their neighbors (cf. Selkirk 1972). The issue is a complex one, since there are tonally similar cases not involving enclitics. Suppose we are asked the distance from Cambridge to Milton. This distance is approximately thirteen miles; if we had some reason to suppose that this information was, or ought to be, well known to our interlocutor, we might reply, employing the surprise/redundancy tune:

```
3.2.6/15        Thirteen miles.
                  |      |\
                  L      H-M L
```

If we wished to heed our schoolteachers' injunctions to use complete sentences, we might answer instead:

```
3.2.6/16        It's thirteen miles.
                       |       |\
                       L       H-M L
```

The existence of the initial <u>it's</u> would also give us a place to put a high initial boundary tone, to indicate vehemence, if we chose; but the initial tone of the /L H-M L/ tonal unit would still go on the stressed syllable of <u>thirteen</u> (unless for some reason, e.g., the existence of a list of such questions and answers, we wished to contrastively stress the pronoun). This treatment of the initial <u>it's</u> would be accounted for by the normal "cliticization" of unstressed pronouns and auxiliaries.

However, we might decide to make our answer even more unnecessarily complex, and throw in the whole noun phrase instead of a mere pronoun:

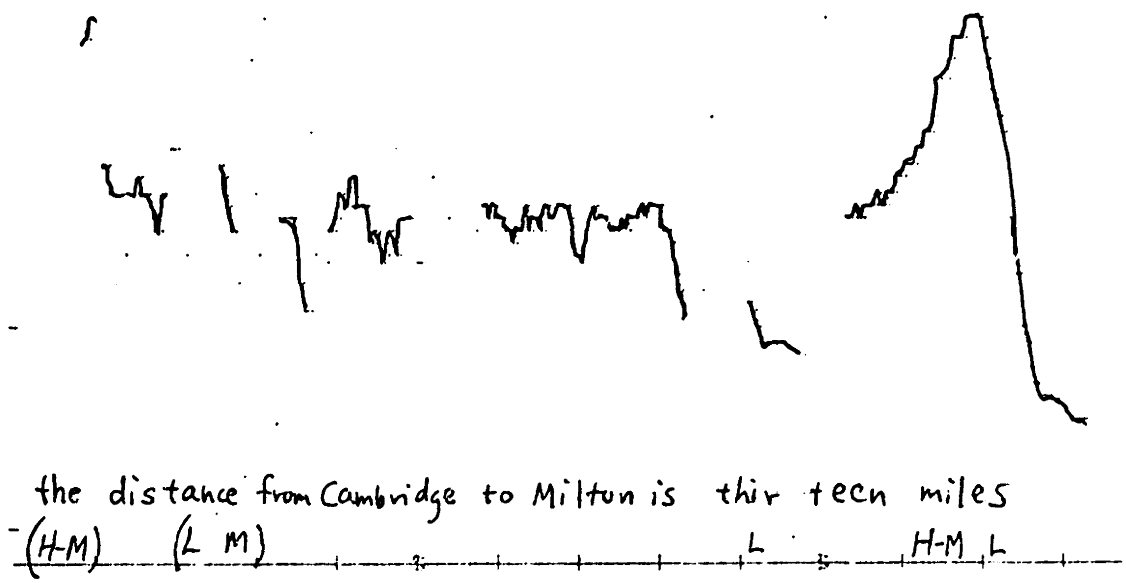
"the distance from Cambridge to Milton is thirteen miles." There are a large number of ways in which such an answer might be intoned -- we wish to examine just one of them, the one whose underlying tonal association is given below:

3.2.6/17 The distance from Cambridge to Milton is thirteen miles.

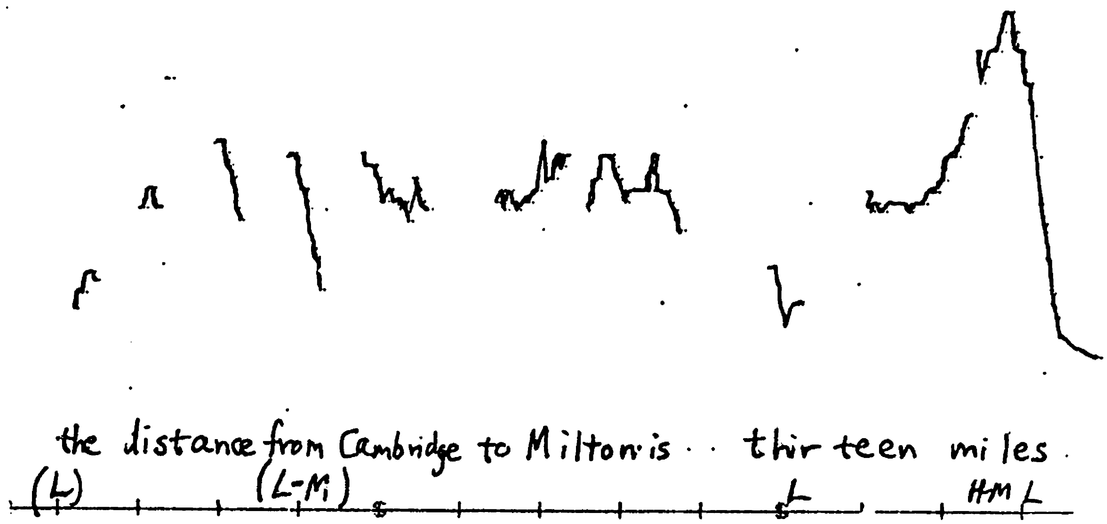
L H-M L

In this example, the subject noun phrase has undergone what is sometimes referred to as "anaphoric destressing," or "redundancy destressing." Depending on our choice of initial boundary tones, the F_0 contour might look like 18a or 18b below:

3.2.6/18a



3.2.6/18b



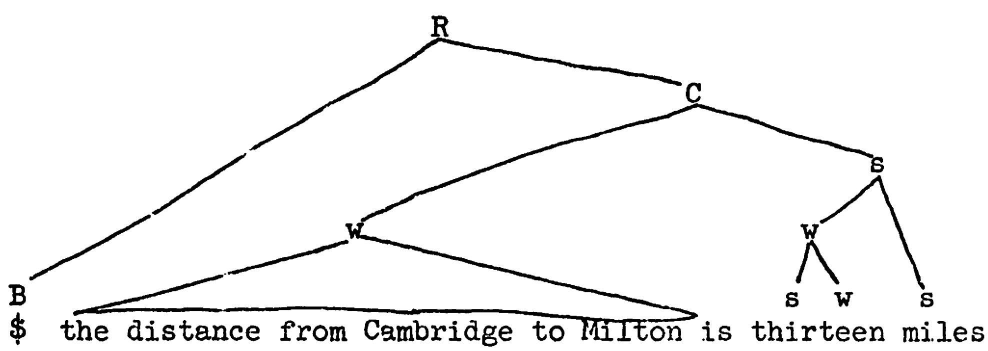
The point which interests us is that (with the exception of the initial boundary tone) the subject NP does not participate in any aspect of the tune that is familiar to us. The tonal unit /L H-M L/ is exclusively associated with the <u>answer</u>, "thirteen miles," which as we observed could have stood alone as our contribution to the discourse. In British school terminology (at least according to the suggestions given in O'Connor and Arnold) the whole subject NP plus copula is in the "pre-head," thus giving us a ten-syllable pre-head.

How are we to represent this phenomenon? It will not do to say that redundant or anaphoric material cannot receive (non-boundary) tones -- if there is nothing "better" around, then they will get the tone, as we have seen. I think that it would also be a mistake to try to assimilate this case entirely to that of the non-lexical words, although they do have something in common both pragmatically and intonationally. In the case of nonlexical words, there are a number of reasons for us to re-attach them lower down in the metrical tree -- not the least of these is that the unstressed monosyllabic cases typically become phonological units with a neighboring word. In the next chapter we will propose a metrical treatment of stress rules; if this proposal is valid (and it has some striking advantages over the traditional view) then metrical structure must exist throughout the phonology. Indeed, if the metrical pattern forms the basis of tune-text association, a process which seems to follow all of the textual phonology, then this tree-structure must still be defined on the output of the phonology, by obvious implication. In our summation (chapter 6), we will present an idea about the nature of metrical structure according to which it is reasonable and even inevitable that such structure should remain defined on the systematic phonetic level.

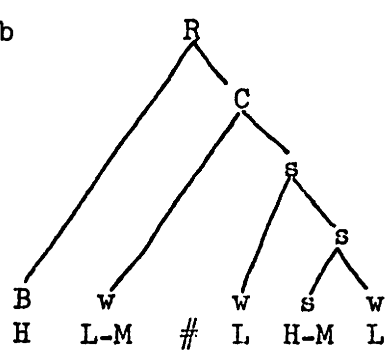
In any case, if metrical patterns are defined at the time of cliticization, as in our theory they must be, then the process of coalescence must be preceded by a rebracketing which has the effect of putting the future clitic into a metrical constituent with the word it will cliticize onto. The "coalescence" process itself (which ensures that the clitic becomes a single word with its parent) may, on this view, be a separate phenomenon, with distinct conditioning; thus the constraints on cliticization may be an intersection of the constraints on metrical rebracketing, and the constraints on a boundary-deletion process. This is rather different from the traditional view, but will not be defended further at this time.

Whatever the true nature of cliticization, it will not do to represent 3.2.6/18 entirely as a case of "super-cliticization." For one thing, the "lack of tonal specification" of the long "pragmatic prefix" emerges in a rather specific way, namely as a low-mid tone. This is rather similar to an aspect of the treatment of tags (which could be viewed as "pragmatic suffixes") that will be pointed out in section 3.3. Thus there is perhaps some reason to view 3.2.6/18 as a morphologically complex tonal word, consisting of an initial boundary tone and neutral low-mid tone, followed by the expected /L H-M L/ "stem" of the tune. If this view is correct, then the attachment of the "stem" of the tune to the portion of the text that constitutes the "answer," would not in this case require any special operations on the metrical pattern:

3.2.6/19a



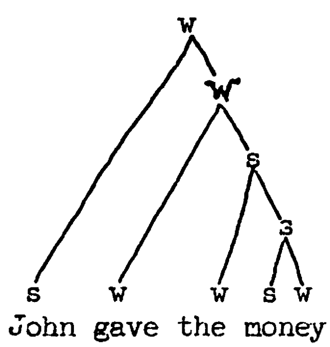
19b



N.B. The # in the tune is intended only to indicate the presence of some kind of morphological structure, an issue to which we will return briefly in the next section.

19a and 19b will define the pattern of tune-text association seen in 18a, without any deformation of the metrical structure of the text. But this will not always be so easy -- for example, if the question had been "Who did John give the money to?" we might have answered "John gave the money to Susan's uncle," putting the "stem" of the tonal word on the answer "(to) Susan's uncle," and associating the L-M prefix with [John gave the money] , which is not a syntactic constituent under any normal conditions.

My belief is that the correct approach to such cases is roughly the following: a certain class of material (including redundant elements and non-lexical words, among other categories) is "cut loose" from the metrical tree and readjoined, towards its left if possible, with neighboring constituents. This is the process that allows unstressed auxiliaries to contract on the end of the subject NP, which is not underlyingly in a constituent with them. This is also the way in which I would propose to arrange for the necessary constituent structure of "John gave the money to Susan's uncle," under the circumstances described above. The structure resulting from "cutting loose" the redundant elements John, gave, the money and readjoining them leftwards is something like



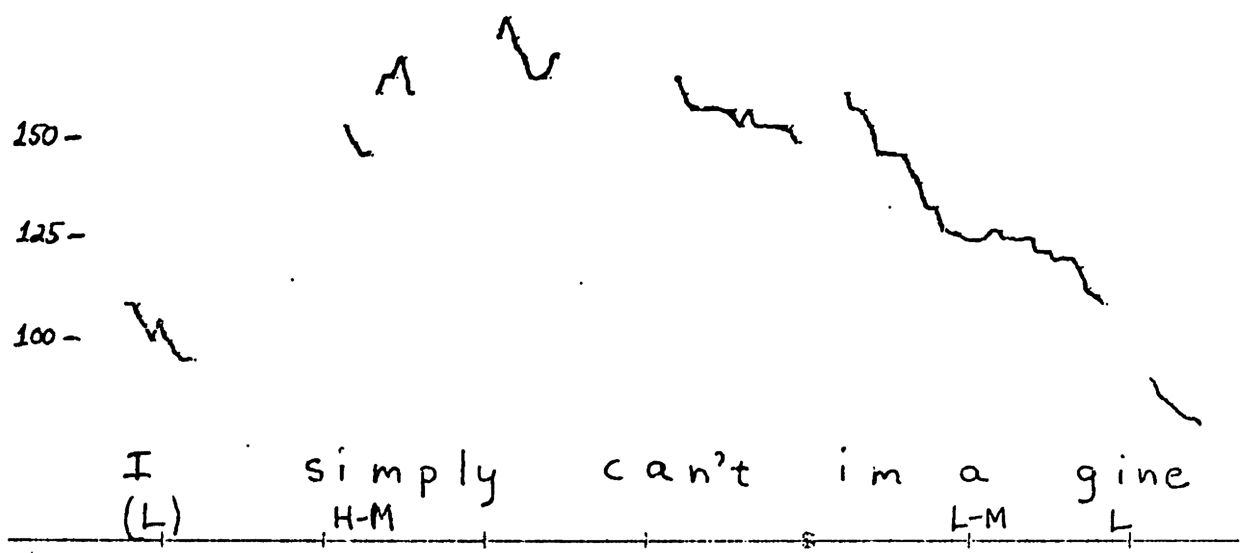
(the highest constituent is sw for rhythmic reasons -- such phenomena are discussed further in 4.3 and in chapter 5).

We will point out in the next section that the description of cases like these immediately involves us in a number of complex theoretical issues of the highest importance, which cannot be discussed apart from problems of syntax and semantic interpretation which are out of place in this work.

So we will in these cases allow ourselves to make descriptively necessary adjustments in metrical structure, to get the tone assignments (and also timing, as discussed in chapter 5) to come out right.

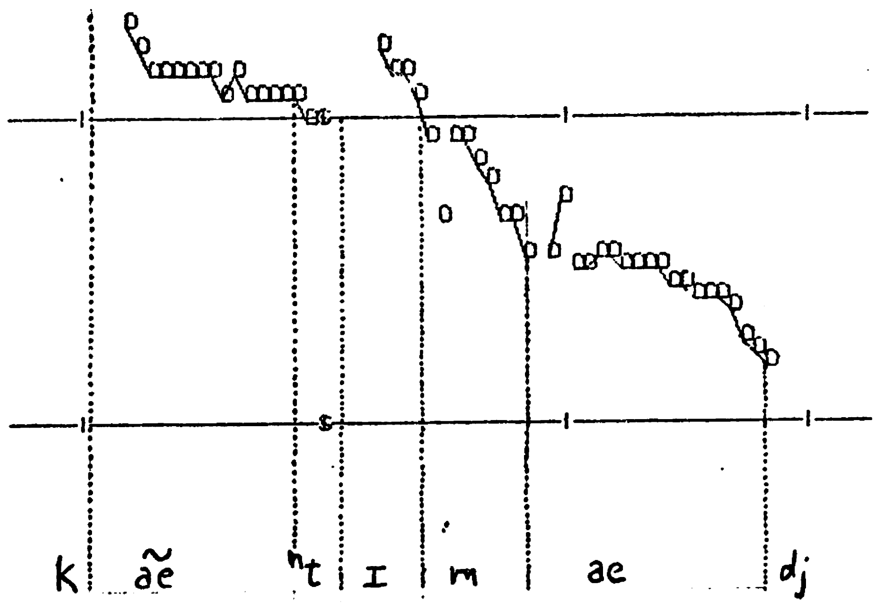
We now return to our discussion of examples of Tone Group 2, the "judicious tone."

3.2.6/20



An interesting feature of the above example is that it has what O'Connor and Arnold would call a "stepping head." They give the "steps" as the underlined syllables in "I <u>simp</u>ly <u>can't</u> imagine." Our theory simply assigns H-M tone to the stressed syllable of <u>simply</u>, and L-M tone to the stressed syllable of <u>imagine</u>. The intervening stretch is presumably subject to the kind of interpolation that we discussed in relation to the various examples of sequences of "free syllables" earlier in the chapter. If it is really true, as O'Connor and Arnold assert, that there is a special character to the treatment of free syllables in this case (namely that they step downwards by increments defined on stressed syllables), then we might have a reason for abandoning our tentative nonlinguistic theory of such interpolations, in favor of a linguistic one. Some aspects of examples of this type suggest that the "stepping head" may really exist -- it is often true that the stressed syllables in such a sequence are relatively level intermediate stages, with most of the fall occur-ing on the unstressed syllables. Thus if we look closely at "can't imagine," we see that <u>can't</u> is relatively level, while the low-stressed first syllable of <u>imagine</u> takes most of the fall to the low-mid pitch on the following syllable:

3.2.6/21



In listening to this example, as well as to cases with more "steps," one feels that there is something to the idea of a "stepping head," although it also seems that a rendition with more of a smooth, "glissando" interpolation is also possible. In order to describe a "stepping head" in our theory, we might want to introduce a class of interpolation rules for the treatment of free syllables (for a different approach to the same problem, see the discussion of complex tonal words in section 3.3). Suppose we have a situation as in the diagram below, where X_1 - X_n are syllables, and T_1 - T_2 are tones:

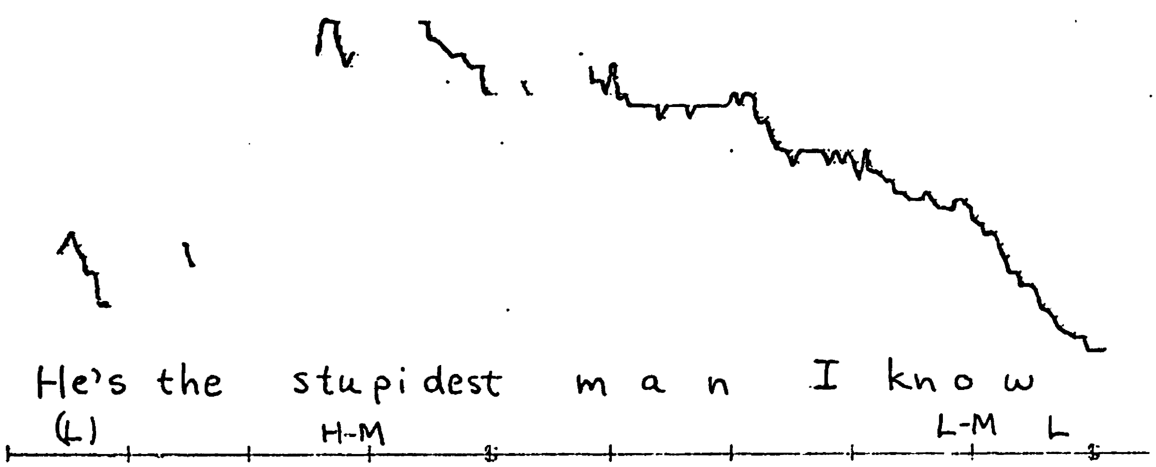
3.2.6/22

$$\begin{array}{ccc} X_1 & \ldots\ldots\ldots\ldots & X_n \\ | & & | \\ T_1 & & T_2 \end{array}$$

Then our interpolation rule would have to create a set of subsidiary tonal entities, T', T" etc., which would in some way partition the interval created by the particular mode of realization of T_1 and T_2. These subsidiary tonal entities would have to be assigned to (some of?) the stressed syllables in the stretch $X_1 - X_n$. We will carry this suggestion no further, since considerable research would be required to determine whether in fact such interpolation rules exist, whether this is the correct way to look at what they do, and exactly how it is that they do it, if do it they indeed do.

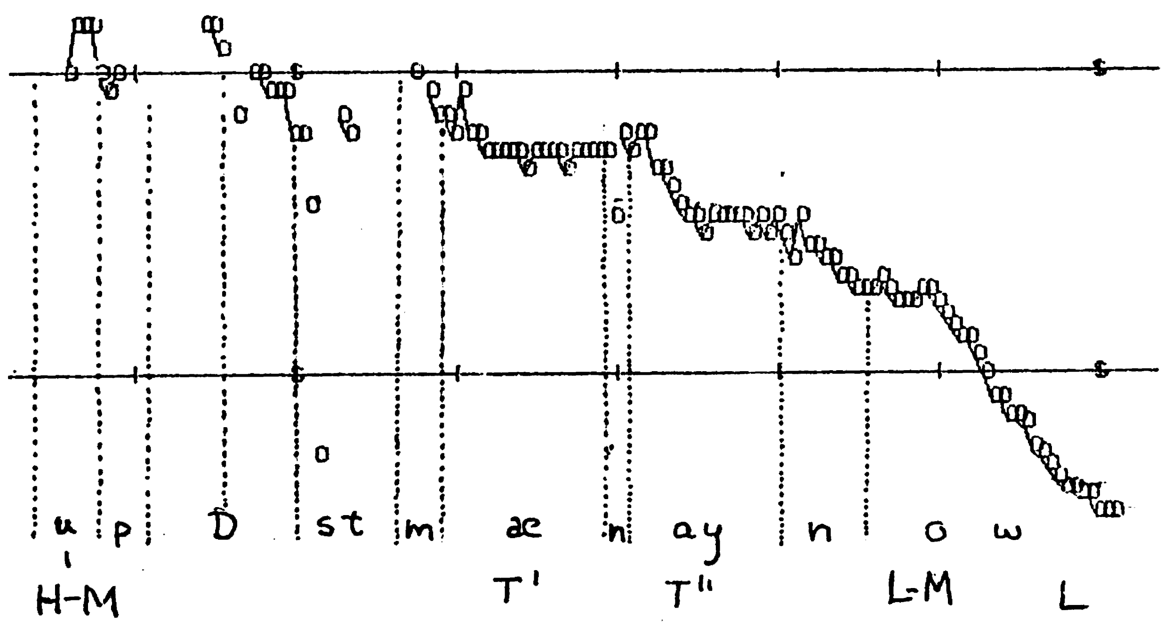
We will, however, cite one further example of an example notated by O'Connor and Arnold with a "stepping head," in which there do indeed seem to be steps, and in which the steps do seem to correlate with the syllabic nuclei of stressed syllables:

3.2.6/23



Looking at a more detailed version of the stretch from the point of attachment of the H-M tone to the point of attachment of the L-M tone, the existence of "steps" seems fairly clear:

3.2.6/24



We have now gone about as far as we intend to go in our examination of simple "tonal words" in English. We have not completely exhausted the family of falling tunes, and haven't even begun to consider the family of rising tunes, but no new theoretical points would arise in carrying such a survey out to the end. Our purpose has been served; we now are in a position to answer at least some of the questions raised in chapter 1.

We have proposed an underlying form for tunes. So far we have only examined simple tunes -- concatenation and compounding of various sorts commonly occur, a phenomenon on which we will rest our gaze briefly in the next section, but the general form of our theory expands reasonably well to cover these cases, at least to the extent that I understand them.

We have proposed a method for associating tunes and texts. This method has two advantages -- first, it works, correctly predicting the tonal effect of stress pattern changes, etc., and second, it is extremely general, consisting (primarily) not of a list of idiosyncratic rules. but rather of a series of appeals to the propensity of human beings to impose certain sorts of structure on sequences of events in time. We have had to make certain basic choices in deciding how to represent this propensity formally, but if we are correct, then it may be that evolution has made the same choices. The value of our proposal should therefore be determined not so much on its ability to give the right output -- there are undoubtedly many other ways in which this could be done -- but on the strength of its claim that the particular patterns observed in the output are not arbitrary.

We have remarked that the intonational lexicon seems to share many properties with the better-understood category of ideophonic words. We proposed that these similarities stem from the fact that both are semi-conventionalized systems in which the basic mode of meaning is iconic. We have implied that much of the iconic value of "tunes" stems from some kind of relationship to the human gestural system. There is considerable precedent for such a view -- to give one example, Harris writes that "in incidence and meaning, these features border closely on gesture." It is certainly true that it is extremely difficult to "perform" intonational contours without appropriate facial expressions, hand motions, etc. (although it has been suggested to me that the reason for the great wealth of intonational expression in English is that the Ur-Englishman, forbidden by social convention to wave his hands, chose to wave his larynx instead).

In addition to the underlying, metrically generated tune-text associations, we have suggested that there might be additional processes involved in interpolating tones for certain (stressed) free syllables. There are probably further considerations involved in deriving the phonetic representation (the lowest systematically significant level of representation) of tonal patterns in English; for example, the specification of changes in tessitura.

Our next task is to make a few remarks about the nature of complex tonal patterns. Once this is accomplished, we will proceed to discuss two very important issues: the nature of the metrical patterns associated with linguistic entities in English, and their relationship to the more conventional notions of stress patterns and of intonational phrasing; and the nature of the metrical grid, and its relationship to the problem of the rhythms (in the ordinary language sense) of speech.

3.3 Complex Tonal Patterns

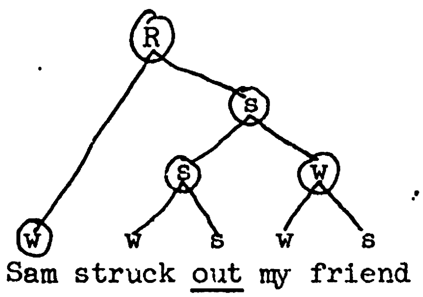
This section will be much too short to do justice to its subject. One reason for this is the unfortunate pressure of time; another reason is that the subject is not very important to our present inquiry. The subject is crucially important, of course, for an understanding of the interrelationship of intonation, syntax, and semantics in any reasonably complex case; but our attention is focused on somewhat narrower issues.

It is expositionally impossible to separate the description of complex tunes from considerations of the form and meaning of sentences. Our theory of congruence makes a claim about textual metrical patterns, given an observed tonal association pattern; by implication, a claim may be made about the syntactic structure of the text, to the extent that this is believed to determine the metrical pattern. This implication arises most strikingly in the case of sequences of tonal words separated by boundary tones, which by the nature of our theory implies the independent presence of corresponding boundaries in the text. In some cases, those boundaries fall where it has been traditional to assume the existence of important syntactic boundaries, particularly NP and S boundaries. An example of this case is the boundary separating sentences from certain kinds of post-sentential tags, an instance of which was mentioned in chapter 1 (the example "Sam struck out, my friend.")

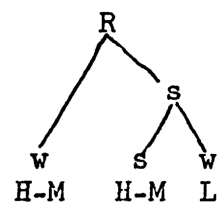
In reference to this particular example, it should be pointed out that the existence of a boundary in the position represented by the comma, will automatically produce the facts about tune-text association in this example, as opposed to the example "Sam struck out my friend," which were pointed out in chapter 1.

First let us derive the case without the boundary, in which out receives the main stress of the phrase:

3.3/1a



1b



1c
```
Sam struck out my friend.
 |          \       |
H-M        H-M      L
```

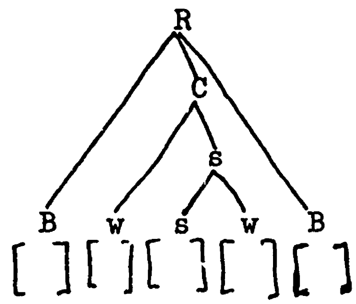
This is exactly the association pattern observed in 1.2.4/4.

Before we derive the version with the "comma" between out and my, we need to say a few words about what the necessary tonal structure will look like. We observed earlier that the maximal structure of a simple tonal word in isolation was:

3.3/2
```
 B   w   s   w   B
[ ][ ][ ][ ][ ]
```

where the B's represent boundary tones. Since "boundary" and "content" are not relational notions, but refer to the intrinsic nature of the elements they dominate, there is no reason not to allow triple expansion of the root in representing such cases:

3.3/3



Indeed, this corresponds rather well to our intuitive concept of what "boundaries" are like, structurally.

When a "tag" is added, given an initial structure like the above, we get a terminal string like this:

3.3/4 [B][w][s][w][B][?][B]

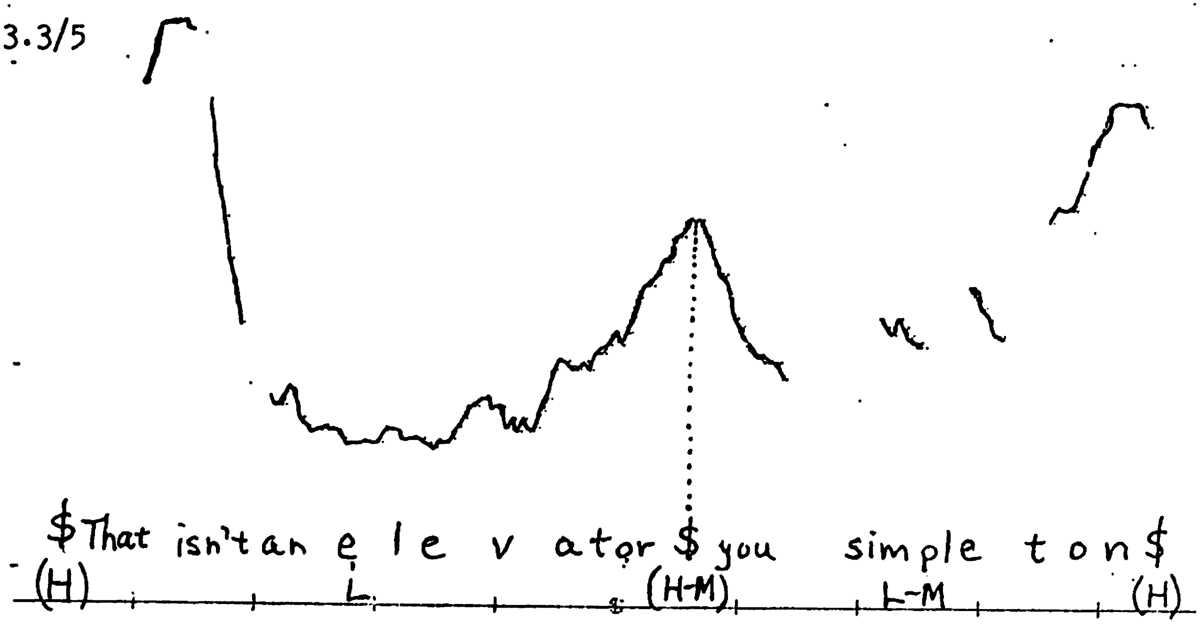
The "body" of the tag generally has only a single tone, which must be the unmarked low-mid tone, or some "allophone" of it; e.g. in the environment

$$\overbrace{\begin{array}{ccc} B & ? & B \\ | & & | \\ \text{high} & \text{low-mid} & \text{high} \end{array}}^{\text{Tag}}$$

the pitch of the tag will not plummet to an acoustic "low-mid" area, but simply sags a little; which is reasonable enough if the "low-mid" tone is actually a kind of null articulation.

In any case, the tag does not get a separate tonal word of the normal type; on the other hand, there is an intonational boundary, often with a noticeable boundary tone, intervening between it and the main clause:

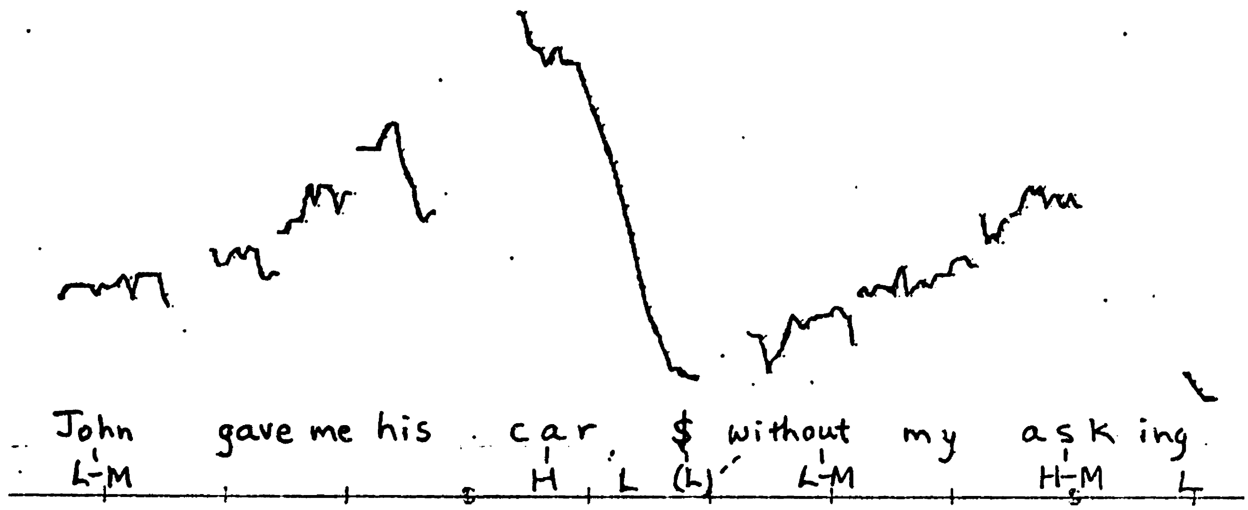
3.3/5



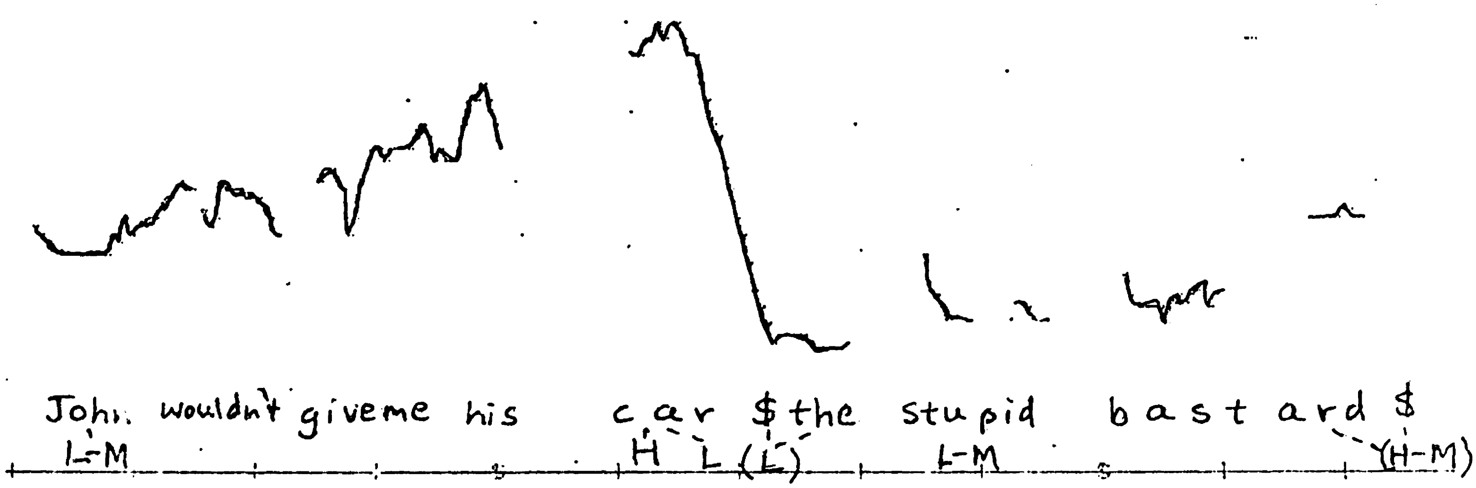
Thus the tonal counterpart of the tag is a sort of tonal affix, appended to the tonal word in a way quite analagous to the attachment of the tag to the main clause. Since our theory requires structural congruence of tune and text, this is not especially surprising. However, any general specification of the theory of such congruences immediately raises innumerable questions about the nature of the representation at issue.

We will give just one example. In the following pair of examples, the first consists tonally of a pair of complete "words," while the second has the same <u>word</u> + <u>affix</u> structure as the previous example. This is not a matter of optionality -- added-on comments, like the material following the comma in 6a, do not take the <u>word</u> + <u>affix</u> tonal structure; and epithet-tags, like the material following the comma in 6b, do not allow the <u>word</u> + <u>word</u> structure (without changing their nature and becoming added-on comments).

3.3/6a



6b



Nevertheless, the addenda in the two examples have the same metrical structure, as near as makes no difference; so some non-metrical source must be found for their differing intonational behavior. We could deal with this situation in two ways. Suppose we classify and label syntactic structures, according to principles like "takes word + affix;

takes word + word; etc." If this classification turns out to reflect real classes of arguably different syntactic structures, then these syntactic structures should be taken to define the basic forms of intonational phrasing, with some reflection of them being defined on tonal representations.

Alternatively, we could conclude that the difference between 6a and 6b above does not correspond to any difference in syntactic structure. Rather, two different tonal forms have been mapped onto the same syntactic structure; the effect of the word + affix form is to cause the material following the intonational break to be interpreted as a "tag" (whatever exactly that is), while the effect of the word + word form is to cause the same material to be interpreted as an added-on comment.

We will assume that under both hypotheses, the boundaries corresponding to the "commas" are represented in some form in both strings -- this issue is independent of whether the difference in tonal form corresponds to a difference in textual syntax.

Both of these alternative hypotheses have extremely interesting consequences. If the first hypothesis is true even in part, then we are provided with an invaluable probe into the nature of at least certain types of syntactic structures. If the second hypothesis is true, even in part, then the effect is to define a second syntax on the terminal string of any sentence which has an intonation. Since the elements of the tonal representation have meanings, we know already that (under any hypothesis) a kind of "second semantics" is defined on the string. The range of tonal meanings covers only conditions on appropriate usage, to our knowledge, so that a better term is probably "pragmatics." In any case, this overlay of meaning, defined by the "words" of the tune, undoubtedly depends for its effect on the particular association of tonal words with textual constituents, as anyone who has ever tried out a few tonal variants on a complex sentence knows. This has to be true under any theory; properly understood, it can simplify the description of textual syntax and semantics considerably.

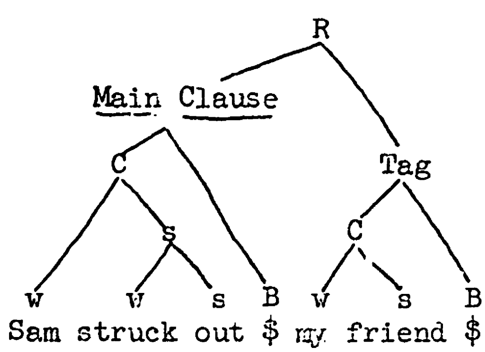
But the consequences of defining an independent tonal interpretation of utterances, on this rather minimal basis of projecting some pragmatic vector from a tonal word onto the textual constituent to which it corresponds, are minor compared to the possible consequences of defining an independent tonal syntax on the textual string. If the tonal system has an independent specification of form, this form presumably has an interpretation. Such a supposition would be necessary to account for the difference between 6a and 6b on the basis that it is only the tonal form, and not the textual form, which is varied; and this was an extremely simple example of what such forms would have to be like.

The most likely candidate for the interpretation of this putative tonal form, would be a structuring of the content of the text according to the role of the utterance in the discourse. This kind of "information structure" clearly exists; if its specification could be systematically separated from the aspects of semantic interpretation which deal with truth and reference (the traditional domain of semantics), then many complexities of the determination of "meaning" in ordinary language, could be factored into the interaction of a simplified system of textual interpretation, with the system interpreting tonal structure. The potentialities in this regard of an actual <u>intonational syntax</u> (as opposed to simple concatenations of the pragmatic value of individual tonal words) are very great.

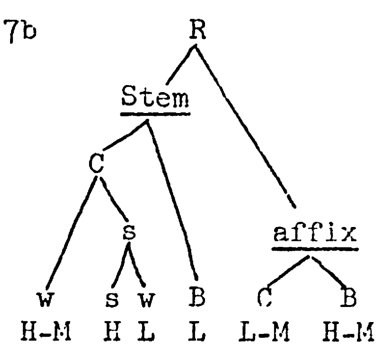
Since either of the two possible hypotheses in this area (the dependence of tonal form on syntactic structure, or its independent status) has such interesting consequences, we would like to pursue the matter further. However, the domain of data involved is simply too vast for us to attempt any further investigation, in the context of an inquiry to which these issues are tangential.

Returning to the issue of the effect of the comma in "Sam struck out, my friend" on tune-text association, we can safely ignore the issues just raised -- whatever else is true, we maintain the assumption that a boundary of some sort exists in both tune and text. Thus the intonational derivation goes something like this:

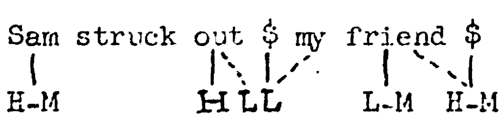
3.3/7a



7b



7c



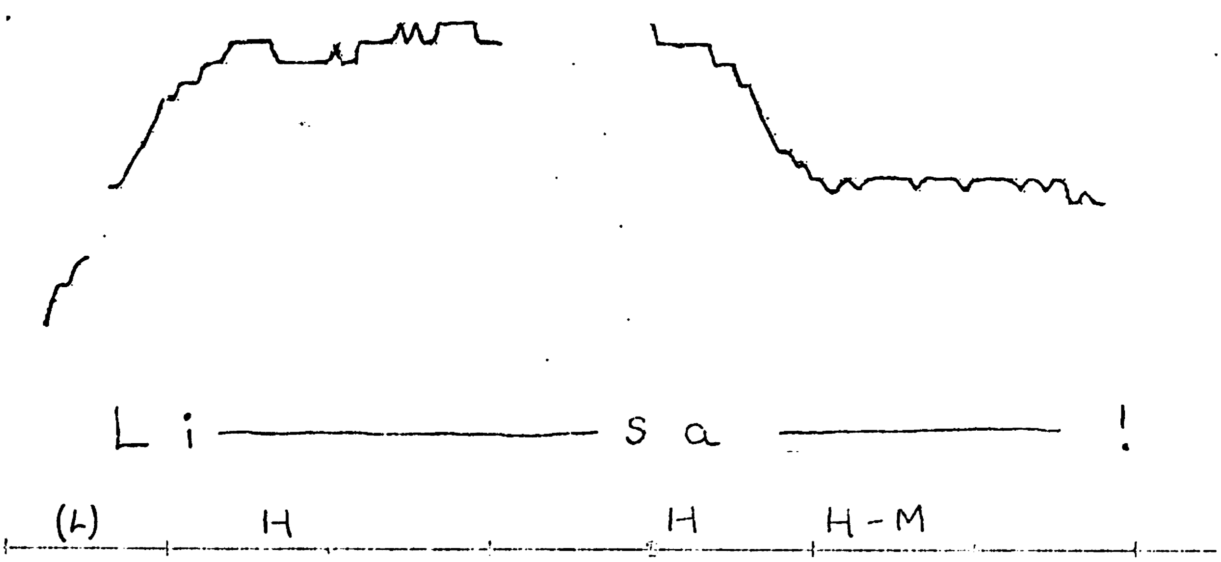
The fact of the existence of the boundary will "trap" the Low tone on out, just as we observed empirically in reference to examples 1.2.4/3b and 1.2.4/5. The rest of the derivation is just as we have done in the past, assuming some solution to the "tonal syntax question," except that we have allowed the medial boundary tone to associate forward onto the free syllable my, rather than backward onto the already overcrowded out. This rightwards association of medial boundary tones is quite common even when the syllable to the left is free.

There is a great deal more of a purely descriptive nature, to be said about complex intonational words (which may be expanded internally by reduplication of various sorts, and externally by the addition of "suffixes" and "prefixes" like the tag-intonation already discussed). An example which might suggest a reduplicative analysis is the following form of the vocative tune /L H H H-M/

3.3/8a

```
         Lisa !!!
     (L) H H H-M
```

3.3/8b



There are two different ways in which we could represent such elaboration in our theory: they might be termed the morphological method and the phonological method. Our hypothesis about the interpolation of intermediate tones, to account for "stepping heads," would be an example of the phonological method, according to which the underlying form of the tune is simple, and there exists some procedure, operating after the initial tune-text assignment is completed, which creates the "elaborated" form of the tune by specifying the nature and location of non-underlying tones. We might attempt to explain examples like 8 above as the result of this sort of processing. Alternatively, we could take the position that there exist morphological processes, in the intonational lexicon, which create complex intonation words out of simple ones, for example turning /L H H-M/ into /L H H H-M/, and which operate prior to tune-text association.

We could invoke such a procedure to create "stepping heads," for example, by turning the simple tune /H-M L-M L/ into /H-M H-M L-M L/, proceeding with tune-text association in the standard way, and then attributing the "downstepping" character of the sequence of high-mid tones to some set of rules which determine the phonetic realization of various sorts of tone sequences.

The advantage of the morphological method is that it enables us to take advantage of the existing apparatus of tune-text association in order to give the correct assignment (always to stressed syllables, suspiciously enough) of the "interpolated" tones. The primary difficulty would seem to be that we need to arrive at a number of tones in the underlying representation of the tune which will suit the nature of the text. We will have to face a somewhat similar problem in any case, in dealing with the generation of complex tonal phrases by stringing together tonal words, so that this difficulty should not dissuade us from pursuing the morphological method. The necessity of attributing a non-equivalent phonetic realization to underlyingly equivalent tones (in the case of the "stepping head") should not dissuade us either, since there is a clear precedent for such a treatment in the phenomenon of "downstep" in certain African tone languages.

In these languages, in a sequence of the form H L H, the second high tone will be downstepped, i.e. realized on a pitch lower than that of the first. In addition, there is a form of high tone (generally called "mid") which by itself has the same realization as an ordinary high tone, but which exhibits "downstep" in sequence with itself, so that in the sequence M M M the first tone is realized at the normal high level, while the succeeding tones are realized at successively lower levels.

The preceding highly truncated discussion represents all that will be said in these pages about the important issue of complex tonal structures. The first draft of this work dealt almost exclusively with the domain of complex tonal phenomena, since it had not occurred to me that there was anything interesting happening in the simple cases, where all that was at issue was the seemingly trivial task of determining the point of association of two or three tones. Later, I came to see the details of this trivial task as a microcosm of the intonational system as a whole; on my present view, the regularities on which I based the earlier (broader and less abstract) system, are an artefactual result of factors quite different from those which I believed to be at work. In other words, what was previously a theory has been downgraded to the status of a set of observations.

One last point on elaborated tunes: Crystal (in Prosodic Systems and Intonation in English, pp. 217-220) establishes categories of "complex" and "compound" nuclear tones. Judging from his descriptions, these cases would all fall with a single tonal word for us (as for him they are elements of a single "tone-unit"); we would describe them as basically resulting from T_2, T_3 and a terminal boundary tone, under various conditions of textual association, with the possible addition, in the case of the compound tones, of a "reduplicated" tone like the additional high tone in 3.3/8. So for us they represent a largely epiphenomenal category. Crystal analyzes (most) nuclear tones as kinetic, in about the same way as O'Connor and Arnold do, so that he has primitive categories like "low fall," "low rising to mid" etc.

For this reason, his remarks on the observed properties of the class of complex and compound nuclear tones are extremely interesting. He observes that in both categories "the kinetic tones must display an 'endocentric' relationship." In other words, complex and compound nuclear tones can be fall + rise, rise + fall, rise + fall + rise, etc., but they can never be fall + fall, rise + rise, fall + rise + rise, etc.

Falling and rising kinetic tones must alternate; sequences without this property are "exocentric," and he says that "'Exocentric' sequences of tone-units are interpreted as either separate or subordinate."

Thus a sequence fall + fall, according to his intuitions, is necessarily the result of two different tonal patterns, juxtaposed either as equals, or with one subordinated to the other. Nothing about his descriptive system requires this result; he includes it simply as an empirical criterion which seems to him to help differentiate intuitively unitary tonal sequences from intuitively separate ones.

He is exactly right in this intuition; and it has a simple source. Any sequence of static tones (reinterpreted as the sequence of kinetic pitch-movements they define) is necessarily "endocentric." Thus the sequence high + low could be considered to define a "falling tone"; but if we repeat the sequence, high + low + high + low, we do not define two "falling tones," but rather the pitch-movement fall + rise + fall. In order to define two successive "falling tones," by means of static tonemes, we must resort to the sequence high + low (break) high + low; and the necessary "break" in the pattern will only occur between basically distinct tone-units.

Thus this empirical observation by an author who is committed to kinetic tonal primitives, strongly supports the idea that the real primitives of the tonal system are static.

4. Metrical Patterns and Stress.

The purpose of our theory of metrical patterns (as rooted, oriented trees with binary branching and "relational" node labels s and w) was to represent the stress patterns and "phrasing" of texts in a way which would allow us to accomplish tune-text association in a natural and general way. We wanted to avoid complex lists of *ad hoc* rules for tone assignment, out of a belief that the process of associating tunes and texts is natural (i.e, does not have to be learned, other than perhaps in a few points of detail) and general (i.e. is not entirely restricted to language and speech, but related to parallel problems of associating independent serially ordered structures in music, dance, etc.). Finding any system at all that would work was also a strong motivation, but we have preferred to look for simplicity at the root of the problem, rather than collecting complexities at its periphery, because of this belief that a natural and general system does exist, not just some more or less elaborate agglomeration of diverse and arbitrary machinery.

If our system is indeed natural, then we would expect its constructs to have validity apart from their role in the specific process which we set out to understand, the process of tune-text association. For example, we would expect that the metrical patterns which we argued to play a central role in tune-text association, would have some independent status as well. In particular, it is clear that they have an intimate relationship to the phenomena which we discussed in chapter 1 under the commonly accepted name of stress patterns.

In this chapter, we will undertake a preliminary exploration of the usefulness of the idea of metrical patterns in representing stress and patterns of stress in English. Especially in the domain of word stress, this exploration will be rather tentative, but we will try to construct a series of arguments, some theoretical and some empirical, that metrical patterns of the sort we have been employing are a promising way to represent these phenomena. These arguments will be of three types, concerning the cyclic nature of prosodic rules, the representation of certain types of stress-shifting rules, and the representation of the interaction of structural and nonstructural factors in determining stress patterns.

4.1 The "Cyclicity" of Stress Rules.

There are two interlocking arguments in favor of the "cyclicity" of stress rules in the phrasal domain: first, the structure-dependence of relative stress, and second, the preservation of relative stress under embedding. Chomsky and Halle put it like this: "it is natural to suppose that in general the phonetic shape of a complex unit (a phrase) will be determined by the inherent properties of its parts and the manner in which these parts are combined, and that similar rules will apply to units of different levels of complexity[1]." For example, this natural supposition would predict that the general tendency to give a head noun higher stress than its modifying adjective will hold true regardless of whether the units in question are simple or complex, and that if they are complex, their internal stress patterns will be determined by principles similar to those that apply in the larger domain.

In order to express this natural (and undoubtedly correct) supposition formally, Chomsky and Halle had recourse to a system with the following properties:

4.1/1a Stress assignment rules dependent on structure of various sorts.

1b The principle of cyclic rule application.

1c The principle of stress subordination.

1d Stress patterns represented as sequences of numbered stress levels, each a feature of some segment, generated by 1a, 1b, 1c and certain additional minor rules.

In a theory in which stress patterns are represented as metrical patterns of the sort we have been discussing, the same effect can be achieved with these properties:

4.1/2a Stress assignment rules dependent on structure of various sorts.

2b Stress patterns represented as hierarchical organizations of (relatively) strong and weak positions.

2a represents simply the idea that structure plays a role in stress assignment, and is common to both formalisms. 2b represents simply the definition of a metrical pattern. The metrical theory does not need: the principle of the cycle; the principle of stress subordination; any nonbinary features.

To see why this is so, consider the "nuclear stress rule," given below in conventional and metrical versions:

4.1/3a System 1: a) word stress rules: assumed.

b) Nuclear Stress Rule: $\begin{bmatrix} \text{V} \\ \text{1 stress} \end{bmatrix} \rightarrow \text{1 stress} / \_\_ \text{Q}]$

(applies only to phrasal constituents)
condition: Q contains no 1 stress.

c) Principle of cyclic application of NSR.

d) Principle of stress subordination: "when primary stress is placed in a certain position, then all other stresses in the string under consideration at that point are automatically weakened by one."

3b System 2: a) word stress rules: assumed.

b) Nuclear Stress Rule (applies only to phrasal nodes): In every phrasal constituent, put the strong position on the right.
formalization: $\text{N} \rightarrow \text{S} / [\text{M} \_\_]$

Applying these rules to a simple case, the NP "madison avenue address," we get the following derivation from system 1:

4.1/4 [[madison avenue] address]

madison	avenue	address	
1	1	1	word stress
2	1		NSR on inner cycle, plus stress subordination.
3	2	1	NSR on outer cycle, plus stress subordination.

If we lacked the principle of cyclic application, we would get derivations like the following:

4.1/5 [[madison avenue] address]

madison	avenue	address	
1	1	1	word stress
2	2	1	NSR on outer constituent, plus stress subordination.

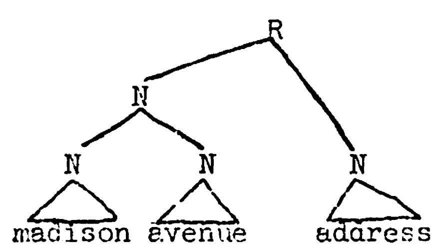
No further applications of the stress rule are possible, since the rule, crucially, can only reassign 1 stress to a 1 stress vowel. In the simple phrase "madison avenue address" this is not, perhaps, so terrible, but it would be unfortunate to generate a stress pattern

2 2 2 2 1

like "rather fancy madison avenue address," which would be permitted by a noncyclic theory. Furthermore, a noncyclic theory (on these assumptions) would have a hard time stating the principle of stress subordination, since without the notion of "cyclic domain at the point of application of the rule," the whole idea of stress subordination doesn't make a great deal of sense. There is no obvious way, for example, to prevent stress assignment on a cyclically lower domain from subordinating stresses at a cyclically higher level (and thus giving the wrong main stress to the output), except by having a cyclic principle of rule application.

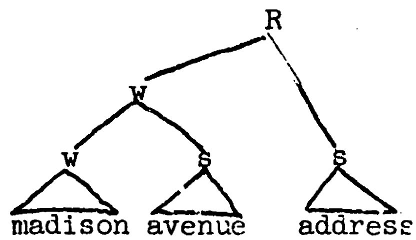
A metrical theory approaches the same problem in a different way. We assume that the syntactic constituent structure gives us a tree, as below, with the metrically neutral node label N:

4.1/6



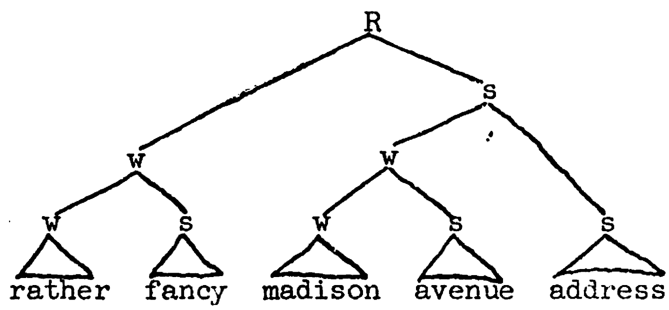
The metrical version of the NSR simply tells us that every constituent is of the form [w s], which implies the node labels:

4.1/7



The same trivial principle will give us:

4.1/8



It doesn't matter how we apply the "metrical NSR" -- it can be bottom to top, top to bottom, from the middle out, randomly, or all at once -- the same result is achieved in all cases.

The metrical theory also lacks any notion of stress subordination, and indeed makes do with the single binary opposition strong/weak. Furthermore, the formal statement of the rule need not involve any variable. One possible formalism for such a rule would write it as follows:

4.1/9 $\quad N \rightarrow S / \; [M \; \_\_]$

where N and M are simply symbols for nodes in the syntactically given constituent structure. The nature of the metrical node labels s and w is essentially relational, as we have emphasized from the beginning; there cannot be a node <u>s</u> except in association with the complementary node <u>w</u>, and vice versa. This implies that the real meaning of assigning the node label <u>s</u> to the second element in a metrical constituent, is the assignment of the <u>pattern</u> ws to the constituent as a whole.

The fact that the metrical theory makes do with a binary (rather than n-ary) opposition, that no special principle of stress subordination is required, and that the metrical NSR requires no variable in its statement, are all interesting points that argue in its favor. However, the most interesting issue here has to do with the phonological cycle. For the conventional theory, the phonological cycle is an optional additional assumption, not logically related to the nature of the particular rule involved. Thus it is as reasonable to postulate it for segmental rules as it is for prosodic rules.

The metrical theory lacks any apparatus for the cyclic application of rules -- <u>it accounts for the "cyclic" properties of prosodic phenomena on the basis that stress is a hierarchically defined relation</u>; that is, on the basis of the inherent nature of the phenomenon itself. Segmental rules do not have this character -- rather than defining hierarchical relations among phonological entities, they define intrinsic properties of the phonological entities themselves. Therefore, the metrical theory predicts that prosodic rules will necessarily, by their very nature, have a "cyclic" character (although actual cyclic application of rules is not necessary), but it makes no prediction at all for segmental rules. In view of this fact, I think it is quite interesting that the arguments for the segmental cycle are so few, recherché, and controversial, while the evidence for the "stress cycle" can be recreated in a convincing way for important and productive processes in almost any language one chooses to look at.

An important argument for the plausibility of a segmental cycle has always been that since stress rules provided such a convincing case for the cyclic principle, the most reasonable and parsimonious approach was to assume that all rules apply cyclically unless proven otherwise. If it is true, as our treatment of the NSR suggests, that there is no principle of cyclic rule application for stress rules, their apparently "cyclic" properties (e.g. preservation of relative levels of stress under embedding) being instead an ineluctable result of the nature of the phenomenon itself, then the burden of proof shifts rather dramatically onto those who wish to demonstrate the existence of cyclic rule ordering for segmental phonology, or indeed anywhere in phonology at all.

Of course, we have not shown that cyclic ordering of stress rules is unnecessary, merely that it is unnecessary in the case of the NSR. So let us shift our sights down a notch and consider the Compound Stress Rule (CSR):

4.1/10a System 1 (conventional):

a) word stress rules: assumed.

b) CSR: $\begin{bmatrix} \text{V} \\ \text{1 stress} \end{bmatrix} \rightarrow \text{1 stress} / \_\_ \text{Q}(\#\#\text{P})]$

(applies only to lexical constituents)

conditions: Q contains no 1 stress

P contains no ##.

4.1/10a c) principle of cyclic rule application.
d) principle of stress subordination.

10b System 2 (Metrical):

a) word stress rules: assumed.

b) CSR: $N \rightarrow s/\ [\_\_ M]$
(applies only in lexical constituents).
condition: M is a single word.

c) NSR as before.

Some derivations:

4.1/11a [[community center] [finance committee]]

1	1	1	1	word stress
1	2	1	2	CSR (plus stress subordination) on inner cycles.
2	3	1	3	CSR on outer cycle.

11b [[[law degree] requirement] changes]

1	1	1	1	word stress
1	2			CSR on inmost cycle.
1	3	2		CSR on medial cycle.
1	4	3	2	CSR on outer cycle.

The reason that the main stress in these four-word examples is "community center FINANCE committee," but "LAW degree requirement changes," is simple: at the beginning of the last cycle, we have

4.1/12a $\overset{1}{\text{community}}\ \overset{2}{\text{center}}\ \underbrace{\overset{1}{\text{finance}}}_{Q}\ \underbrace{\overset{2}{\text{committee}}}_{P}$

4.1/12b $\overset{1}{\text{law}}\ \underbrace{\overset{3}{\text{degree}}\ \overset{2}{\text{requirement}}}_{Q}\ \underbrace{\overset{1}{\text{changes}}}_{P}$

The variable P will take exactly one word on the right, if it can -- it can in both cases, and so does. The variable Q can contain no 1 stress, so that it can cover only the syllable -nance in the first case. In the second case (because of the different derivational history, caused by the different constituent structure) it can cover the stretch degree requirement, and so (being interpreted maximally) it must. Thus the locus of application of the CSR will be as marked by the arrows, in the two cases.

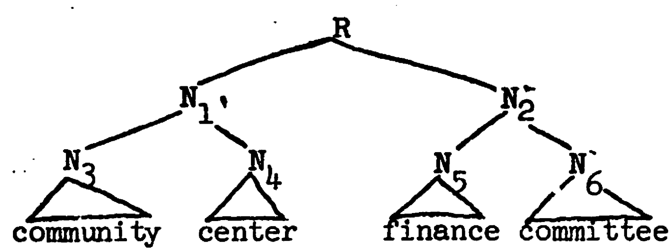
This ingenious solution obviously depends crucially on the principle of the cycle. For example, if we had started with the largest constituent in the second case, we would get:

4.1/13a $\overset{1}{\text{law}}\ \overset{1}{\text{degree}}\ \overset{1}{\text{requir}}\underbrace{\text{ement}}_{Q}\ \underbrace{\overset{1}{\text{changes}}}_{P}$

13b yields: $\overset{2}{\text{law}}\ \overset{2}{\text{degree}}\ \overset{1}{\text{requirement}}\ \overset{2}{\text{changes}}.$

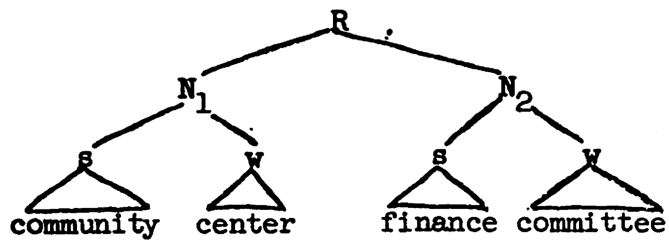
Now let's derive the same examples in the metrical theory.

4.1/14



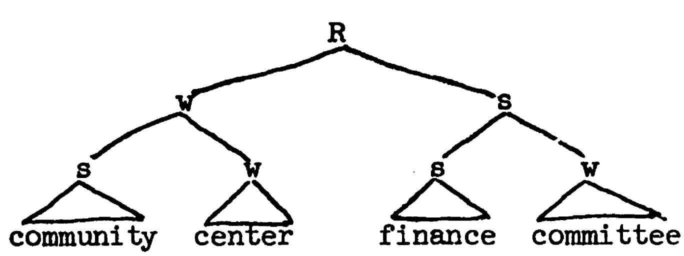
If we check over the tree in 4.1/14 for all possible points of application of the metrical CSR, $N \rightarrow s/[\_\_M]$, we find that N_3 and N_5 meet the conditions of the rule, but none of the other nodes do; N_6 and N_2 fail because nothing follows them, and N_1 will fail because what follows it is not a single word. So applying the rule simultaneously to all possible places (with the by now familiar understanding that making a node strong, by definition, makes its sister weak), we obtain:

4.1/15



In order to metrically specify N_1 and N_2 , we could have utilized disjunctive ordering (the notational convention of parentheses) in the way that the other theory did -- however, to avoid the complex and somewhat separable issue of defining disjunctive ordering for simultaneous rules (for which subtle and interesting proposals have been made by Vergnaud, Halle and Prince), we prefer to allow the NSR to step in and give us:

4.1/16

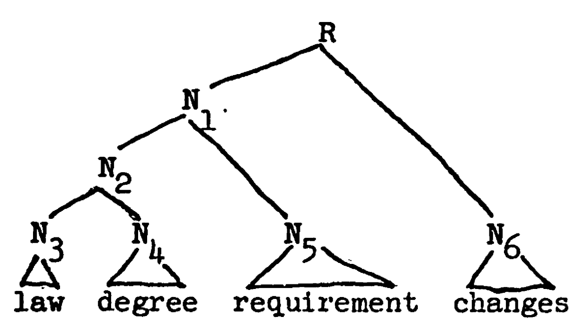


The interesting feature of this derivation is that it involved no principle of cyclic rule application. It depends for its success on the maintenance of the constituent structure of the text throughout the derivation, but since the whole idea of metrical patterns is that they represent a relation strong/weak hierarchically defined, i.e. defined on a constituent structure, this is an inherently necessary feature of the process.

Also, the idea of letting the NSR work to give the main stress of "community center finance committee," rather than a disjunctively ordered "second half" of the CSR (which happens to be identical) is not without merit. This way of doing things reduces the (structure-dependent) rules of English stress, above the word level, to two simple rules, without any variables, parenthesis, angled brackets or curly braces. The CSR says: "put the stress on the left in any lexical constituent whose right hand member is a single word;" the NSR says "put the stress on the right." However, later in this section we will discuss a possible collapsing of the CSR with the main node-labelling principle for word stress, in terms of which the higher-level assignment in "community center finance committee" would be accomplished by one case of the combined rule.

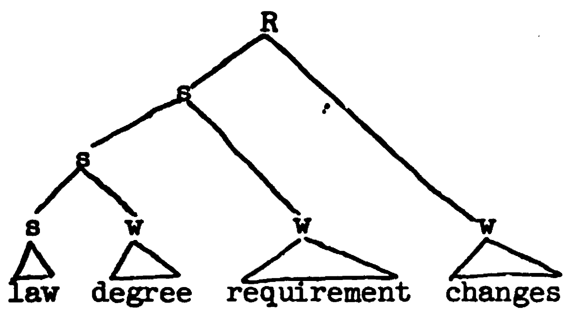
Applying our principles to "law degree requirement changes," we get:

4.1/17



The statement of the CSR (which, in our present formulation, must be ordered before the NSR, a fact perhaps predictable from its restriction to lexical categories) tells us to look for nodes in the environment $\left[\_\_ M\right]$, where M is a single word -- in 4.1/17, nodes N_1, N_2 and N_3 all meet this condition, while none of the others do. Making these three nodes simultaneously strong yields:

4.1/18



The metrical pattern is now completely defined, so that no issue of the application of the NSR arises.

The crucial difference between the metrical theory and the conventional theory is that the metrical theory deals with the specification of nodes, while the conventional theory deals with the specification of segments. This difference is what allows the metrical theory to do without variables, cyclic rule application, and stress subordination.

Since stress rules above the word level, in the conventional system, are unique in assigning n-ary features, which do not represent any intrinsic property of the segment to which they are assigned, but merely its relationship to the stress levels of other segments, it is in many ways a step forward to treat stress patterns as represented by features of nonterminal nodes, rather than as features of phonological segments.

The fact that the metrical theory provides a statement of the NSR and CSR which is somewhat more straightforward than that of the conventional theory is interesting, and is a point in its favor, I think. The most important point, however, seems to be the issue of the phonological cycle and the typology of rules. Our discussion so far suggests the hypothesis that there is no phonological cycle -- stress rules, by their nature, involve "annotations" of constituent structure, and therefore their output depends on that structure in a way which has been (incorrectly) interpreted as reflecting cyclic operation. Segmental rules involve annotations of sequences of feature matrices, and therefore do not have these "cyclic" properties.

Before pushing this hypothesis very strongly, one would like to take a look at word stress -- we will not attempt here the gigantic project of reformulating the entire system of English word-stress. However, we will try to address a few remarks to the question of whether any arguments from word stress and its interaction with other rules can be concocted in favor of the existence of the phonological cycle, given a metrical theory of stress assignment.

To begin with, it is necessary to point out that the situation with respect to word stress differs in three important ways from the system of phrasal stress:

1) the rules for stress assignment are subject to a certain number of arbitrary exceptions.

2) a constituent structure suitable for metrical annotation is not always defined, in the absence of additional principles.
3) word stress patterns are not entirely relational -- in addition to the distinction between stronger and weaker levels of stress, there is the distinction between stressed and unstressed, which has effects intrinsic to the segments on which it is defined (e.g. vowel reduction, etc.). The metrical relation strong/weak is always what might be called a syntagmatic opposition -- when one node is of the first type, its sister is invariably and necessarily of the other type. The [$\pm$ stress] distinction on the words level is not entirely of this character -- for example, disyllabic words in which both syllables are [+stress] are possible, although metrically one must be strong and the other weak.

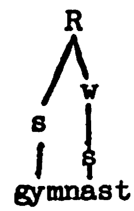
Point #1 above is just one of many examples of how the Lexicon infects with arbitrariness everything it touches. Playing the role of the village idiot savant, it mimics the operations of syntax and semantics in its own peculiar way, which gives rise to complex and semi-productive systems -- it is hardly surprising that its forays into the realm of prosody are likewise a complex mixture of productive and arbitrary principles.

Point #2 means that any metrical theory of word-stress will have to specify a constituent structure, or else abandon some of its assumptions. We will make a stab in the direction of the first alternative.

Point #3 suggests that in addition to the relation strong/weak, which is defined on nodes (nonterminal elements in the metrical pattern) there must also be a feature [± stress] which is defined on segments, and which is related in some way to the metrical distinction, but is not identical to it. Consider the pair of words $\overset{1\ 0}{\underline{\text{modest}}}$ and $\overset{1\ \ 3}{\underline{\text{gymnast}}}$. Both must be metrically of the form [s w]. The only strictly metrical possibility open to us would be to represent them as follows:

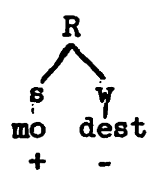
4.1/19a

19b

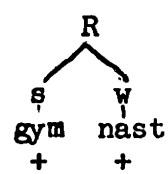


According to the view of metrical patterns we have been developing, 19b is incoherent and meaningless -- the only opposition that the s over the second syllable participates in is an opposition with what might have been substituted for it. This kind of paradigmatic opposition is typical of segmental features, but is foreign to the nature of the metrical relation strong/weak as we have conceptualized it. Instead, we propose the following representation, in which the +'s and -'s under the words represent the distribution of the segmental feature [± stress] :

4.1/20a



20b



There is obviously some relationship between the distribution of +'s and -'s and the metrical pattern. We will suggest as a working hypothesis the idea that this interrelationship is defined by the following view of stress assignment at the word level:

1) initially, there is an assignment of the segmental feature $[\pm \text{ stress}]$.
2) second, there is a metrical bracketing of the word.
3) third, there is an assignment of metrical node labels, on various principles.
4) fourth, there is an adjustment of the distribution of the segmental feature $[\pm \text{ stress}]$.

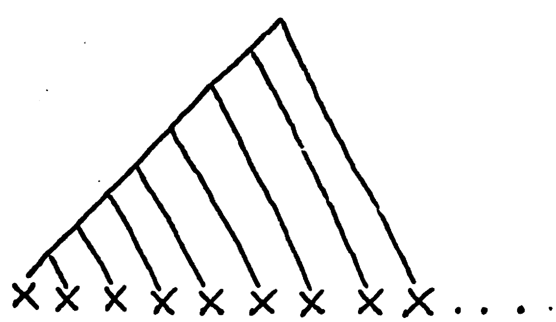
We will now make a proposal intended to flesh out this skeleton with particular rules for 1) - 4). I am not entirely sure that what follows is exactly the right way to do it, and undoubtedly a number of points of detail will have to be changed. The point of the exercise, however, is to investigate the issue of whether any principle of cyclic rule application has to be invoked, and our preliminary conclusion will be that it does not.

The stage is set by the initial assignment of $[\pm \text{ stress}]$ to the string of segments. One obvious principle is that tense vowels and vowels in front of strong clusters are +, while everything else is -; we will suggest certain additional principles for initial $[\pm \text{ stress}]$ assignment shortly. The central portion of the task is to impose a metrical bracketing on the string, and to label that bracketing. The set of rules by which this is accomplished should be seen as a well-formedness condition on such bracketings and labellings -- we will give them as a list of separate rules, and for convenience in exposition we will state them in such a way that they can be applied algorithmically to generate the correct results, but their true nature, as I understand them, is not that of a process but rather that of a filter. What this means theoretically is that issues of ordering, etc. are not allowed to arise.

Once a bracketing and labelling consistent with this complex well-formedness condition has been achieved, a set of further rules will adjust the distribution of the segmental feature [± stress], whose primary role is in determining vowel-reduction.

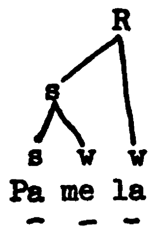
We will divide the well-formedness condition on metrical structure into two parts, bracketing and labelling. The basic form of bracketing in English words is left-to-right -- in other words, a string of syllables (subject to no other constraints) will always be bracketed in a way which groups the initial syllable with its neighbor to the right, then groups the constituent so formed with its rightward neighbor, and so forth:

4.1/21



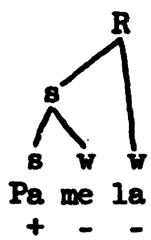
Since the basic labelling principle will assign strong to the left daughter of any constituent, this will produce, e.g.

4.1/22



Since the most important of the $\pm$ stress redistribution rules will say that $\begin{array}{c} s \\ | \\ - \end{array} \longrightarrow \begin{array}{c} s \\ | \\ + \end{array}$, this last example will emerge from the system as:

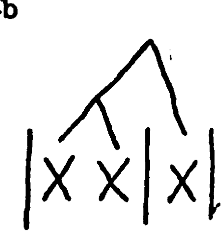
4.1/23



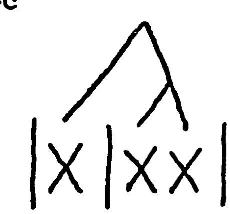
Not all cases, obviously, will be this simple. The first source of complexity is a set of conditions on bracketing. We will describe these conditions as rules inserting a "foot boundary," symbolized $|$, into the string of segments. The meaning of "foot boundaries" is that they divide the string into groups which the metrical bracketing cannot violate. For example, the configuration given in 24a below is not well-formed, but the configurations in 24b and 24c are allowed:

4.1/24a *

24b



24c



There are basically only two "foot-boundary insertion rules" (really conditions on well-formed bracketings):

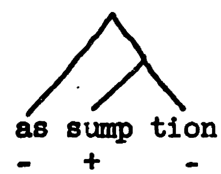
4.1/25a All # boundaries are foot boundaries.

25b A foot boundary is inserted in the environment $\_\_ / [+][-]$

The first of these rules is probably universal. The second seems very common, and is hopefully one of a very small number of possibilities.

An example of the bracketing imposed by 25b:

4.1/26 | as | sump tion | <u>which</u> <u>yields</u>:
        -   +   -



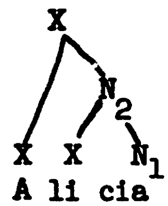
In order to impose the correct node-labelling on the above example, we need to state the node-labelling rule. In its most general form, it is this:

4.1/27 In any lexical metrical constituent $[MN]$, N is strong if and only if it is complex.

The purpose of the restriction to lexical constituents is to differentiate this rule (and the CSR, with which we will attempt to collapse it shortly) from the phrasal labelling rule, the NSR.

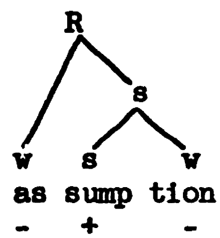
The meaning of <u>complex</u>, fundamentally, is "dominating non-terminal material." For example, in the case below, N_1 is simple (non-complex) while N_2 is complex:

4.1/28



The meaning of the biconditional in Rule 4.1/27 is that in a constituent [MN] , N is strong if it is complex, but weak if it is simple. This rule will impose the following labelling on the bracketing defined by 4.1/26:

4.1/29



The reason for this labelling is simply that the node [-tion] is simple, while the node [-sumption] is complex.

In order to derive a quite wide variety of cases without much further comment, we need to add to the initial distribution of [$\pm$ stress] the following rule:

4.1/30 Case 1: $[\text{-stress}] \rightarrow [\text{+stress}] \;/\; \_\_ \; [\text{-stress}] \; [\text{-stress}] \; \#$

Case 2: $[\text{-stress}] \rightarrow [\text{+stress}] \;/\; \_\_ \; [\text{-stress}] \; \#$

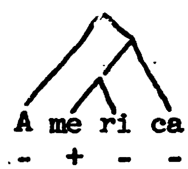
The two cases of this rule are disjunctive -- roughly the generalization is that Case 1 applies to nouns while Case 2 applies to verbs, but there are exceptions on both sides, so that many words will simply have to be lexically marked for which case they take.

An example of Case 1:

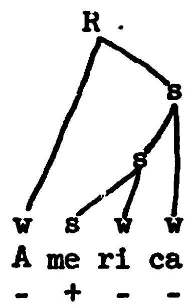
4.1/31 A me ri ca (initial $\pm$ stress assignment)
- - - -

A me ri ca (Rule 30, Case 1)
- + - -

| A | me ri ca | (foot boundary insertion)
| - | + - - |



(bracketing)



(labelling)

Given the system as we have set it up, we could have achieved this same result by means of a rule inserting a foot-boundary in the environment / __ [-stress] [-stress] [-stress] #; we will stick with the formulation in terms of [+stress] assignment, in the interests of keeping constraints on bracketing to a minimum, although I am not absolutely sure this is the right choice.

What we have given so far should be seen as the basic, productive, portion of the English word stress system. Various kinds of morphological and lexical exceptionality will be added, but first, it is appropriate to illustrate the system in its simple form.

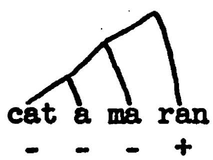
4.1/31

cat a ma ran
\- - - +

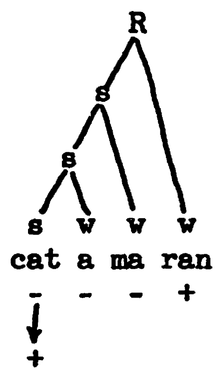
(initial ± stress assignment -- the terminal + must be lexically specified).

| cat a ma ran |
| - - - + |

(foot-boundary insertion)



(bracketing)



(labelling)

$$\left(\begin{array}{c} s \\ | \\ [-] \end{array} \rightarrow \begin{array}{c} s \\ | \\ [+] \end{array}\right)$$

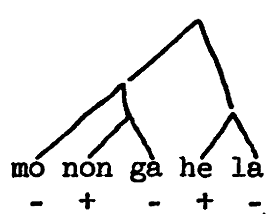
4.1/32

mo non ga he la
\- + - + -

(initial [± stress] assignment -- we assume the penultimate vowel is underlyingly tense).

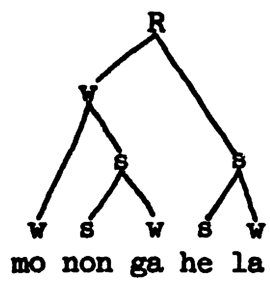
| mo | non ga | he la |
| - | + - | + - |

(foot-boundary insertion)



(bracketing)

4.1/32

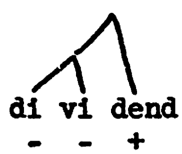


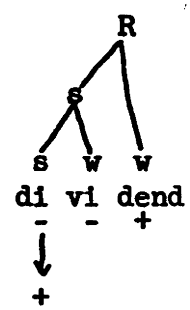
(labelling)

4.1/33

di vi dend
\- - +

|di vi dend|
\- - +

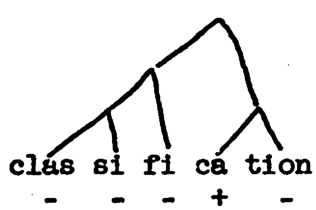


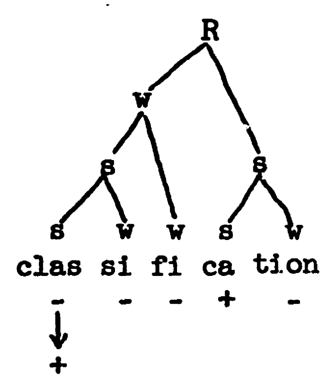


4.1/34

clas si fi ca tion
\- - - + -

|clas si fi|ca tion|
\- - - + -



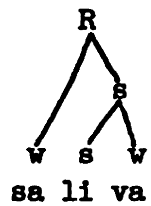


4.1/35

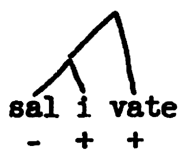
sa li va
\- + -

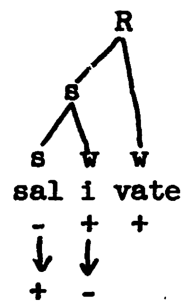
|sa|li va|
\- + -

4.1/35



4.1/36 sal i vate $\qquad$ |sal i vate|
- + + $\qquad$ | - + + |





This last example raises a point of interest. The fact that the initial syllable of "salivate" becomes [+stress] is a regular result of the rule we have referred to previously. The fact that its second syllable becomes [-stress] is the result of another rule:

4.1/38 [± stress] <u>adjustment rules</u> (<u>follow bracketing</u> and <u>labelling</u>):

<u>stressing</u>: $\begin{array}{c} s \\ | \\ [-] \end{array} \longrightarrow \begin{array}{c} s \\ | \\ [+] \end{array}$

<u>destressing</u>: $\begin{array}{c} w \\ | \\ [+] \end{array} \longrightarrow \begin{array}{c} w \\ | \\ [-] \end{array} \; / \; \underline{\quad} CV.$

The next point of interest is the existence of various refinements in the notion of "complexity" referred to in the labelling rule 4.1/27. The listing below is not complete, but will do for our present purposes:

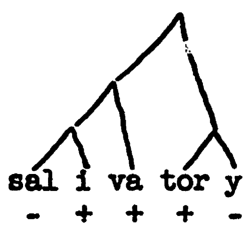
4.1/39a [-ory] is not complex.

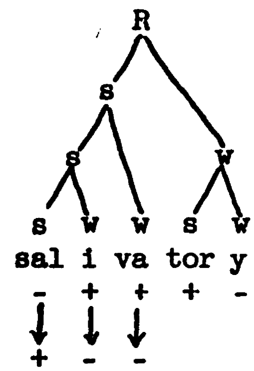
39b [-esque] , [-esce] (et. al.) are complex.

39c Tense high vowels are often complex.

Examples follow:

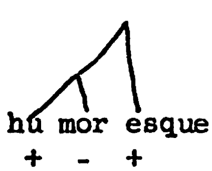
4.1/40 | sal i va | tor y |
      - + + + -

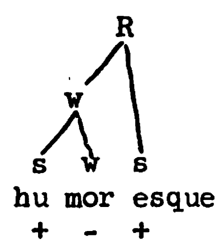




In the example above, the constituent [-ory] would normally be strong, except that 39a specifies that it is not complex -- the result is the "main stress" <u>sálivatory</u>, not <u>salivatóry</u>. The adjustment specifying the segmental feature [+stress] for the first syllable, and [-stress] for the second and third syllables, is regular by the rules in 38.

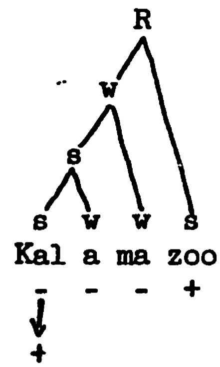
4.1/41 | hu mor esque |
      + - +





In this last example, the fact that -esque is <u>strong</u> depends on the specification in 39b that this affix is to be treated as complex, which yields main stress <u>humorésque</u> instead of <u>húmoresque</u>.

4.1/42 |Kal a ma zoo|
- - - +



The fact that -zoo is <u>strong</u> reflects the fact that the labelling rule 4.1/27 is treating it as <u>complex</u>. High tense vowels often are treated this way -- other examples are <u>guarantee</u>, <u>engineer</u>, <u>kangaroo</u>, <u>macaroon</u>, <u>Japanese</u>, etc. Some of these come from nearly exceptionless classes (-<u>ese</u>, -<u>oon</u>, etc.), while others are more variable (<u>guarantee</u> vs. <u>chickadee</u>). The exact nature of the rule(s) involved is irrelevant to our present discussion.

Another type of morphologically-conditioned deviation from the basic stress patterns imposed by rules 25a, 25b, 27 and 30 is introduced by a pair of rules which affect the distribution of the segmental feature $[\pm \text{stress}]$ in the neighborhood of certain affixes:

4.1/43a $$V \rightarrow [+\text{stress}] \;/\; \left\{ \begin{array}{l} \text{pre=} \\ \text{de=} \\ \text{con=} \\ \text{trans=} \\ \underline{\text{etc}}. \end{array} \right\} C_o \;\_\_\_$$ (the morphological class involved is the class of Latinate prefixes)

4.1/43b $V \rightarrow [+\text{stress}] \;/\; \_\_ \; C_0 \begin{Bmatrix} + \text{ic} \\ + \text{id} \end{Bmatrix}$

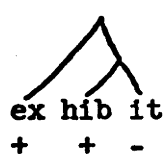
N.B. The issue in these cases is not the actual etymology of the word, but how it is treated synchronically by the morphology.

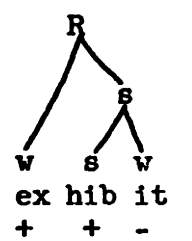
The effect that will be produced by the above rules could also have been produced by special foot-boundary insertion rules; again, we prefer to keep the constraints on bracketing as natural as possible, here introducing instead a pair of basically segmental rules of a familiar morphologically-conditioned type.

Some examples of the effect of these rules:

4.1/44 ex hib it (prefix rule) |ex|hib it|

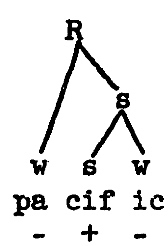
+ - - → + (on hib) |+|+ -|





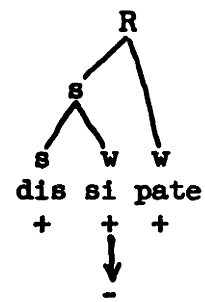
4.1/45 pa cif ic (-ic rule) |pa|cif ic|

- - - → + (on cif) |-|+ -|



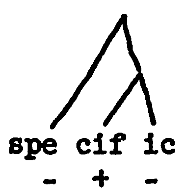
It is worth pointing out that our prefix rule (or the -ic rule, for that matter) will affect the metrical bracketing of the word only if the syllable <u>following</u> the syllable to which it assigns +stress is -stress . Thus in contrast to <u>exhibit</u> we have:

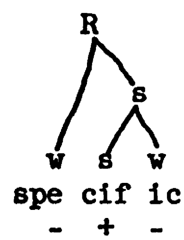
```
4.1/46      dis si pate                 |dis si pate|
             +  -  +                     +   +  +
                ↓
                +
```



An example which shows an alternation that results from this fact is given below:

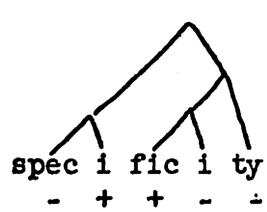
```
4.1/47a     spe cif ic    (by -ic rule)  |spe |cif ic|
             -   -  -                      -  | +   -
                 ↓
                 +
```



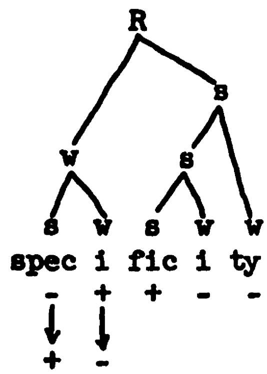



```
but 47b     spec i fic i ty    (by -ic rule and rule 30, Case 1)
             -   - -   - -
                 ↓ ↓
                 + +

            |spec i |fic i ty|
              -   + | +  -  -
```



47b



We are now in a position to evaluate, provisionally, the arguments for the phonological cycle, with respect to word-level stress rules.

The first issue we need to consider is what it would mean for metrical stress-rules to apply cyclically. The idea would be to have the whole system (initial [± stress] assignment, labelling and bracketing, and final [± stress] adjustment) apply, as a whole, first to inner cyclic domains and then to outer ones. But it is not entirely clear what that would mean -- a bracketing which is well-formed on one cyclic level may become ill-formed when additional material is added, and the prospect of adding selective re-bracketing rules is a messy one. There are two alternatives, basically, that would make sense as a theory of cyclic application of metrical word-stress rules. The first alternative would be to restrict cyclic domains to those established by #'s -- on this assumption, no rebracketing would ever be necessary, since #'s already constrain the bracketing process in the way stated in rule 25a. The second alternative would be to say that only segmental information is carried over from one cycle to the next, the metrical bracketing and labelling being reconstructed *de novo*, but on the basis of a distribution of the feature $\pm$ stress which depends in part on the metrical structure imposed on the earlier cycle.

The question to ask is whether any evidence exists for either of these modes of rule application. The alternative is simply to have the word-level stress rules apply freely to their maximal domain, as we argued to be the case for the CSR and NSR.

We will suggest as a provisional conclusion that there is no compelling evidence for either of these injections of cyclic ordering into the metrical theory, and that it is best to continue with the idea that metrical rules apply to their maximal domain, with only that degree of "cyclicity" inherent in the fact that they are rules that assign features to nodes rather than to segments.

There are a number of different sorts of arguments which could be (and usually have been) adduced to prove the cyclicity of word-stress rules. Most of these arguments have the characteristic that they attempt to explain some aspect of the stress pattern of a given morpheme, in a case in which it is part of a larger morphological structure, on the basis of how stress would have been assigned (and on the cyclic hypothesis, <u>was assigned</u>) to a simpler structure.

Let's take a relatively simple case. Consider the pair of words "elasticity" and "devastational." Both have five syllables; in both the main stress is antepenultimate; the secondary stress is on the initial syllable in both cases. There is, however, an interesting difference -- the second syllable of "elasticity" bears some residual stress, and therefore its vowel does not reduce to schwa; by contrast, the vowel in the second syllable of "devastational" is completely unstressed, and becomes schwa. In a conventional theory, this difference might well be attributed to the fact that on an earlier cycle, the main stress was elástic, which will "protect" the second vowel from reduction even though on a later cycle main stress assignment and stress retraction will make it the least stressed of the three syllables under consideration. According to this theory, "devastational" differs crucially, in that the second syllable never bore stress at any time in the derivational history of the word -- the earliest main stress assignment was dévastate. The details of the derivation do not matter here -- the point is that a generalization is claimed to exist that links the stress pattern of the "inner word" with an aspect of the phonetic shape of the "outer word," namely whether or not a certain vowel becomes reduced.

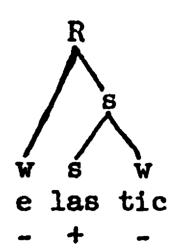
What happens to this generalization in a metrical theory of word stress? Let's begin by deriving the "inner words:"

4.1/48a

e las tic
\- $\rightarrow$ + -

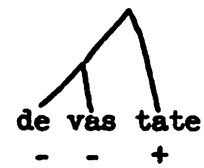
(by -ic rule -- we are assuming that -st- does not in general cause the preceding vowel to be +stress cf. Galveston, Hollister, minister, cannister, bannister, registrar etc.)

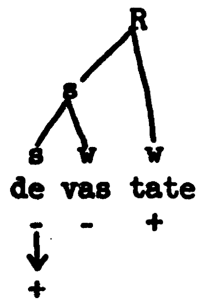
| e | las tic |
| - | + - |



4.1/48b

|de vas tate|
 - - +

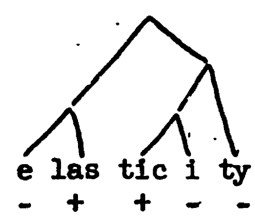


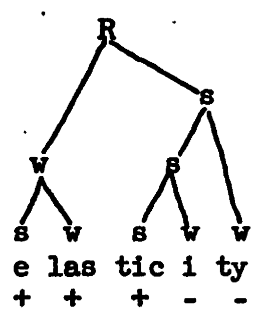


It should be clear that the factor producing the difference in the stress pattern of these "inner words" -- namely the nature of the affix -<u>ic</u>---remains constant in the domain of the "outer words:"

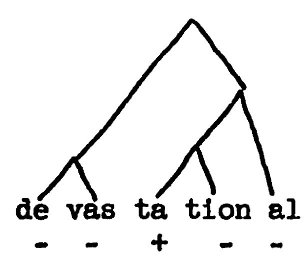
4.1/49a e las tic i ty (by -<u>ic</u> rule; by rule 30 Case 1)
- - - - -
↓ ↓
+ +

|e las|tic i ty|
 - + | + - -

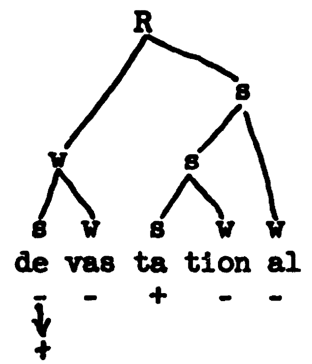




49b |de vas|ta tion al|
 - - | + - -



4.1/49b



In getting these results (the same results achieved by the cyclic analysis) it is not necessary to first derive the "inner word" and then derive the "outer word" -- it is only necessary that the factor producing the result exist in both derivations. We are not claiming that the position of main stress in "elastic" and the failure of the second vowel to reduce in "elasticity" are unrelated phenomena. On the contrary, these two facts have the same explanation, in our theory; but that explanation does not involve cyclic rule application.

A brief note on the facts just discussed: the second vowel of "elasticity" is not destressed by our adjustment rule 38, because it is followed by two consonants. This pattern is not altogether regular --
alongside exaltation (3 4 1), relaxation (3 4 1), etc. we find consultation (3 1), transformation (3 1), etc. These cases require exceptionality of some sort, somewhere, in any system of stress rules I know of -- the question of just where in our theory it would be most appropriate to introduce this exceptionality is not relevant to the present discussion, and will be left open.

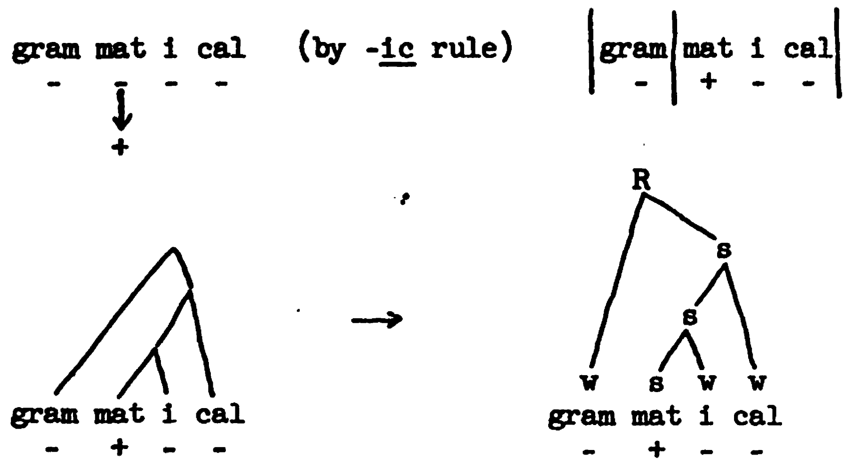
There are other classes of cases in which the position of secondary stresses in complex words correlates with the place where main stress would have been assigned in simpler words. For example, we have

1 / 3 1; 1 / 3 1

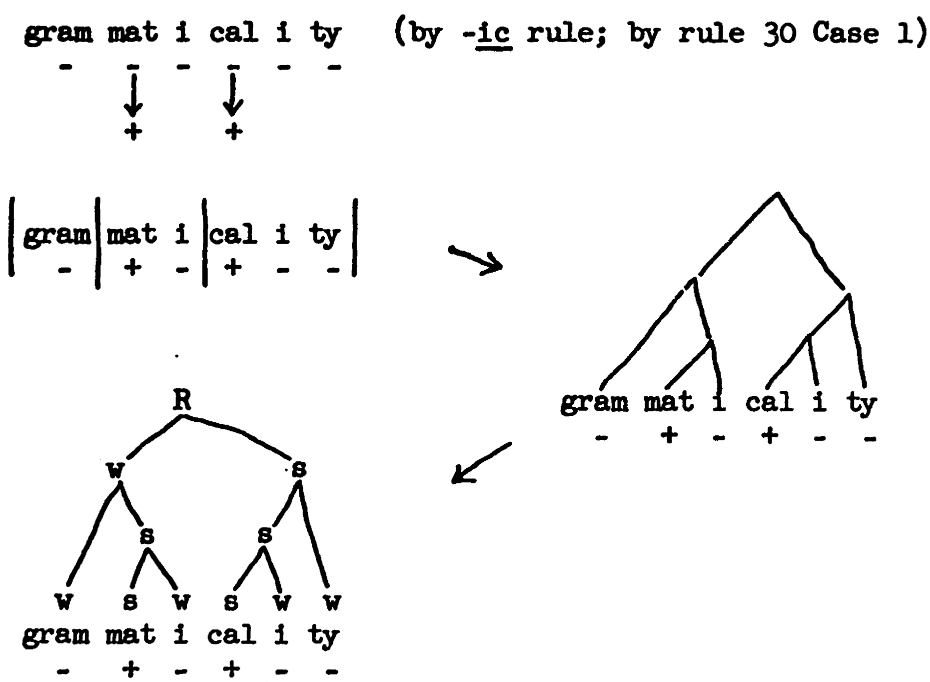
grammatical / grammaticality; despicable / despicability.

These cases will also be derived regularly, both in the smaller and larger word, without the need for cyclic rule application:

4.1/50a



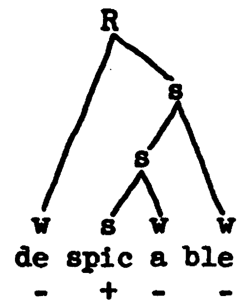
50b



50c de=spic a ble (by prefix rule -- also consistent with rule 30 Case 1)

$$\begin{array}{cccc} \text{de} & \text{spic} & \text{a} & \text{ble} \\ - & - & - & - \\ & \downarrow & & \\ & + & & \end{array}$$

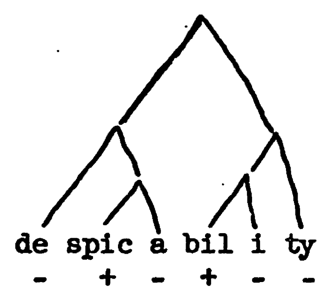
$$\left|\,\begin{array}{c} \text{de} \\ - \end{array}\,\right|\begin{array}{ccc} \text{spic} & \text{a} & \text{ble} \\ + & - & - \end{array}\left.\right|$$

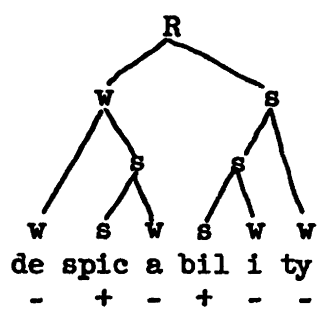


50d de=spic a bil i ty (by prefix rule; by rule 30 Case 1)

$$\begin{array}{cccccc} \text{de} & \text{spic} & \text{a} & \text{bil} & \text{i} & \text{ty} \\ - & - & - & - & - & - \\ & \downarrow & & \downarrow & & \\ & + & & + & & \end{array}$$

$$\left|\,\begin{array}{c} \text{de} \\ - \end{array}\,\right|\begin{array}{cc} \text{spic} & \text{a} \\ + & - \end{array}\left|\,\begin{array}{ccc} \text{bil} & \text{i} & \text{ty} \\ + & - & - \end{array}\,\right|$$

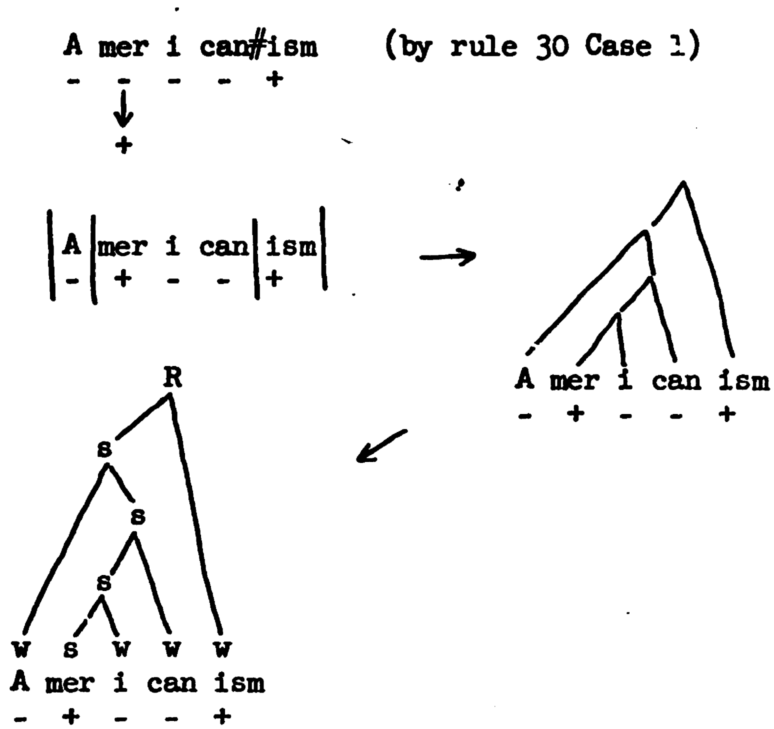




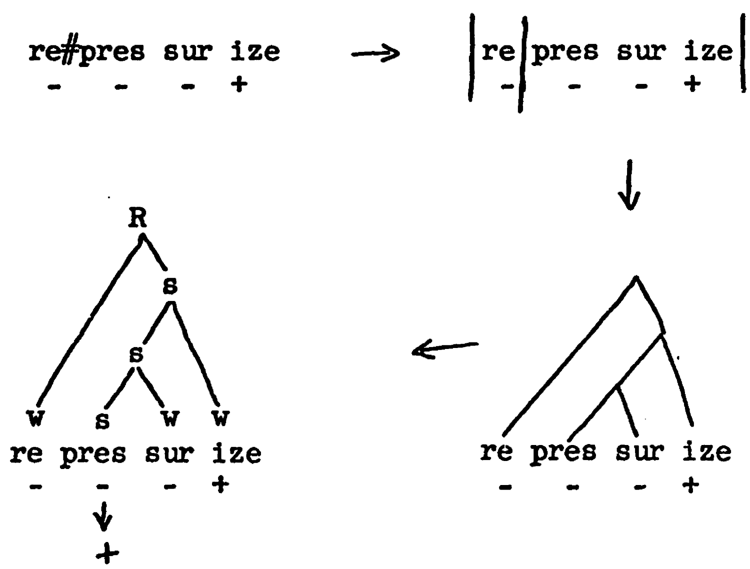
Another class of cases in which cyclic rule application might seem to be indicated are the cases involving productive affixes such as <u>re</u>- , -<u>ism</u> , -<u>ness</u> etc. These affixes (in their productive manifestations) never affect the stress pattern of the parent word.

This result will follow immediately and automatically (and again, without cyclic application of rules) if we assume (as is the common belief) that these affixes are separated from the parent word by the boundary #, which (probably universally) constrains metrical bracketing:

4.1/51a

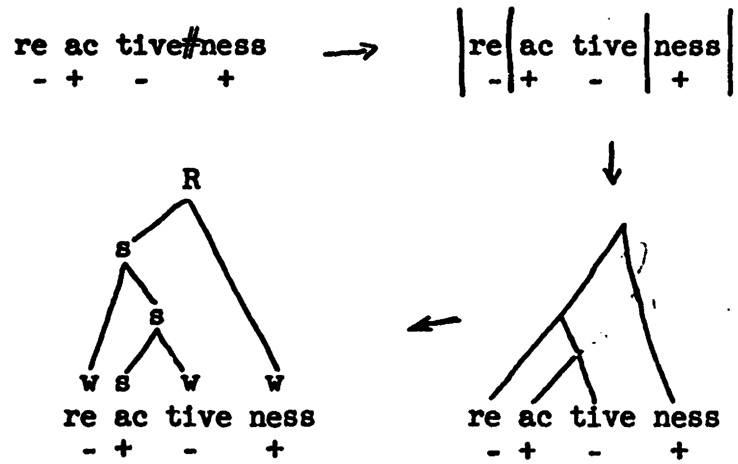


51b



actually [illegible]

51c



There is another important class of cases which have been taken to argue for the word-stress cycle -- the class of noun-verb stress-doublets like contrast[1] / contrast[1 3], implement[1] / implement[1 3]. Within the conventional framework, serious questions have been raised about the cyclic treatment of the phenomena involved (<u>Ross</u> 1972, 1973, <u>Aronoff</u> 1974), and it has been suggested on a number of grounds that a non-cyclic treatment of the same facts is preferable. Since the facts are so complex and exception-ridden in the conventional theory, whether treated cyclically or not, it does not appear that any compelling argument for the phonological cycle will emerge from such cases. This is especially true in light of Aronoff's argument that the typical noun-verb differentiation cannot be explained on the basis that the nouns in question are deverbal (a key assumption of the cyclic solution), since in most cases the verb is clearly denominal (e.g. fragment[1] / fragment[3 1]). Aronoff's argument is bolstered by the historical fact that the nominal pattern is original, with the verbal pattern developing (gradually, it appears) not so very long ago. Joshua Steele, writing in 1775 (<u>An Essay Towards Establishing the Melody and Measure</u>

of Speech), says: "In the foregoing list of words it will be seen, that the syllables in some verbs are of a different POIZE from the same syllable in the kindred noun. This useful distinction is, I believe, not of very long standing. I remember when it was in fewer words than it is now; and, I think, it is a good deal in the power of the learned, by art, to make it almost, if not quite, general."

Indeed, the tide seems to have receded somewhat since Steele's day, judging from the list of words he gives. I have no doubt that it is now in the power of the learned, by art, to give an almost (if not quite) general account of the phenomenon. It seems unlikely, however, that this account, in terms of whatever theory it is expressed, will provide much evidence for the cyclic application of stress rules.

We will not attempt to give such an account here, so that the bearing of the phenomenon of noun-verb stress doublets on the issues under discussion will have to be left unsettled.

The phonological system of English, of which we have been treating a fragment, is large and complicated. It is always dangerous to come to sweeping conclusions on the basis of a partial analysis of such a system -- every theoretical choice has far-reaching (and often unexpected) consequences, in remote corners of the description. Therefore, a firm conclusion about the possibility of replacing the concept of the phonological cycle with a metrical theory of stress patterns, must await further research.

However, it is worthwhile to underline two points -- 1) a metrical theory of stress patterns (in which stress levels are treated as a structural, rather than a segmental phenomenon) predicts a kind of implicit "cyclicity" for stress rules, without any apparatus of cyclic rule application; 2) such rules, applying simultaneously to their maximal domains, have been successfully (and indeed easily) written for the English phrasal stress system, and for a medium-sized fragment of the word-stress system. Given the well-known typological differences between prosodic and segmental rules, the explanation offered by the metrical theory (that stress rules are of a fundamentally different type from segmental rules, and thus have inherently different properties, which include the misleading appearance of cyclic application) must be granted a certain initial plausibility.

One last point. In his article <u>Stress Rules in English -- a New Version</u> (LI IV 4 1973), Morris Halle noted a similarity between the "detail rule" proposed by Schane (whose effect is to "place main stress on the last stressed syllable in a word that precedes the final syllable") and the Compound Stress Rule. He proposed that they should be collapsed into a sort of super CSR, for which he gave the formulation below:

4.1/52 $$\begin{bmatrix}\text{1 stress}\\ \text{+ syl}\end{bmatrix} \rightarrow [\text{1 stress}] \;/\; \_\_\_ \; Q\; (\#\#P)(ic{+}at)\; V\; C_o({+}y)\,]_{N,\,A,\,V}$$

In our system, Schane's "detail rule" represents a special subcase of the operation of the word-level node-labelling principle, "in a metrical constituent [M N], N is strong if and only if it is complex." It is interesting to compare this word-level labelling rule to our formulation of the CSR, "in a lexical constituent [M N], M is strong if N is a single word." This could easily be reformulated, "in a lexical constituent $[M\ N]$, N is strong if and only if it contains more than one word." Our definition of "complex" in the case of the word-level rule was "dominating non-terminal material" (with a few kinds of lexical exceptionality). If we adopt the rather natural position that from the point of view of the domain of a single word, terminal elements are syllables, while from the point of view of the domain of a lexical compound, the terminal elements are words, then our word-level node-labelling rule and the compound node-labelling rule become one single rule, basically "in a lexical constituent $[M\ N]$, N is strong if and only if it is complex (i.e. dominates nonterminal material."

We will still need lexical input to the notion "complex," on the word level, but the fundamental identity of the two rules should be apparent. On this view, there are really only two (metrical) stress rules in English -- one for phrasal constituents (the NSR) and one for lexical constituents (the CSR). At the word level, there are certain additional segmental rules for assigning the feature $[\pm \text{ stress}]$ (e.g. our rules 30 and 43), but these are of a fundamentally different character.

Thus the picture that results from reinterpreting the English stress system in terms of a metrical theory is quite a promising one, although much work remains to be done. It is, I think, a strong validation of our metrical theory of tune-text association, that its constructs work so well in areas rather far removed from their origin.

4.2 Stress Shifting Rules.

In the preceding section, we argued that a metrical approach to stress has the advantages that rules can be stated more simply (without variables, etc.) and that no principle of cyclic rule application is required. We argued that the appearance of cyclic application arises only because stress rules deal with nodes rather than segments, and therefore that similar properties of apparent cyclicity would not be expected for segmental rules.

However, in these cases the advantages of the metrical approach have been largely theoretical -- a potentially more elegant treatment, the possibility of restricting the class of allowable stress rules more narrowly, and an interesting hypothesis about the typology of prosodic vs. segmental rules. In the conventional sort of system, in which stress is treated as a segmental (rather than structural) feature, ingenious methods have been devised for representing generalizations about stress patterns, methods intended to allow the statement of desirable rules, and disallow "impossible" (i.e. non-occurring) kinds of stress rules. By and large these methods (the cycle, various sorts of variable, etc.) are descriptively excellent, in the hands of a skillful operator, and usually can be argued to capture real generalizations.

Therefore, at least in the domain of data we have examined up to now, the main benefit of the metrical approach is simply that it does a little more naturally what the conventional kind of system already did very well indeed. We will now examine, rather briefly, a kind of phenomenon which can be represented quite naturally in a metrical system, but which lies outside the competence of the conventional way of dealing with stress. It is not claimed that the facts cannot be described, in a conventional system, although such description is often quite difficult, but simply that the rules that result are a horribly complex way to represent a very simple phenomenon, and require a kind of expressive power which would allow the statement of many humanly impossible stress systems.

The primary case we will inspect is the rule in English which

3 1   2 4 1

produces "thirteen," but "thirteen men." Here are some other examples of its operation, in numerological notation:

	3 1		2 4 1
4.2/1a	absolutely	<u>but</u>	absolutely clear
	3 1		2 4 1
1b	tennessee	<u>but</u>	tennessee Ernie
	3 1		2 4 1
1c	anaphoric	<u>but</u>	anaphoric reference

Now, what is going on is pretty clearly that main word stress is veing shifted backwards, under certain conditions, to what normally would be a secondary stress. The reason for the shift seems pretty clearly to be the desire to maintain alternating patterns, but the teleology of the rule needn't concern us for the moment. Obviously, the rule cannot take place on the cycle which deals with the lower word (assuming this way of assigning phrasal stress), since at that point we have no way of knowing whether or not the word is terminal. Furthermore, it makes a difference where the main stress of the second word is:

4.2/2 preferential (3 1) but preferential (2 4) treatment (1)

but preferential (4 2) recommendation (3 4 1)

So the rule has to be applied on the upper cycle. We will state it as if it applies after stress assignment has taken place on this upper cycle, for the reason that it does not usually apply at all, if for some reason the left-hand constituent receives main stress:

4.2/3a the Mississippi (2 4) Delta (1) -- but, where we have been talking about deltas in general,

"how about the Mississippi (4 1) Delta (2)?"

3b a person who loves Tennessee is a "Tennessee-lover (4 1 2)," not a "Tennessee-lover (1 4 2)."

For the cases we have considered so far, the rule might be written as a sort of stress-number metathesis, formulated approximately as follows:

4.2/4 S.D. X 4 (0) 2 (0) 1 Y

S.C. X 2 (0) 4 (0) 1 Y

However, this will not do -- although we have seen the rule apply in cases of the form "4 2 1" (thirteen men), in cases of the form "4 0 2 1" (Tennessee Ernie), and in cases of the form "4 0 2 0 1" (anaphoric reference), we have not seen any cases of the form "4 2 0 1."

There is a good reason for this gap in the set of examples presented -- in such cases the stress does not, cannot, shift back:

```
            4 2 0  1         2 4 0  1
4.2/5a      fantastic deal  not  fantastic deal

            4  2 0  1 0       2  4 0  1 0
    5b      Montana license  not  Montana license
```

Thus the rule stated above must be modified:

4.2/6 S.D. $X\ 4 \langle(0)\rangle_b\ 2 \langle(0)\rangle_a\ 1\ Y$

S.C. $\quad 2 \qquad 4 \qquad 1$

<u>condition</u>: <u>if</u> <u>a</u> <u>then</u> <u>b</u>.

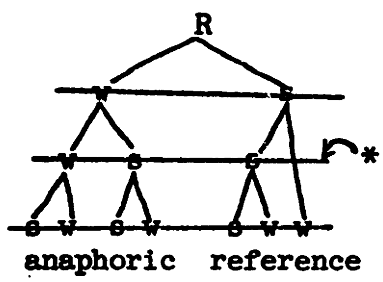
There are undoubtedly other ways in which the rule would have to be modified; for example, there can be more than one syllable intervening between the 2 and the 1 (although in such cases the rule becomes somewhat more optional):

```
            2    4          1                   4    2          1
4.2/7       absolutely expressionless    or    absolutely expressionless
```

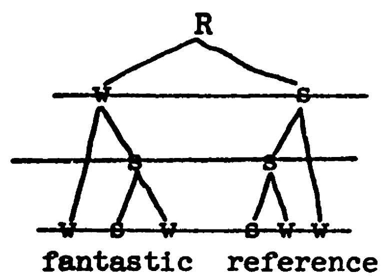
N.B. I definitely prefer the first version, even in this case, unless the adverb *is* being emphasized.

We will not, however, attempt to finish the task of tracking down all of the details of the rule. The point we want to make is that the condition on rule 6 (which allows 4 2 1, 4 0 2 1 and 4 0 2 0 1, but prohibits 4 2 0 1) is actually a reflection of the basic nature of the rule, which is to create alternating patterns. If we treat such stress patterns as sequences of an n-ary segmental feature, this generalization is very difficult to state. In a metrical theory, however, in which the stress pattern is represented as a labelling of the nodes of a tree, the nature of the phenomenon is rather clearer. Consider the distinction between "anaphoric reference" and "fantastic reference," as they are represented metrically before the operation of the stress shift rule:

4.2/8a



8b



If we mark hierarchically equal levels in the metrical tree (computing from the bottom up), as we have done with the horizontal lines in the above figure, then we notice that in 8a one level (the one marked with a star) has the sequence w s s, and that the change necessary to produce the observed stress pattern is to turn the first of the two <u>s</u>'s to <u>w</u> (with the concomitantly required change in its sister node). No level in 8b has this character. As we will see, this property differentiates all of the examples where stress shift occurs from those in which it doesn't. The rule may be written as follows:

4.2/9 S.D.

$$\begin{array}{ccccc} & & W & S(\text{or } R) & \\ & & | & | & \\ X & W & S & S & Y \\ 1 & 2 & 3 & 4 & 5 \end{array}$$

S.C. $3 \rightarrow W$

N.B. The S.D. is defined on hierarchically equal levels of the tree, as explained above, and is thus not a conventional condition on proper analysis of the terminal string. The reason for this has to do with the nature of rhythm (which is the real basis of the rule), and will be discussed in the next chapter.

Condition: the effect of the rule is not to create the configuration

$$\begin{array}{c} s \\ | \\ [\text{-stress}] \end{array}.$$

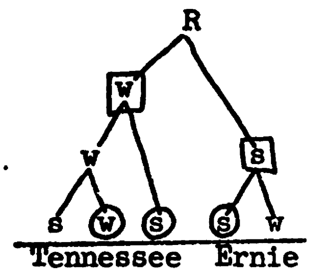
Granting, for the moment, the allowability of S.D.'s on hierarchically equal structural levels, the only aspect of the rule which is not simply an expression of the intuition that secondary stresses are moved back when they phrasally precede a primary stress, is the condition that the stress not be shifted back onto a stressless segment. This is a highly natural and reasonable requirement -- one aspect of the cyclic idea that seems to be true is that sentence stress rules can never change the segmental phonology of the word, even though (as in this case) they may change its metrical structure.

This rule (or the somewhat modified version of it to be discussed in the next chapter) has three extremely important advantages over the rule in 4.2/6:

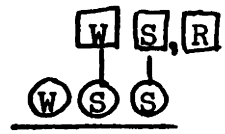
1) it captures the generalization hidden in the angled-bracket part of 4.2/6 -- that the nature of the rule is to break up sequences of strong elements, in order to create alternating patterns.
2) it needs no angled brackets, parentheses or internal variables in its statement.
3) its structural change is a change in a single element ("3 $\longrightarrow$ W"), not a metathesis of stress numbers, or however exactly it is to be done in the conventional framework.

To demonstrate the approximate validity of our rule, we will list without comment a number of other cases where it applies, as well as other cases where it does not:

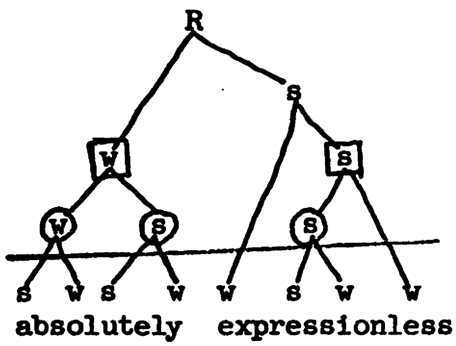
4.2/10a



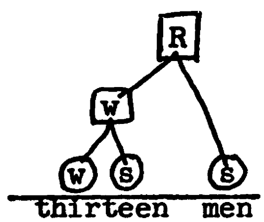
Approved Configuration Indicated in This Way:



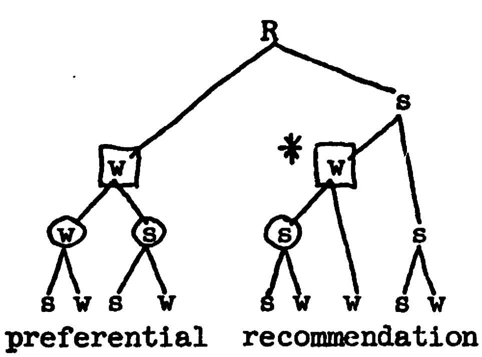
10b



10c



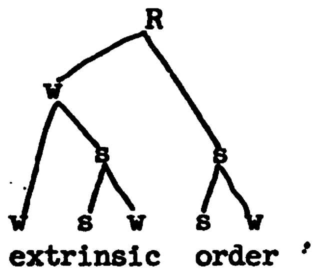
4.2/11a



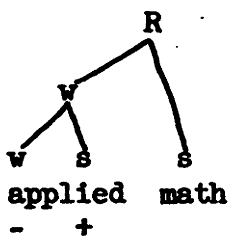
N.B. in the above example, there is a sequence w s s, but the second strong position is immediately dominated by a weak position:

w w
| |
w s s, which means it fails to meet the rule's S.D.

11b



11c



N.B. the S.D. is met, but application of the rule would shift word-stress back onto an unstressed syllable; therefore, the rule cannot apply. The constraint that syllables immediately dominated by a strong metrical node must be stressed, is a pervasive feature of the derivation. At the word level, this constraint appears in the form of the rule that changes

s to s
| |
[-] [+] ; at the phrasal level, the constraint functions to prevent the application of the stress-shift rule, if that application produces the forbidden configuration. This state of affairs suggests the truth of the familiar idea that phrasal stress rules cannot alter the segmental features of lexical items. This is a partial equivalent, in our theory, of the "strict cyclicity" of the conventional theory.

There is more to be said about this rule, but we will postpone the discussion to the next chapter, where we will try to explain why such a rule should exist, and why it has the particular form that it does. The main point that we want to make in this section is that rules that shift stress can be written in a very natural way in a metrical framework, since a local adjustment (e.g. changing a strong node to weak) will automatically cause an adjustment of the label of the sister of the adjusted node. Therefore, it is not necessary to indulge in complex "structural changes" such as stress-number metathesis. The fact that we can also capture a generalization about the environment in which the rule applies is a distinct bonus.

Rules of this kind are not uncommon. Other Germanic languages (German, Swedish) show a similar pattern, as Paul Kiparsky has pointed out to me. In Biblical Hebrew, there is a phenomenon known as "Nasog ʔAḥor" (the standard example, comparable to our "thirteen men," is tōkál léḥem → tṓkal léḥem "she eats bread"). Given the refusal of the English rule to restress unstressed vowels, it is interesting that the Hebrew case shows the vocalism that would be associated with the original stress pattern, rather than the derived one. This is not apparent in the example given above, for irrelevant reasons, but in ʔōḵḗl léḥem → ʔṓḵēl léḥem "eater of bread," we would expect the first word to have a short second vowel, ʔṓḵel, if its surface stress pattern fed the appropriate rules. (These facts, and their interpretation, were provided to me by Alan Prince, in whose thesis the reader will discover a penetrating analysis of these (and other) aspects of Hebrew phonology).

In chapter 5, we will argue that the similarity of the English and Hebrew phenomena should not be considered an accident, nor a subtle effect of the Diaspora, but rather the independent emergence of similar solutions to similar problems.

4.3 Some Other Influences on Phrasal Stress Patterns

In the treatment of phrasal stress given in 4.1 (the discussion of the NSR) the only factors affecting the "stress pattern" which emerged from the node-labelling process were the constituent structure and the order of elements in a given constituent. We purposely abstracted away from the many other factors which, interacting with the "Nuclear Stress Rule," yield the result that the patterns of stress observed in everyday speech violate the NSR nearly as often as they follow it. We do not conclude from the prevalence of patterns which violate the NSR that no such principle exists, but only that there are other factors which can pre-empt it. On this view, the NSR becomes a sort of "last resort" or "null hypothesis" principle -- "put the strong element on the right in any given metrical constituent, if you have no good reason to do otherwise." Thus the view of [ws] as the "normal" phrasal constituent does not imply that it is necessarily the most common one in the usage of any given individual, but simply that [sw] constituents will occur only where there is a good reason to prefer them.

What such "good reasons" might be, we will consider shortly. For some reason, the idea that there should be a "null hypothesis" stress-assignment principle seems to draw opposition like flies to honey; perhaps it is because of general hostility to "rules" as "narrow" and "uncreative," especially when the "rules" involved are stated in a formal way. The usual counterargument to the idea of "stress rules" (setting aside a more fundamental objection, that denies the existence of any category of phrasal <u>stress</u> separate from tonal inflection) is that phrasal stress is purely and simply a method of emotional or informational highlighting. We are about to take a brief look at cases in which such influences determine stressing; but why should it be so hard to accept the idea that in the absence of other considerations, some simple and useful structural principle should intervene? In the phrase "a red cow," I might want to "highlight" the animal's redness, in which case I would give <u>red</u> the highest stress, from which would regularly follow its association with the "nuclear tone;" I might also want to highlight the animal's cowness, in which case (interestingly enough) I would probably be forced to use an extra-high pitch in the "nuclear tone" on <u>cow</u> (assuming a basically rising-falling type of tune), an option that was not especially necessary in the previous case, when the main stress was on the NSR-violating left branch.

However, I might simply want to speak of a <u>red cow</u>, desiring neither to highlight the animal's redness nor to highlight its cowness, but purely and simply to point out that it is a <u>red cow</u>, or to refer to it by that designation in the context of some more general observation. In this case, the "nuclear tone" (and the perception of greater stress) will settle on the rightward element, <u>cow</u>, an observation which (when generalized) is equivalent to the hypothesis that the "nuclear stress rule" defines the unmarked stress pattern of the phrase. I find it inconceivable that the prosodic/tonal system of English should prohibit me from using the phrase "a red cow" in a neutrally designative fashion, if I so desire; if a "null hypothesis" stress rule did not exist, I should personally feel perfectly justified in unilaterally inventing one, to suit the needs of everyday usage. It seems clear that I have been anticipated in this opinion by a fair number of previous generations of humanity, with the result that I do not need to intervene personally in the structure of English, but can make use of the simple and effective convention (one of only two possible choices, according to the theory of metrical patterns) which is in common use in the language.

The concept of a "null hypothesis" requires, of course, that other options exist. In what remains of this section, we will take a brief tour through some of the circumstances in which other factors besides left-right come into play.

We begin with a class of cases which do not really belong in the tour; these are the cases of systematic constructional "irregularity" of the sort mentioned in chapter 1 (madison avenue/madison street, steel warehouse/steel warehouse). These cases are of two types; in one, the stress pattern is determined solely by the head of the construction (<u>avenue</u>/<u>street</u>), while in the other, it is the nature of the modifier-head relationship that seems to be crucial.

In dealing with irregularity of the <u>avenue</u>/<u>street</u> variety, we have two options. The first is to specify that [madison avenue] is an NP, while [madison street] is a compound noun. This option seems highly undesirable, since they do not show any appropriate difference in syntactic behavior. The second (and preferable) alternative is to find a way to specify their irregular prosodic behavior by some mechanism specifically tailored to represent such phenomena.

Exactly such a mechanism exists, by virtue of our introduction of lexical exceptionality into the definition of the notion "complex" for the word-level node-labelling rule (4.1/39). The fundamental definition of "complex" was "dominating non-terminal material;" we were forced to add to this a list of systematic (but morphological) criteria, like "dominating one of the affixes -esque, -esce, etc."

At the end of section 4.1, we argued that the compound stress rule and the word-level node-labelling rule should be collapsed. This principle will assign strong on the right in a given constituent if the right-hand element is complex, and on the left otherwise, where (at the level of lexical compounds) words are "terminal elements" for the purposes of computing "complexity." We have the option of modifying the definition of "complex" in lexically systematic ways at this level, if we desire; so we could add to the category of complex nodes (which already includes a list of affixes) a list of words, specifically avenue and its fellows.

On this view, madison avenue and madison street are both compound nouns, whose stress patterns are related roughly in the way that humoresque is related to humorous. Since any kind of avenue, from Dorchester to Honolulu, always takes stress on the right, just as any kind of -esque does, this mode of description seems appropriate to represent the systematic lexical caprice which seems to underlie both the stress-attracting affixes and the stress-attracting nouns. Of course, this "stress-attracting" property will only operate in cases where the placement of stress depends on a rule which mentions "complexity." Since the NSR cares for nothing but left-right order, avenue will "attract stress" only if it is in a lexical compound, and (even then) only if it is on the right.

The second case of systematic constructional "irregularity" involves noun + noun collocations which (contrast etc. aside) have one meaning with stress on the left, and a different meaning with stress on the right. The example we gave in chapter 1 was stéel warehouse, "a warehouse for storing steel," vs. steel wárehouse, "a warehouse made of steel."

> (N.B. We are assuming that warehouse is synchronically a single word, not a real compound -- those who doubt this can choose some other example, like BRICK store vs. brick STORE).

Given the systematic difference in meaning that characterizes these cases, it seems quite appropriate to analyze the ones that take stress on the left as noun-noun compounds, and the ones that take stress on the right as involving an inherently adjectival modification, both syntactically and semantically. There are a number of more specific arguments in favor of this position, I believe, but they would be out of place in this discussion. On the assumption that (stéel warehouse) is a lexical constituent, while (steel wárehouse) is a phrasal constituent, their stressings will be regular. Presumably the internal structure of the second case is something like

4.3/1 $$\left[_{NP} \left[_{Adj} \left[_{N} \text{steel}_{N}\right]_{Adj}\right] \left[_{N} \text{warehouse}_{N}\right] _{NP}\right]$$

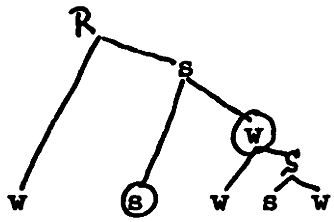
The configuration $\left[_{Adj}\left[_{N}\text{steel}_{N}\right]_{Adj}\right]$ is seen as resulting from a productive morphological process that turns nouns into adjectives, with a predictable effect both on stress and on meaning.

The preceding discussion is really out of place in a tour of the non-structural aspects of English stress, since the factors involved are actually of a purely formal nature, with no possible relation to any issue of "highlighting" or "relative importance," and therefore demonstrate all the more conclusively that there are non-pragmatic factors at work in determining English stress above the word level. I have included them in the discussion because I have several times had the unsettling experience of hearing such cases placed in evidence by people who were trying to argue against the idea that there are any rules at all in this domain, other than "what is meant" or "what is important." Using these facts in such an argument strikes me as comparable to claiming that one's fingerprints on the murder weapon are proof positive of innocence.

These remarks behind us, we will proceed to the other two points of interest on our tour, which are the area of "communicative dynamism," and the area of the distinction lexical/nonlexical/nonexistent.

Under the heading of "communicative dynamism" (a Prague school term) we include those phenomena which are commonly called <u>contrastive stress</u> and <u>anaphoric</u> (or <u>redundancy</u>) <u>destressing</u>. A phenomenon that might be called <u>emphatic stressing</u>, which relates pragmatically to the idea of degree of communicative importance, but which we regard as arising from a special kind of alignment with the metrical <u>grid</u>, rather than from any change in metrical <u>pattern</u>, will be discussed briefly in chapter 5 (in relation to example 5/35). We will say very little about anaphoric destressing, except to point out that it creates exceptionalities to the NSR:

4.3/2



If John were here, Sam'ld hit the bastard.

Presumably the reason that the constituent which contains the circled nodes has the form [sw] , instead of the "null hypothesis" form [ws] , is that anaphoric or redundant material is considered intrinsically *weak*, as opposed to anything that adds new information to the conversation. The other side of the coin is obviously that crucial or extra-important new information will want to be *strong*. Often it is hard to determine, in a particular case, what represents the result of destressing old information and what represents the result of stressing new information. One not unreasonable hypothesis would permit free assignment of either [ws] or [sw] to any phrasal constituent, but would stipulate that the pattern [sw] will be assumed to have some special pragmatic justification. The alternative is to postulate some set of features like [± contrastive] , assigned to nodes, which influence prosodic and tonal phenomena in some way. The main difficulty in this area is that nobody has ever gotten the facts straight, to my knowledge. There are four factors whose description should be kept separate (since otherwise it is hard to evaluate the question of whether they spring from a single source, "contrastive stress," or whether they represent a complex of phenomena incorrectly lumped together in that catch-all category). These phenomena are 1) stress differences; 2) tune differences; 3) register differences; 4) rhythmic differences. The fact that the association of different tunes with "contrasted" items plays a role in their interpretation is discussed in Jackendoff 1972; in terms of our theory, the differences he describes represent in each case the association of a certain tonal word with the contrasted constituent, with variations in the terminal boundary-tone that result in large differences in interpretation. His discussion suggests that not only the stress pattern of the text, but also the choice of associated tune, the issue of whether the tune is simple or complex, and the issue of which parts of a complex tune are associated with which parts of the text, are all important factors in determining the final interpretation. The same factors, with scarcely any modification, can also play a role in determining the interrelationship of clauses and groups of clauses, in examples that no one (as far as I know) has ever considered to be associated with any issue of "contrastive stress."

It should also be clear that expanding or contracting the range of the voice makes a difference -- in positions in which the nuclear stress would normally make strong anyway, an expanded pitch-range seems to be almost the only way to indicate "contrastive stress." This expanded pitch range (which marks "contrast" by a higher high tone, or a lower low tone, depending on the tune involved) is also possible in positions which would not normally be assigned main stress by the NSR.

However, as we pointed out in 3.3, such cases are tonally differentiated from "normal stressing" anyway; it is unclear to me whether expanded pitch-range adds anything other than vehemence to this differentiation.

As we will point out in chapter 5, it is also possible to add weight or emphasis to a stress by purely rhythmic means; this means may be employed either with or without a change in metrical pattern.

It is quite possible that all of these complicated and inter-related aspects of "intonation" are regulated in such cases by a single feature (+Focus) or something of the sort. It is also possible that there ought to be a more differentiated analysis, in which different intonational features have systematically independent distribution but (in certain cases) similar effects. I personally lean towards the second approach, but no arguments will be presented here.

It was the hope of coming to grips with problems of this sort that led me to study intonation in the first place. My initial plan for this thesis involved developing what I expected to be a relatively trivial notational system, and then using it to help unravel the tangled threads of syntax, semantics and pragmatics. The task of defining the terms of the enterprise has expanded, for good or ill, to fill up the bulk of the document, and I have decided to omit any report of my preliminary attempts to get on with the rest of the task, although I hope that they will bear some expositional fruit in the future.

So we will simply observe, at this point, that a contrasted constituent will always be metrically <u>strong</u>, if it can be. We might also point out in passing that if prosodic features are features of structure, rather thar .eatures of segments, the "strength" of a contrasted constituent will automatically be felt as governing the "stress-level" of its designated terminal element. In Deep Structure, Surface Structure and Semantic Interpretation, Chomsky observes that a phrase which he notates as

4.3/3 an ex-convict with a red SHIRT

could be taken as "focusing" on any of the constituents in the list below:

4.3/4 (i) an ex-convict with a red shirt

(ii) with a red shirt

(iii) a red shirt

(iv) shirt

It is a consequence of our approach to prosodic and tonal phenomena that any attempt to add extra "strength" to any of the listed constituents will give prosodic and tonal prominence the word which is the designated terminal element (by simple application of the NSR) of all of them, namely SHIRT.

As a result, it is neither necessary to add a rule interpreting the "focused constituent" as any constituent dominating the "capitalized" word" in surface structure (we mean by "capitalized word" the locus of some perception of "supranormal prominence"), nor to add a phonological rule adding "extra stress" to the 1 stressed vowel of a focused "stress-level" of its designated terminal element. In Deep Structure, Surface Structure and Semantic Interpretation, Chomsky observes that a phrase which he notates as

4.3/3 an ex-convict with a red SHIRT

could be taken as "focusing" on any of the constituents in the list below:

4.3/4 (i) an ex-convict with a red shirt
(ii) with a red shirt
(iii) a red shirt
(iv) shirt

It is a consequence of our approach to prosodic and tonal phenomena that any attempt to add extra "strength" to any of the listed constituents will give prosodic and tonal prominence the word which is the designated terminal element (by simple application of the NSR) of all of them, namely SHIRT.

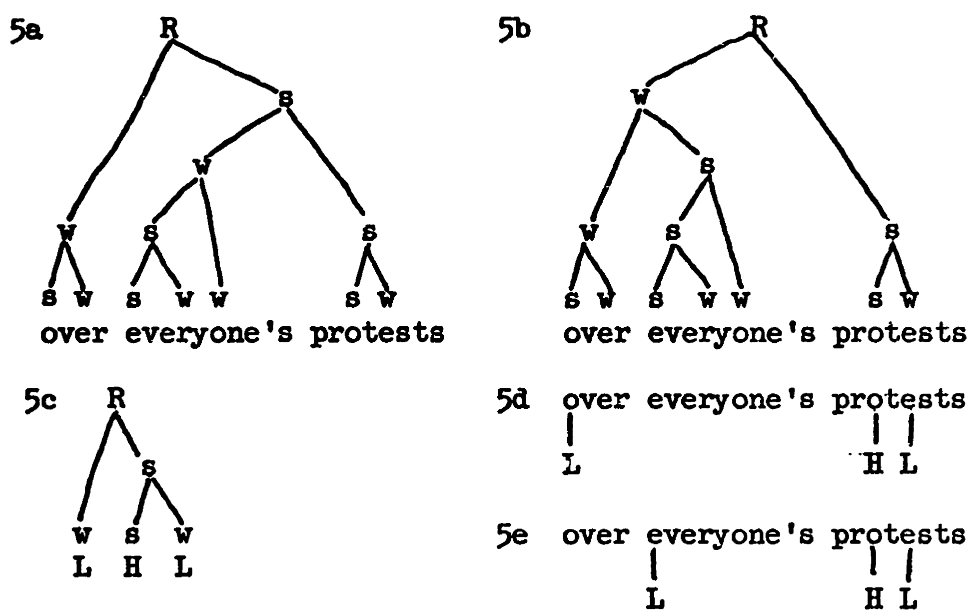
As a result, it is neither necessary to add a rule interpreting the "focused constituent" as any constituent dominating the "capitalized" word in surface structure (we mean by "capitalized word" the locus of some perception of "supranormal prominence"), nor to add a phonological rule adding "extra stress" to the 1 stressed vowel of a focused constituent. If it is true (and it is hard to doubt this) that a surface constituent may be an "information focus," and that such focused constituents will be considered to be metrically strong, with special care being taken in performance (by means of tonal and rhythmic exageration) to make sure that their strength is appreciated, then the realization of this situation as "capitalization" (= being the locus of association of a "nuclear tone" with expanded pitch-range, having rhythmic prominence, etc.) of the highest-stressed word in the focused constituent, follows without further ado from the ideas developed in this thesis. The notion of "rhythmic prominence" is a natural extension of ideas about timing developed in the next chapter; the rest of this account of the effect of "focusing" a constituent depends on the principles we have stated for tune-text association, given the assumption that "focused" material will be tonally exagerated.

The last case we will consider in which other factors besides the NSR and CSR intervene in determining the labelling of metrical patterns, involves consideration of the failure of contraction before "deletion sites." Much of the discussion (which will be as brief as possible) involves matter unrelated to metrical node-labelling, but one crucial step at the end does depend on a motivated pre-emption of the NSR, so that there is a reason to include the whole business in this section.

We pointed out in section 3.2.6 that in certain cases, in order to derive the correct tune-text association, it seemed to be necessary (given the assumptions of our theory) to rebracket the metrical pattern of the text. This rebracketing in general had the effect of "demoting" a node to a lower position in the tree--for example, we suggested that the words <u>it's</u> <u>an</u>, in "it's an unusually dark night," should wind up adjoined to <u>unusually</u>.

This rebracketing was noted to have some relation to the phenomenon of cliticization, in which certain originally independent words (in English, the class of monosyllabic function-words such as is, the, an etc.) become attached, in a destressd form, to some adjacent word, and interact phonologically with it in a way that does not normally occur across word-boundaries. We also abserved, however, that the postulated rebracketing was not restricted to such clitic elements -- we gave a couple of rather complex examples involving destressed redundant material, but we might simply have cited cases involving, say, a dissyllabic preposition like over:

4.3/



The syntactically-motivated bracketing, in 5a, yields the tune-text association in 5d (we utilize the /L H L/ surprise/redundancy tune, since the locus of its first pitch is rlatively easy to intuit). This is a perfectly possible way to combine the text and tune we have selected. But there is another way, shown in 5e; and this second association pattern, given our assumptions, suggests the rebracketing in 5b, in which over is attached to the word on its right, rather than to the root of the prepositional phrase.

Of course, we have other options for dealing with facts of this type; we could revise our theory of tune-text association, or we could make freer use of "elaborated" tunes such as the one motivated for 3.2.6/18, and discussed in 3.3.

But it is worth pointing out that (within the approach we are advocating, in its most general form) there is good reason to accept the existence of rebracketing in at least some cases of this type, whatever else may be going on.

Standard theories of phonology have generally assumed that some syntactic constituent structure is wiped out before the phonology begins, and that the rest is destroyed in the course of the phonological derivation, so that the level of phonetic representation has no organization except as an ordered sequence of segments. In SPE (p. 5) Chomsky and Halle specify phonetic representation as "a two-dimensional matrix in which the rows stand for particular phonetic features; the columns stand for the consecutive segments of the utterance generated; and the entries in the matrix determine the staus of each segment with respect to the features." Except for the vertical elaboration of the feature-content of the string of segments, this view seems to be common to almost all phonological theories which take the trouble to define their terms with enough care that it is possible to determine what their view is.

Our approach has implicitly contradicted this view -- tune-text association almost certainly applies to the output of the textual phonology, and nevertheless we have supposed it to depend on a very considerable structuring of the text, what we have called its <u>metrical pattern</u>.

Thus the most basic assumptions of our theory depend on the idea that the phonological component is not so much a destruction of structure, which maps a complex tree onto a simple serial ordering of segments, as a transmutation of structure, which maps a structure suitable for operations in one domain (syntax and semantics) onto a structure suitable for use in another domain (the motor control of articulatory gestures, and its perceptual analogue).

That the motor control of articulatory (or any other human) gestures should be a variety of Markov process that hops, dully and inexorably, from one feature matrix to the next, issuing orders on the basis of the list of features in the particular segment it finds itself in, and then jumping to the next segment, where the process is repeated -- this idea strikes me as absurd. The most telling argument against it is that it provides no reason at all why <u>rhythm</u> (which is a complex relation among segments, not anything intrinsic to any one of them) should be so commonly and indeed inevitably a feature of gestural sequences of all kinds.

In terms of the Markovian model of motor control, in which the gesture has no organization but as a sequence of segments, rhythmic patterns could only be introduced by performing enormously complex calculations designed to minutely alter the durations (and other aspects) of individual segments, and requiring the motor control itself to be able to realize n-ary durational features specified to a very high n. An alternative theory seems preferable, in which motor control is a much more abstract and structured kind of process.

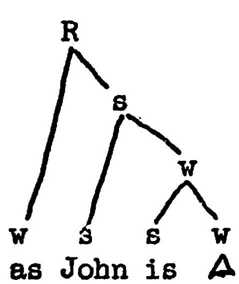
The point of this digression has been to make plausible the idea that phonetic representation should impose far more organization on sequences of segments than just what is implied by the fact that they are *sequences* and *of segments*, and therefore to make plausible our claim that the structures of syntax are transmuted by the phonology, rather than destroyed by it.

If this is true, then we have good reason to want to perform certain rebracketings. In a theory in which phonology progressively destroys structure, the fact that an auxiliary, originally attached to a verb-phrase or verb, ends up as part of the last syllable of the subject, requires only that an excessive number of word-boundaries do not intervene in the string, and that the auxiliary form be destressed. In the approach we are suggesting, a (metrical) rebracketing of *John* [*is coming*] to [*John is*] *coming* is a necessary preliminary to the derivation of *John's coming*.

We have not tried to give a formal account of either the preconditions or the effect of the rebracketing operation; nor will we. But surely one of the preconditions will be that the constituent which is "cut loose" be metrically *weak*.

On this assumption, it is possible to give an interesting account of the failure of contraction before "deletion sites." Consider the subordinate clause in the sentence "Sam is as tall as John is."

4.3/6



We assume, along with most right-thinking syntacticians, that this clause has an empty position in surface structure, in place of the understood adjective tall. Whether this empty position was there from the beginning, or derived by deletion, and how it should be symbolized, are irrelevent to the argument.

Now, in this particular example, $[\text{is } \Delta]$ is metrically weak, since it is redundant; we might have constructed the example as "Sam isn't tall, but John is," in which case it would have been metrically strong. It doesn't matter, since in either case, there is a constituent $[\text{is } \Delta]$, and as weak as is may be, in comparison to lexical material, it obviously wins out over nullity in the competition for metrical value. The scansion $\begin{bmatrix} s & w \\ \text{is} & \Delta \end{bmatrix}$ violates the NSR, to be sure; but the NSR only said to put the strong position on the right if we have no good reason to do otherwise; and the fact that there is quite literally nothing on the right in this case, constitutes as good a reason to do otherwise as the distinction between new and old or important and unimportant.

Given this scansion, the auxiliary is in 4.3/6 cannot be "cut loose" to rebracket with the subject NP, because it is metrically strong. If the sentence had been "John's tall," then is would have been metrically weak (since tall, both by the NSR and on "communicative dynamism" grounds, would have been the strong element), and therefore would have been free to rebracket to the left.

Of course, the constituent $\begin{bmatrix} s & w \\ \text{is} & \Delta \end{bmatrix}$ is itself metrically weak, and contains no lexical material -- on these grounds it might be considered free to rebracket. In the example we have given, that would make no difference, since it is already attached to John. But even if placed in a configuration where rebracketing would affect it in some way, $\begin{bmatrix} s & w \\ \text{is} & \Delta \end{bmatrix}$ is not monosyllabic, or at least not "monometrical" -- and therefore its failure to merge phonologically with a neighboring word can be charged to the same account as the similar behavior of $\begin{bmatrix}\text{s w}\\ \text{over}\end{bmatrix}$ and other dissyllables.

5. The Rhythms of Speech.

We have made considerable use, in the preceding sections, of the idea that it is natural for people to impose metrical patterns on sequences of events. These metrical patterns are in essence a hierarchical grouping of relatively strong and weak elements, in which a more or less complex pattern, on one level, serves as a single aspect of a simple opposition on a higher level. This concept has proven its value in helping to define the basis of tune-text association, and in providing a new perspective on stress rules.

As we pointed out in chapter 2, this notion of metrical pattern, in itself, provides no constraint on the arrangement of its terminal elements in time, other than the constraint of sequentiality (i.e. that the ordering defined by the orientation of the tree is an ordering in time).

This constraint of sequentiality is not entirely trivial -- not every aspect of phonological representation corresponds to a distinct position in time, phonetically or even intuitively. The various constituent features of a segment, for example, do not (in general) have temporally separate identities, although they are otherwise theoretically independent.

Empirical observation suggests that only those phonological entities which are temporally distinct can function as terminal elements of a metrical pattern. The contrary case is certainly conceivable, e.g.

5/1

$$\begin{array}{c} R \\ \left[\begin{array}{cc} s & w \\ \begin{bmatrix}+\text{round}\\+\text{back}\\+\text{high}\\+\text{syll.}\end{bmatrix} & \begin{bmatrix}+\text{voc}\\-\text{cons}\\-\text{low}\end{bmatrix}\end{array}\right]\end{array}$$

where the strong and weak positions in some particular case would correspond to disjoint subsets of the feature specification of a single segment. These disjoint subsets of features would, in this hypothetical case, be simultaneous in their temporal realization; the division implied by the metrical pattern might be determinable, say, from their distinct behavior with respect to some phonological process which could most generally be said to apply to features of metrically adjacent positions, whether the positions in question were temporally distinct or nay.

I don't know of any cases which would compel such an analysis, and I would be surprised to find any. The association of the well-ordering of the elements of a metrical pattern with the well-ordering implied by the sequential arrangement of events in time seems to me, intuitively, to be an essential property of meter.

However, we are not required to state this property as we did in a preceding paragraph, by saying that "only those entities which are temporally distinct can function as terminal elements of a metrical pattern." Another way of looking at the matter would be to say that those entities which function as the terminal elements in a metrical pattern are temporally sequential (and therefore temporally distinct).

It may be possible to make this principle do some phonological work, for example in the treatment of so-called <u>breaking rules</u>, in which a single segment splits (under certain conditions) into two segments which seem, in some sense, to partition certain of its features. Thus, we find cases in which æ → iə , ɔ → uə .

Irrespective of its possible utility in the analysis of such examples, the principle that the ordering of terminal elements in a metrical pattern corresponds to an ordering of events in time, seems to be empirically valid.

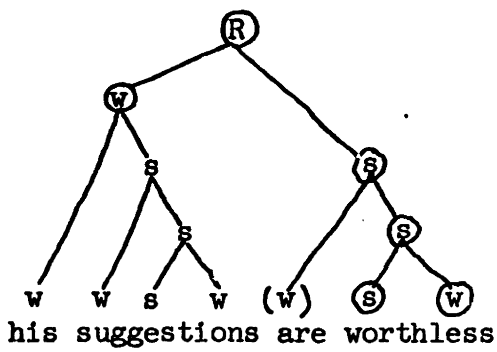
However, from the very beginning of linguistic investigation, it has been claimed that stress patterns (reanalyzed by us as <u>metrical patterns</u>) imposed much greater constraints on temporal patterning in speech than simply the requirement that syllables are arranged in a certain order, which has been taken to be so trivial as not to need comment. The reason for this pervasive belief is the intuition, shared by almost everyone, that speech has its <u>rhythms</u> (in the ordinary-language sense of the word), and that in a language like English, these rhythms are influenced (or even defined) by stress patterns.

Unfortunately, attempts to verify this intuition instrumentally have not met with very good success, with the result that most researchers have been forced to conclude that the whole concept of stress-timing is a kind of collective hallucination, or at best a tendency which influences the temporal patterning of speech in a relatively marginal way.

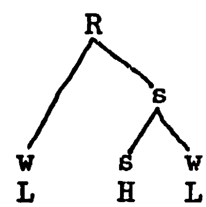
In this chapter, we will attempt to argue the position that the intuition of stress-timing in English reflects a fundamental and important aspect of the linguistic system, namely, the existence of what we have called the metrical grid. We will also try to give some reasons why this intuition cannot easily be verified by the discovery of objective isochrony of successive stresses.

We introduced the concept of the metrical grid, in an embryonic way, in chapters 2 and 3. Our reason for bringing it in at that time was to explain an otherwise arbitrary aspect of tune-text association, namely the fact that a tone which corresponds to a complex metrical constituent of the text, is realized in alignment with the "main stress" of that constituent, its "designated terminal element." Consider the example below:

5/2a



2b



The establishment of congruence between the metrical patterns of tune and text tells us only that the initial low tone corresponds to the textual constituent [his suggestions] . We offered a special "Metrical Association Rule" (2.4.3/29) to give the empirically correct result, that the initial low tone associates with the mainstressed syllable of the constituent, the stressed syllable of suggestions.

We then argued that this result was not an arbitrary one, but could be explained in terms of the nature of metrical grids, with which we argued the metrical patterns of tune and text were to be aligned.

This metrical grid, we suggested, is a kind of intuited structure of time itself, while metrical patterns are intuited structures of events in time. Neither could really exist without the other, but they are distinct. This distinction (parallel to the distinction between meter and rhythym in music) is a difficult one to grasp, but it is crucial that it be made. We will assume that the argumentation in previous chapters has established the plausibility of our idea of metrical patterns; we will begin this chapter with a general discussion intended to clarify the basic idea of a metrical grid as an intuited structure of time, and then we will attempt to show that this idea has empirical consequences which justify its acceptance. In the initial discussion we will take our examples from music, since the issues involved are more clearly defined in this domain.

The only way in which a structure can be conceived for time, is as some way of dividing it up -- as Cicero says, "numerus in continuatione nullus est: distinctio, et aequalium et saepe variorum intervallorum percussio, numerum conficit;" there is no meter in continuation: the distinguishing and performance of equal and (often) differentiated temporal intervals establishes meter. It would be a mistake, however, to conclude that there is no intuited structure to time other than that which inheres in the particular sequence of temporal intervals which are distinguished and performed on a particular occasion. Anyone who has had the slightest musical training knows that the "meter" of a piece is something more abstract than the sequence of temporal intervals defined by the note-values of a particular fragment -- the first three full measures of the following song (the beginning of Schubert's Das Wandern) are entirely different in terms of the sequences of temporal intervals they define, but they are all quite straightforwardly bars of $\frac{2}{4}$ time:

5/3

The further fact that the linguistic (and melodic) phrasing of the song cuts across the bar lines, also is irrelevant to the nature of the metrical grid defined by the concept of $\frac{2}{4}$ time. The linguistic and melodic phrasing of the song (which are pretty much the same, in this case) have an existence of their own, comparable to what we have been calling metrical patterns; the point to be made here is that the metrical grid defined by the time-signature and bar lines also has intuitively real existence, and in a way which is independent of the metrical organization of tune and text, although the interrelationship of pattern and grid (what we have referred to as *alignment*) is constrained in certain ways.

There is no surer indication of this independence than the fact that such alignments can be changed, without changing the basic nature of either the metrical pattern or the metrical grid. We gave some cases of different possible grid alignments of vocative chant examples in chapter 2, when we first introduced the idea of *metrical grids*. On a higher aesthetic plane, we might cite the two different alignments of the NP *das Wandern* in the previous example; both share the property that the main stress of the NP goes on the down-beat of the measure, but the alignment of the post-tonic syllable is different.

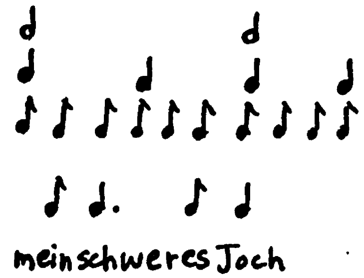
A musical cxample showing more complex differentiation of metrical alignment can be found in two treatments of the NP *mein schweres Joch* in J. S. Bach's aria *Hört ihr Augen auf zu weinen*:

5/4a

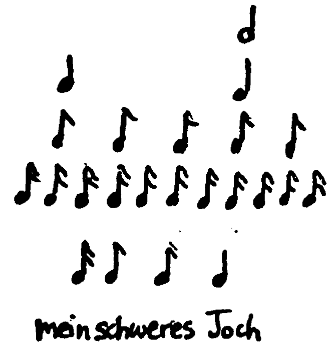
4b

Both of these settings of the noun phrase presuppose the structure imposed by $\frac{3}{8}$ time, and neither gives rise to conflict between the stress pattern of the text and the inherent stress pattern of this meter. Shorn of melisma, the alignments are these:

5/5a

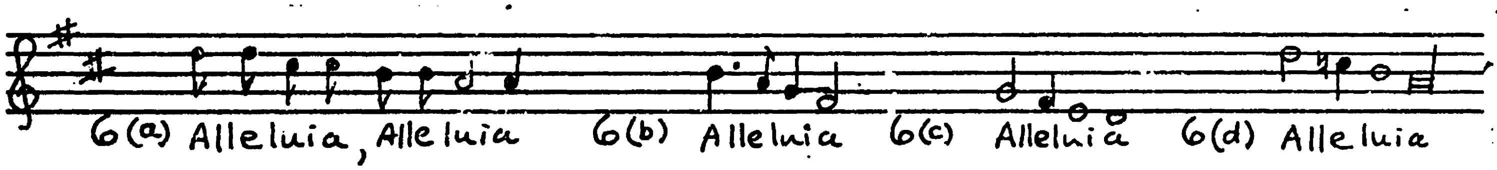


5b



In both cases, Joch is in the strongest metrical position, the two syllables of schweres are aligned with lesser positions, while the possessive pronoun mein is the weakest of all. We do not know this from a simple consideration of the sequence of temporal intervals defined by the note-values involved, which give (to the four syllables of the text) durations in the ratios 2:6:2:4, 1:2:2:4, but from a consideration of the way these ratios fit with the metrical grid defined by the time signatures and bar lines.

The following settings of the text Alleluia, taken from Schütz's Bringt her dem Herren, are interesting for a number of reasons:



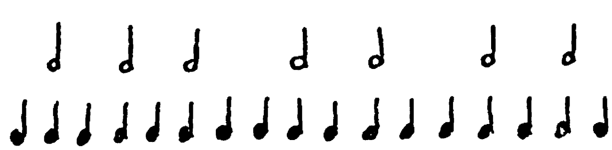
The first interesting point is that they are all set, in essence, to the same tune -- four descending notes of the scale. In the first case, two words of text are set to a single instance of this tonal pattern; in the last three cases, the tune-text relation is one-to-one.

The word *Alleluia* is, all in all, assigned five different sequences of note values, in the ratios 1:1:1:1, 1:1:4:2, 3:1:2:4, 4:2:8:8, 4:2:8:16. (Actually, in the piece these examples are drawn from, there are a number of other variants as well).

A second interesting point is that the piece (I believe) originally was written without time-signatures or bar lines. The modern edition from which I took the examples has supplied them, the time-signatures changing, throughout the piece, back and forth between bars of four half-notes and bars of three half-notes, with the internal organization of (these often rather long) measures being also (judging from the textual and harmonic structure) somewhat variable. This should not be taken to mean that Schütz intended the meter of his piece to be a simple sequence of note-values, without further structure; on the contrary. What it does mean is that the structure of the metrical grid, for music of this type, was not constrained in the way represented by a modern time-signature -- rather, it was intended to be patterned in a way which was believed to be implicit in the text.

In this respect it is often said to be like speech (appropriately enough, since it was an essentially *declamatory* style). The metrical grid fragment given below is not a well-formed example of any particular musical time-signature, since the number of "quarter-notes" (of constant duration) which intervene between the positions on the next level up (symbolized as half-notes) is sometimes one and sometimes two:

5/7



This kind of metrical structure is nevertheless typical of declamatory music. Sometimes (as in French Baroque practice) the changes in meter are carefully notated by frequent changes of time-signature; sometimes there is no barring at all; sometimes there is an arbitrary barring, which the performer is expected to disregard. John Hoyle, Dictionary of Music, 1770, writes of recitativo that "notwithstanding this sort of composition is noted in true time, the performer is at liberty to alter the Bars, or Measure, according as his subject requires. . . " Hoyle does not say that the performer should act as if there were no bar lines; merely that the bar lines should be altered "as the subject requires."

It is worth our while to step back for a moment, and consider the nature of metrical grids which incorporate irregularities of this sort. We have argued that a metrical grid is an intuited structuring of time, through which the intuited structuring of complex events (represented by metrical patterns) is given temporal patterning. There are two different (and sometimes conflicting) ways in which people are prone to regard the temporal status of events -- first, as durations, entities which occupy temporal intervals of varying lengths; and second, as temporal points, as entities which occupy (occur at) variously defined points on the continuum of time. These two ways of looking at events are commonly distinguished in the grammatical category of aspect -- they are also commonly distinguished in different cultural approaches to musical rhythm.

Following an established usage, we call these two approaches to rhythm additive and divisive. In an additive system, the basic metrical scheme consists of some repeated pattern of long and short intervals, in some ratio to each other. The structures that result need not have any special periodicity in their internal organization; thus the Bulgarian dance rhythm Puščeno consists of repetitions of the pattern S Q Q S Q , where S (slow) and Q (quick) are in the temporal relation of 3:2 (Singer 1974). Such a system suggests the view of events as durations, and the structuring of time into some specified sequence of different temporal intervals.

In a divisive system, we characteristically find a sequence of beats, usually of nominally equal duration, which are classed as relatively strong and weak according to some principle of organization -- the musical examples of metrical grids we have been considering in this section have been examples of a divisive system. Such a system suggests the view of events as points, and the structuring of time as a hierarchy of "pulses" of different periodicities. 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.

We will not have very much more to say about the additive view of rhythm, since it has no application to English. It is worth underlining, however, how fundamentally different this conception is from a divisive system, even when the surface patterns that result can be interpreted in either way. We will therefore take a brief side-trip to chronicle a case in which a musical tradition shifted from one conception to the other.

In the style of twelfth-century polyphony in Europe, a set of _rhythmic modes_ was established, consisting of various patterns of alternation of long and short notes. The basis of the system was a _nota brevis_, of one _tempus_, and a _nota longa_, of two _tempora_. Modes containing only such notes were termed _modi recti_, or "correct modes." Other note-values were possible, for example a _longa_ of three _tempora_, or a _semibrevis_ of 1/2 a _tempus_; these note-values were considered _ultra mensuram_, or "beyond measure," and modes in which they occurred were called _modi ultra mensuram_. The _modi ultra mensuram_ seem to have developed, historically, later than the _modi recti_, and to have been derived from them.

According to _Waite_ 1954:

"This system is completely unlike our conception of rhythm. It does not consist of a succession of beats of equal duration. . . on the contrary. . . (it) is founded on the repetition of a given pattern of long and short notes. The pattern itself is an entity, and only by the repetition of this unity is rhythm created. . . At all times the form of the pattern is retained as the metrical unit, while the individual notes receive their value only as they lie within the pattern."

The result of this conception was a notational system in which note-values were not generally represented, but only the relation of a given sequence of notes to the pattern of temporal intervals defined by the rhythmic mode of the piece.

Due to the particular restrictions placed on the system, in almost all cases the basic patterns of the modes can be placed in correspondance with a sequence of six notae breves, arranged in two groups of three.

The reason for such restriction seems to have been that the music was fundamentally polyphonic in character, with different voices proceeding in different modes, and it was felt to be necessary to establish some organizing principle according to which a correspondance between modes could be made.

Nevertheless, according to Waite, it is wrong to "jump to the conclusion that there is in reality but one type of rhythm in all the modes; all the modes are in triple time, and therefore the six modes are nothing but the various possibilities of arranging time values within a given triple-time measure. . . Such a conclusion is far from being the truth of the matter. . . these patterns not only determine the rhythmic flow of organa, but also the melodic construction. . ."

In terms of the surface patterns that resulted, however, the rhythmic modes might easily be reanalyzed as resulting from a divisive system. In the next century (the 13th) such a reanalysis seems to have been carried out, although not without considerable rear-guard polemicizing by traditionalists. When Pseudo-Aristotle proposed that the ternary longa be taken as the normal value, with the original longa of two tempora to be renamed a semi-longa, and thus considered a subdivision of a fundamentally periodic ternary meter, St. Emmeram Anonymous responded violently that

> "item nec in variatione nominationis, quemadmodum semibrevis a brevi vidilicet semilonga vocaretur, quave ipsam nonquam esse credimus imperfectum, sed eam esse dicimus rectam longam et veram et insuper et perfectam."

> "(this note) should not be called semilonga, (differentiated from the ternary longa) as the semibrevis from the brevis, because we can never believe it to be an imperfect entity, but say that it is a correct and true and superior and perfect longa."

However, the inherent logic of polyphony won out, and the fundamentally additive rhythmic system whi h the 12th century polyphonists had adapted from the rhythmic traditions of monody became fundamentally divisive.

We have entered on this digression in order to make the point that there may very well be modes of metrical organization, operating in language, which are fundamentally different from those we will propose for English, so that claims of universality must be made with some trepidation.

The issue confronting us at this point is the nature of metrical grids which are irregular, that is, where the succession of elements at a given level of the hierarchy does not define a beat "obstinately and periodically fixed in time." Actually, there is no reason at all to expect that the metrical grids that play a role in everyday speech should be regular. It is no more necessary that ordinary English is in (say) $\frac{4}{4}$ time than that ordinary classical Greek should have been in Sapphic Strophe, or some other meter of the music contemporary with it. Both $\frac{4}{4}$ time and Sapphic strophe are special and somewhat arbitrary restrictions on meter -- if they were not, no one would ever have thought to notice their existence.

However, while it is not to be expected that the metrical grids that play a role in English are constrained in the way entailed by a time-signature, we have argued that they nevertheless exist. Since we have allied our notion of metrical grid to the divisive conception of musical rhythm, in which the elements are (we argued) considered as temporal points rather than temporal intervals, suppose we adopt a notation for metrical grids in which only points are represented, while maintaining the hierarchy of levels, and the restriction to duple and triple relations between levels. An example of such a grid:

```
5/8                   4
         3            3          3
         2   2        2    2     2
       1 1 1 1 1 1 1 1 1 1 1 1 1
```

We use numbers for place-holders, using a different number at each level of the hierarchy for convenience in reference. It may be objected that we have so far not defined any particular "note-values" for the elements of the grid. This is true: but it is still a kind of abstract structuring of points in time, and it will serve a number of important purposes just as it is. For example, the argument in 2.4.4 (concerning the alignment of a tone with the main stress of the textual constituent it corresponds to), and the argument in 3.2.3

(concerning the explanation of the fact that the initial tone of a $\overset{\dot{w}}{T_1}\,\overset{s}{T_2}\,\overset{w}{T_3}$ sequence deletes when the text has initial main stress, while the final tone never deletes), depend only on the existence of a grid pattern like that in 5/8, not only any particular "note-values" being given to its elements.

One might well ask how it could be possible for a completely abstract grid-structure like that in 5/8 to play such a key role in everyday speech; the time-signatures and bar lines of music, to which we are asked to look for an analogy, may be abstract, but at least (we hypothetically protest) their influence can be seen in the definition of the "beat" and the relationship of the time-values of the various notes to this "beat." But the metrical structures of music are not always realized in a way which results in experimentally verifiable isochrony; it depends on the period, the style and the performer. There have been many combinations of these three factors for which isochrony of "beats" could not have been found at all. Giulio Caccini, in writing of the proper performance of his Nuove Musiche (1602, translated in Playford 1654) says "I call that the noble manner of singing, which is used without tying a man's self to the ordinary measure of time, making many times the value of the notes less than half, and sometimes more, according to the conceit of the words; whence proceeds that kinde of singing with a graceful neglect, whereof I have spoken before."

The specific work of which Caccini writes is nevertheless composed with regular time-signatures and bar lines, and indeed often has kinds of metrical regularity and structure that go far beyond what the time-signature requires. Even in extreme cases of a free rhythmic style, the basic metrical structure remains perceptible, if the performer is skilled, so that an accompanist experiences little difficulty in following the soloist so as to come in on the principal beats exactly "on time." As Donington says of such declamatory singing, "there is a symmetry even in its irregularity, and unless the singer can sense this symmetry, his liberties will sound not expressive, but formless. . . <u>It must be shaped to a pattern which the mind can grasp</u>. . . It is, at best, one of the most exacting skills; but there is nothing more telling than a harpsichordist and string bass player moving as one man, foreseeing every wish of their singer, and joining with him in making every irregularity sound not only natural but inevitable." (<u>The Interpretation of Early Music</u>, 1974 edition, pp. 427-428).

Similar quotations could be found for musical styles of many other times and places. The point I wish to make is that the way a metrical grid divides up time must indeed "be shaped to a pattern that the mind can grasp," but there is no reason to suppose that this pattern will always be robotic isochrony. It is undoubtedly true that the "ordinary measure of time" (to which Caccini refers) is a kind of perceptual isochrony of beats, but even this "ordinary measure" may be far from objectively regular. I do not pretend to know how it is that people, both in music and speech, can understand such structuring of time under conditions of "graceful neglect," but I think it is clear that they can do it.

We thus conclude that the traditional intuition of "stress-timing" has its roots in the alignment of the text-tune combination with a <u>metrical grid</u> whose "beat" is "shaped to a pattern that the mind can grasp." There are a number of obstacles in the way of experimental verification of this "beat": 1) there are many possible alignments, in any given case, even with regularly-patterned grid, many of which would not give rise to isochrony of stresses even if the performance were metronomically regular; 2) the metrical grids involved may have "downbeats" somewhat irregularly defined, like those in recitativ, where the "time-signature" is constantly changing, so that again, even performance to metronome-standards would not necessarily meet expectations; 3) the performance may in fact be characterized by "graceful neglect," a factor which is better appreciated by the human mind than by the sound spectrograph.

Nevertheless, there is empirical content in the claim that an abstract structuring of time, of the kind we are representing by our <u>metrical grids</u>, plays a role in ordinary speech. Arguments for its existence can be found in three areas: 1) the effect of grid alignment on tune-text association; 2) the role of perception of a "beat" in distinguishing stress patterns for which no other cue exists; 3) the relation of the phenomenon of grid-alignment to the stress-shifting rule discussed in section 4.2.

Arguments of the first type are made in 2.4.4 and 3.2.3; we will not recapitulate them here. They constitute the primary reason for our postulation of a role for metrical grids in the intonational system cf English. In what remains of this chapter, we will briefly sketch arguments of the second and third types.

Before beginning these arguments, we will take a moment to observe how formally different our view of metrical grids is from our view of metrical patterns. In the first place, metrical grids are not trees -- we have seen no reason to suppose that they define any particular constituent structure, and there is also no real reason to believe that they are "rooted," that is, that in every case the highest level in the hierarchy consists of a single element. Generally in music one level is chosen as the "beat" and one keeps track of perhaps two or three levels "up," and one or two levels "down" from this starting-point. Something similar happens in speech, I believe, and forms the basis of the common observation that no more than four or five "levels of stress" are normally distinguished, despite the inherent power of a metrical pattern (or theory of cyclic stress rules) to define as many levels as one likes.

We might choose to formalize a metrical grid as an ordered set of ordered sets, $\{L_1 - L_n\}$, and a function F with the property that it maps every member of L_{m+1} onto some member of L_m in a way which preserves order. This preservation of order may be expressed in the constraint that $\forall \lambda_i, \lambda_j \in L_x, \lambda_i > \lambda_j \supset F(\lambda_i) > F(\lambda_j)$. We impose the stipulation that either one or two members of L_m must intervene between adjacent members of the image set of L_{m+1}, for all m, and interpret the mapping defined by F as formalizing what we have been representing as vertical alignment in the grid.

An illustrative example:

5/9

$$L_3\colon \{\lambda_1^3\}$$

$$L_2\colon \{\lambda_1^2\ ,\lambda_2^2\}$$

$$L_1\colon \{\lambda_1^1, \lambda_2^1, \lambda_3^1, \lambda_4^1, \lambda_5^1\}$$

$$F\colon\ \lambda_1^3 \rightarrow \lambda_1^2\ ,\ \lambda_1^2 \rightarrow \lambda_2^1\ ,\ \lambda_2^2 \rightarrow \lambda_4^1$$

The above specification of $\{L_1 - L_3\}$ and F defines a single measure of $\frac{4}{4}$ time with a quarter-note upbeat (or equivalently a single measure of $\frac{2}{4}$ with an eighth-note upbeat -- in the more straightforward notation in which F is represented by vertical alignment, it would be:

5/10

$$\begin{array}{ccccc} & \lambda_1^3 & & & \\ & \lambda_1^2 & & \lambda_2^2 & \\ \lambda_1^1 & \lambda_2^1 & \lambda_3^1 & \lambda_4^1 & \lambda_5^1 \end{array}$$

In the notation in which numbers are used as place-holders on each level:

5/11

```
  3
  2 2
1 1 1 1 1
```

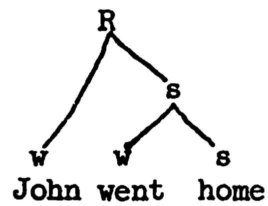
The "inherent stress pattern" of the grid can now be formally represented by comparing the number of times F^{-1} is defined for a given element of a given level -- thus in the example given above, $F^{-1}(\lambda_1^1)$ is undefined, while $F^{-1}(\lambda_2^1)$ is defined, telling us that λ_2^1 is a metrically stronger position than λ_1^1. $F^{-1}(F^{-1}(\lambda_4^1))$ is undefined, while $F^{-1}(F^{-1}(\lambda_2^1))$ is defined, telling us that λ_2^1 is a metrically stronger position than λ_4^1.

To make clear the nature of the restrictions on F: <u>every</u> member of a given level is mapped on to <u>some</u> member of the next level down. The inverse of $F(F^{-1}$, which maps "upwards" in the hierarchy) is thus not defined for all members of a given level. Let the members of a given level for which F^{-1} is defined, be called <u>defined</u>, and the members of a given level for which F^{-1} is undefined be called <u>undefined</u>. Then there may not be more than two adjacent undefined elements; but adjacent defined elements are forbidden. In musical terms, this simply means that a given note-value is always (potentially) subdivided into either two or three parts -- our reasons for conceiving the system in terms of the <u>position</u> of <u>points</u> rather than the <u>relation</u> of <u>durations</u> have already been expounded.

The primary purpose of this formalization we have just sketched is to make it clear that the notion of a <u>metrical grid</u> can be defined in a way which does not make it a tree structure (i.e. it is not necessarily "rooted," no constituent structure is established, etc.). The particular kind of formalization we have chosen (a mapping among elements of ordered sets) is simply an attempt to render formally the intuition of hierarchically-related periodicities.

Given an idea, along these lines, of what a metrical grid is, we can now consider some simple examples of how a text (along with its congruent tune) might be aligned with such a grid. Consider the simple text "John went home," with a metrical pattern as given below:

5/12



The particular choice of tune, and the details of its association, are irrelevant to the discussion, and will be disregarded.

It seems clear that the easiest and most straightforward way to establish an alignment is to match the three syllables of the text with three adjacent elements of the first-level set, as in 13a below. It seems likely that in any grid structure with more than a single element, at least two levels are obligatory; given the existence of a second level, either the alignment in 13b or that in 13c ought to be possible, since either two or three elements of level one can separate the occurrences of elements on level two:

```
                                      2       2                           2
5/13a     1    1    1           13b  1    1    1          13c  1    1    1
          John went home             John went home            John went home
```

Suppose we now decide to "lengthen" John, that is, to align the following syllable not with the adjacent element of level one, but rather with an element one position removed.

5/14
```
 2          2
 1  1 1     1
John went home
```

Now we are *required* to include an element of level two aligned with *John*, because of the prohibition against more than two members of level one separating adjacent members of the image set of level two.

Various other alignments are obviously possible, for example these:

5/15a
```
 2    2         2
 1 1  1    1    1
John (ø) went home
```

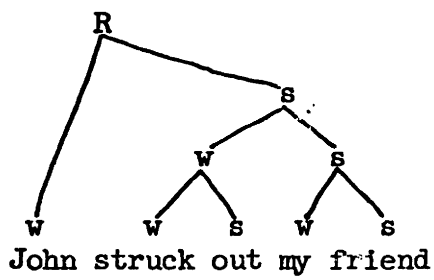
15b
```
 2    2     2
 1 1  1  1  1
John went home
```

The alignment in 15b is unproblematical -- it is simply a "slower," more deliberate version of 13b. The alignment in 15a is a little more interesting. If *went* is placed two positions removed from *John*, but is itself limited to a single position, with *home* following hard on its heels, then the metrical pattern of the phrase requires the following situation:

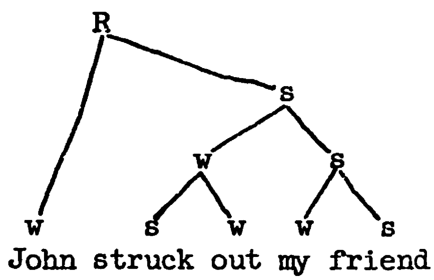
We have elements of level two mapped onto the first and last members of a sequence of five elements on level one. But it is impossible, by definition, for more than two elements on level one to intervene between the positions corresponding to adjacent elements of level two; therefore, there must exist some element of level two which is ordered between these two previously required elements. Since at least one element on level one must separate the image of this new element of level two from that of those previously given, the only possible position for this medial 2-element is as given in 15a. As a result, this "beat" at the 2-level aligns with nothing in the text. This state of affairs is perfectly possible; but I think it would be interpreted to mean the presence of some kind of special "juncture" in the utterance. In other words, the presence of a boundary, like those we have argued to be associated with boundary tones, would be implicated. While I am not in a position to carry this idea much past the speculative stage at the moment, such phenomena suggest to me the idea that the "boundaries" whose existence we have postulated on tonal grounds, in fact have phonetic content even in the textual string, namely the phonetic content of zero. Like zero in arithmetic, they would serve as placeholders in a string, taking up phonetic space without adding any features of their own, other than the effect that they have on the realization (and meaning) of the string in which they occur. The well-known phenomenon of pre-pausal lengthening, in English, might be laid in part to their account. In other languages, there are much more interesting and complex phonological and phonetic processes which occur prepausally, often leading to the postulation of a special class of "pausal forms" (as in Biblical Hebrew) or "final forms" (as in Whorf's description of Hopi). The Hopi phenomenon especially (which seems to consist of the addition of a preglottalized copy of the final vowel, preceding at least one of the other rules of the phonology) lends itself quite naturally to an interpretation which begins from the assumption that a phonetic null occurs at (systematically significant) pauses. We will therefore suggest that alignments of the kind represented in 15a require the presence of such a boundary (symbolized by ϕ in that example) in the string, and conversely that the presence of such a boundary influences the alignment by acting as a place-holder. We will not make any special use of this suggestion in the arguments that follow; nor will we develop it further in this work. We introduce it as half of the raw material for a treatment of the phenomenon of "intonation break," the other half being the idea of boundary tones. The refining of this raw material into a theory is a task that will be postponed to another time and place. We will add here to what we have already said, only the observation that the relationship between a true "intonation break" and an instance where the speaker stops to think of what comes next, is (on our assumptions) exactly parallel to the difference between a rest in music, which is an integral part of the meter of the piece, and an instance in which the musician hesitates because of unfamiliarity or lack of technique, which inserts space that has no relation to the meter. The difference between a rest and a hesitation is generally quite perceptible in music; the difference between an intonation break (a systematically significant pause) and a hesitation due to uncertainty, is generally equally obvious in speech. If the intonation break is associated with special tonal phenomena, as in the examples considered in chapter 3, this difference is all the more marked; but I do not think that such tonal phenomena are necessary to the perception of a systematically significant pause.

Let us now consider the problems of alignment that arise with a slightly more complex text, "John struck out my friend." We will assume that the stress rules operate in the normal fashion; so that friend receives the main stress, and John the second highest stress. However, we will allow one optionality, since we wish to consider its effect on grid alignment -- in the verb-particle pair struck out, we will leave it open which of the two is the strong position:

5/16a



16b



It is a fact that this optionality exists; whatever the explanation, both versions are perfectly acceptable.

We begin the process of determining a grid alignment on the same assumption that we utilized in working on "John went home," namely that the simplest case is the one in which the syllables of the text are aligned with successive elements of level 1 of the metrical grid:

5/17

```
1     1     1   1    1
John struck out my friend
```

The highest constituent of the metrical pattern of the text divides it into (John) (struck out my friend). Since the second of these is the strong position, we know that one of the elements of the (obligatory) second level of the grid must be aligned with its main stress <u>friend</u>; this gives us:

```
                                2
5/18        1    1    1   1     1
            John struck out my friend
```

The existence of four elements in level one, preceding this position, requires that we impose more structure on the grid by adding elements to level two. The next most important stress of the text (the main stress or "designated terminal element" of its highest weak position) is <u>John</u> -- therefore an element of level two must be placed in alignment with this syllable, if there are to be any additional elements of level two at all (and we have argued that there must be):

```
5/19        2                   2
            1    1    1   1     1
            John struck out my friend.
```

Given the grid structure as we have developed it so far, and the general restriction that neither more than two elements, nor less than one element, may intervene between the images of successive elements of the next higher level, there is only one possible way to turn 5/19 into a well-formed grid:

```
5/20
            2         2         2
            1    1    1   1     1
            John struck out my friend.
```

N.B. We leave aside the issue of the definition of a third level, as it is irrelevant to our discussion.

The grid alignment defined by 5/20 has an interesting feature -- it is a perfectly well-formed alignment for the stress-pattern represented by 5/16a, in which out was strong and struck weak, but it is not a possible alignment for the stress-pattern defined by 5/16b, in which struck is the strong part of the verb-particle constituent. The reason for the ill-formedness of 5/20 as an alignment for 5/16b is simple -- by hypothesis, struck is metrically stronger than out, but in the grid alignment under discussion, out is aligned with a grid position which is stronger than the one allotted to struck.

This does not mean that this text cannot be uttered with the stress pattern in which struck is stronger than out. But it does mean that in order to find a well-formed grid alignment for this case, we will have to cancel the assumption with which we began, that the syllables of the text should be aligned with adjacent elements of the first level. It is easy to see why this is so. We know that some element of level two must be established between those aligned with John and friend. From the argument given in the last paragraph, it is clear that this 2-element must be aligned with struck. This gives us the following structure:

5/21

```
2     2             2
1     1      1   1  1
John struck out my friend.
```

But 5/21 is not a well-formed grid structure -- two elements of level one intervene between the second and third elements of level two, which is permitted, but <u>nothing</u> on level one intervenes between the first and second elements of level two, and this is forbidden. The simplest way to resolve this difficulty is to insert the required 1-element.

5/22
```
 2      2             2
 1  1   1    1   1    1
John struck out my friend.
```

Now the relation of textual syllable to 1-elements is more complex -- in particular, <u>John</u> is separated from <u>struck</u> by a syllable-less 1-element. But both the grid structure and its alignment with the text are now well-formed.

Thus the "simplest" grid alignments for the two variant stressings of this text are these (considering only the two lowest levels):

5/23a
```
 2           2        2
 1      1    1   1    1
John struck out my friend.
        w    s
```

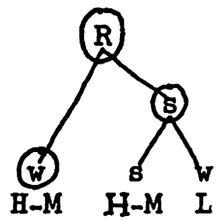
23b
```
 2      2             2
 1  1   1    1   1    1
John struck out my friend.
        S   w
```

Keeping the preceding discussion in mind, I would like to draw the reader's attention to three extremely suggestive facts:

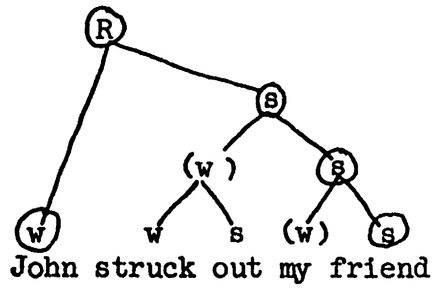
```
                     w   s   w
```
<u>Fact 1</u>) Given a normal three-tone tune, e.g. H-M L-M L, the tune-text associations for these two variant stressings of "John struck out my

friend" are identical:

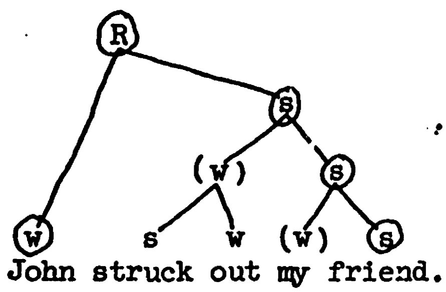
5/24a



24b



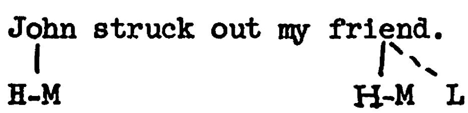
24c



N.B. We reiterate that the only variation allowed, for the sake of the argument, is in the relative stressing of struck and out. If the pattern is changed in some other way (e.g. by anaphoric destressing of John) then everything changes, and the argument loses both premises and conclusion.

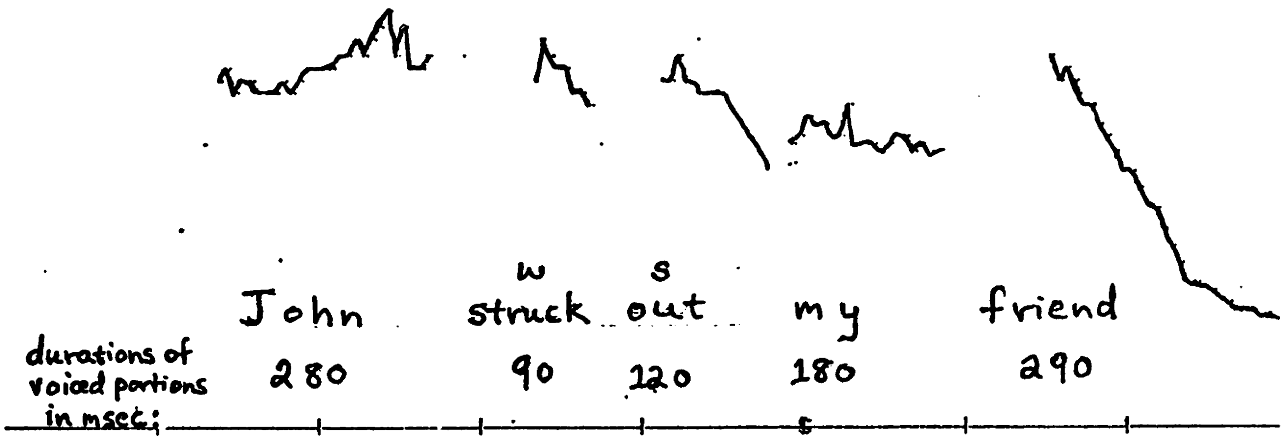
In both cases, the association will be as follows:

5/25

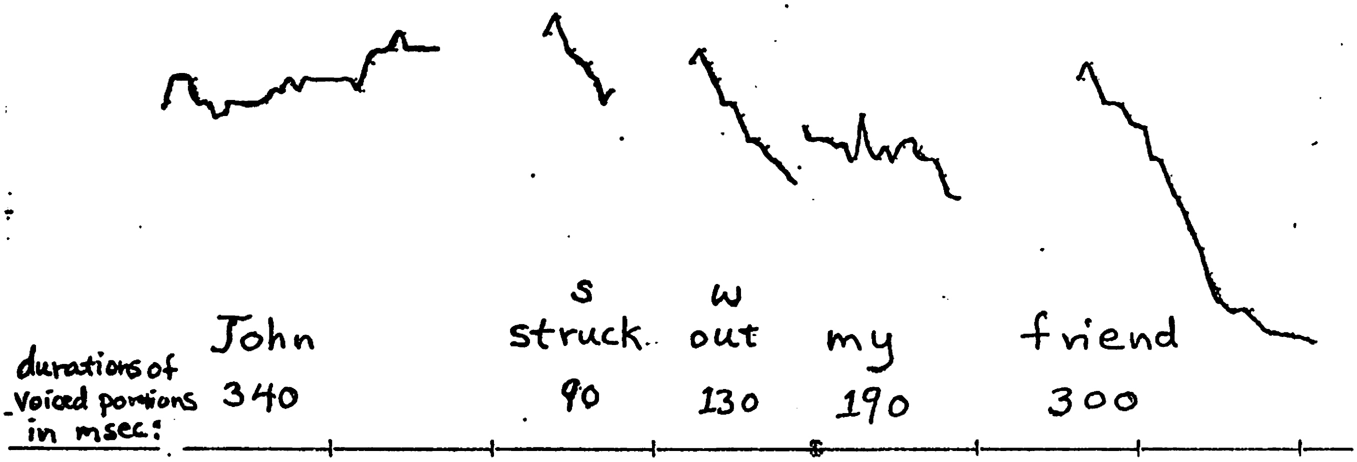


And indeed, the pitch contours show no significant difference in the tonal treatment of struck and out:

5/26a



5/26b



<u>Fact 2</u>) Despite this tonal similarity, the two versions are easily and reliably distinguished in performance -- I have had several people try the task informally, both as speakers and as listeners, with no instances of a listener failing to comprehend the relative verb-particle stressing intended by the speaker.

Fact 3) The communication of the relative stress level of verb and particle (given the assumptions of main stress on friend and secondary stress on John) is not accomplished by pitch difference, nor by intensity difference, nor by any significant difference in the relative durations of the verb and particle themselves. Rather, the difference in relative stressing of verb and particle is signaled primarily (perhaps exclusively) by a difference in the relative duration of the subject noun phrase, whose perceived stressing is not affected at all by the change. Specifically, the subject is longer (relatively) when the verb has higher stress than the particle: exactly as predicted by the "minimal" grid alignments in 23a and 23b.

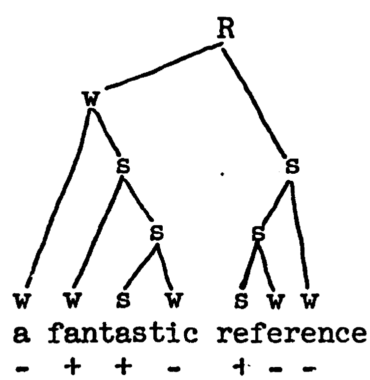
I have no very impressive body of data to submit in proof of the claim just made, although it was quite strikingly true of the dozen or so tokens I have examined instrumentally. (Interestingly, there was no isochrony of adjacent stresses in any of these tokens -- however, the argument we are making depends in no way on any claim of isochronous stress-timing, but only on the existence of a metrical grid whose structure is grasped by both speaker and hearer).

The issue should be an easy one to resolve, both by analysis of natural speech and by the use of synthetic stimuli in a perception experiment. The point of introducing the example into this discussion, even without experimental verification, is that it provides a paradigm for a whole class of arguments designed to lend empirical support to a theory of grid alignment as the basis for the rhythms of speech, with some mysterious but psychologically real effect on its timing. These arguments are of the most classical kind in linguistics -- if an abstract entity (in this case, the metrical grid) serves reliably to differentiate examples which are otherwise minimal pairs, then it is linguistically real. If this abstract entity is difficult to isolate and define instrumentally (as most linguistic entities are), this does not mean that it doesn't exist, but only that the instrumental techniques are not subtle enough.

The last kind of empirical evidence for the existence of metrical grids (at least the last kind which we will consider here) concerns their relation to the phenomenon we dealt with in the last chapter under the heading of the thirteen men rule. We will compare the possibilities of grid-alignment for the pair of examples "anaphoric reference / fantastic reference." We will argue that the nature of the structures that result explains the existence and peculiarities of the stress-shifting rule, although exactly how the grid-alignment and node-labelling processes are interrelated, is a matter we leave open.

We begin with the text "a fantastic reference," with the normal metrical and segmental stress pattern, as indicated below:

5/27



We start in the way we have been starting all along, by the simple expedient of aligning the sequence of syllables in the text with a sequence of elements on level one:

```
5/28        1 1 1 1  1 1 1
            a fantastic reference
```

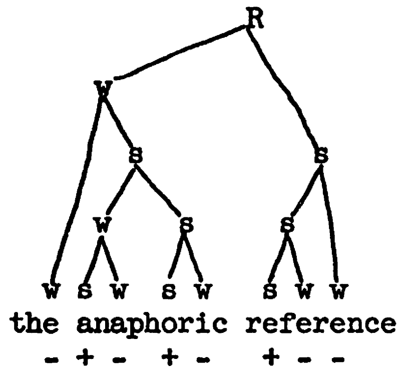
A second level, we have suggested, is obligatory; one of its elements must obviously be aligned with the main stress of the phrase (the stressed syllable of <u>reference</u>), or the resulting configuration could not possibly be well-formed. Since three first-level elements precede the position of this required "2," there must be another "2" to break up the string, and the only possible place for it is the main stress of <u>fantastic</u>, yielding the structure and alignment below:

```
5/29            2     2
            1 1 1 1  1 1 1
            a fantastic reference
```

If we wish to have a third level (and I suspect we would), it could only consist of a single element, aligned with the stressed syllable of <u>reference</u>. The result is well-formed on all counts.

Now we consider the case of <u>anaphoric</u> <u>reference</u>. By hypothesis, the underlying metrical pattern is normal, by the word-stress rules and the NSR:

5/30



We begin by aligning the syllables with a sequence of 1-elements; a second level is required, and as before, we place 2-elements over the stressed syllable of <u>reference</u> and the main stress of the preceding adjective:

5/31

```
              2    2
  1 1 1  1 1   1 1 1
the anaphoric reference
```

But in this case a third 2-element will be required, and the only place it can go is on the initial syllable of <u>anaphoric</u>. When we add a 3-element to the stack on the stressed syllable of <u>reference</u>, we get:

5/32

```
                   3
    2    2         2
  1 1 1  1 1   1 1 1
the anaphoric reference
```

This structure and alignment is in accord with the rules as we have defined them. It fails to mark either of the two stressed syllables in <u>anaphoric</u> as primary, but nothing says it has to. However, it does begin with a rather long and complex upbeat, from the point of view of the 3-level -- rather (in fact exactly) like beginning a piece of music with the sequence

5/33 

Such things happen in music, but relatively rarely, because it is quite difficult to perform such a sequence without causing the listener to falsely suppose that the first bar-line (which precedes the first 3-element, in our notation) is positioned between the first and second eighth-notes. Of course, such a false supposition, in the case under consideration, would lead to the hypothesis that the main stress of <u>anaphoric</u> is on its first syllable; and perhaps this is what is happening in some such cases. But I think that the <u>thirteen men</u> rule is not (in general) a perceptual error, but rather a feeling about metrical structure shared by both speaker and hearer. So we will assume that the speaker attempts to avoid the perceptually difficult situation in 5/33 by adding structure on the 3-level in 5/32.

Given the metrical pattern in 5/30, in which the main stress of <u>anaphoric</u> is in the normal place, the only possible alignment of an additional 3-element is as indicated below, since the 3-element must map onto one of the left-over 2-elements, and the metrical pattern of the text tells us that <u>-phoric</u> cannot be weaker (in terms of the grid alignment of its designated terminal element) than <u>ana-</u> :

5/34

```
             3    3
       2     2    2
    1 1 1  1 1   1 1 1
 the anaphoric reference
```

However, the structure in 5/34 is not well-formed, purely as a metrical grid, since 3-elements map onto adjacent 2-elements. There are only two ways to rectify this situation. The first is to salvage the grid structure in 5/34 by interpolating additional grid elements on lower levels -- the minimal interpolation is given below:

5/35

```
                 3           3
           2     2     2     2
         1 1 1   1 1   1   1   1 1 1
       the anaphoric  (ø )( ø ) reference
```

This solution is perfectly acceptable, and indeed represents one way of saying the phrase; however, it requires the interpolation of phonetic nulls into the textual string. The pragmatic effect of such a rendition seems to be to emphasize the adjective <u>anaphoric</u>, although it is not "contrastively stressed" (in which case the basic stress pattern would have been different, with concomitant changes in tune-text association, and with no requirement for "pause" between adjective and noun).

The phonetic nulls in 5/35 (perceived as a kind of pause) seem to be different in source from the sort of intonation-break we perceive between clauses, or setting off parentheticals, or (in some cases) between subject and predicate. This difference can be observed in the fact that some such "pauses" constitute barriers which semantic operators like negation cannot penetrate -- the "pause" in 5/35 is certainly not of this character.

In any case, there is a second alternative by which we could rectify the impossible situation in 5/34, in which 3-elements map onto adjacent 2-elements. Specifically, we could violate the metrical pattern of the text (or perhaps we should say, *revise it*) by aligning the additional 3-element with the first of the 2-elements:

```
5/36         3       3
             2   2   2
           1 1 1 1 1 1 1 1
          the anaphoric reference
```

The result is well-formed from the point of view of grid structure, and leaves us with a simple syllable-by-syllable mapping from elements of the 1-level onto the textual string; its only marked feature is that it requires us to apply the "*thirteen men* rule." That is, it requires us to readjust the node-labelling of [[ana][phoric]], changing it from [ws] to [sw].

This is the true nature of the *thirteen men* rule -- it has its roots not in the nature of the metrical pattern itself, but in certain difficulties that may arise in aligning metrical patterns (if they have certain properties) with a metrical grid. The rather odd "structural description" that we gave for this rule in the last chapter, and especially its restriction to "hierarchically equal levels" of the metrical pattern, were nothing but an attempt to define *those properties of a metrical pattern which will give rise to difficulties of the sort which we noted in reference to 5/34*. That is, *those properties of a metrical pattern which will make syllable-by-syllable alignment with a metrical grid difficult or impossible*.

I am not sure how to represent this _pattern_ / _grid_ interaction theoretically. There are various possibilities; from the standpoint of our present discussion, the answer doesn't matter. The point of the argument is that the nature of the _thirteen men_ rule cannot be understood except in terms of such interaction of _metrical pattern_ and _metrical grid_. Neither concept by itself provides any rationale for the particular properties of this very natural rule; but together they make it not only natural, but very nearly inevitable, given certain initial characteristics of the system.

6. Summation.

6.1 The Problem Redefined

The inquiry we have begun in the preceding pages, as St. Augustine says of rhythm, nullum certum habet finem, "has no fixed stopping-place." The pages themselves, however, must come to an end. We will prepare the final cadence by recapitulating the themes introduced in chapter 1, summing up what we have said about the questions raised there, and adding a few additional remarks. A general concluding statement concerning the nature of the formal devices we have postulated, will sum up the essence of the whole (to the extent that this is possible) by advancing a general hypothesis about action and perception. It may strike some as obvious, and others as absurd, but its implications for language have been neither obvious nor entirely without value, so it merits a more general evaluation.

In chapter 1, we noted three pre-theoretically striking aspects of "how a sentence is said" -- its stress, its tune, and its phrasing. We asked for an underlying representation for each of these, a phonetic representation, and a means of getting from the one to the other; and we wished to know how the interaction of stress, tune and phrasing was determined. Finally, we asked for the whole system (or set of systems) to be put in its proper place in the grammar as a whole.

The task has turned out to be a large one, as should perhaps have been obvious, and the answers cannot entirely be structured according to the categories defined by the questions. The organizing principle of our theory has been an ars metrica, a metrical system. This system consists of two complementary parts: a theory of metrical patterns, which imposes metrical structure on complex events, and a theory of metrical grids, which imposes metrical structure on the dimension of time.

Metrical patterns are formally defined as trees, with uniformly binary branching, and a relation strong/weak defined on the two elements of each nonterminal constituent. Metrical grids are defined formally as an ordered set of ordered sets, representing the levels in a grid, of which the lowest level (the initial set) is considered terminal; and a function F that maps each nonterminal level onto a proper subset of the next level down, in such a way that the image set conforms to the ordering relations on all levels. There is an additional stipulation on F that no two members of its image set, and not more than two members of the complement of its image set, can be adjacent.

Formal structures of this kind are not commonly found in linguistic work, with the result that we have had to define the notion of grid for ourselves. However, a formalization of the theory of time-signatures and bar-lines in music would be nearly identical.

The effect of a grid is to define a multidimensional structure on the unidimensional flow of time. One aspect of this structure is that the elements of the terminal level are distinguished in importance according to how many times F^{-1} is defined on them; in the visually convenient grid notation, importance is defined by the number of higher-level elements vertically superimposed. The intuition behind the move to define "importance" on an ordered set in this way, is that the levels of the grid represent intersecting periodicities.

In performance, pattern and grid each require the other's existence. Their interaction has been termed _alignment_, since it aligns the terminal elements of the metrical pattern relative to the grid's structuring of the time-line. Alignment is subject at least to the condition that the weak member of a constituent, in the metrical pattern, cannot be aligned with a more important grid position than the strong member of that constituent. If these nodes are nonterminal, this alignment condition is defined (by necessary proxy) on their designated terminal elements (the terminal positions reached by a path intersecting no _w_ nodes). Properly stated, the alignment condition has this result without special stipulation.

We have argued that in English, _text_ and _tune_ are underlyingly separate. The metrical system (with the addition of certain specifically linguistic and language-specific principles) defines metrical patterns for text and tune independently, and then combines them (into a single abstract gesture) by the establishment of _congruence_. Formally, _congruence_ is a set of correspondances between nodes of different metrical patterns, defined recursively by setting root in correspondance to root, setting daughters of corresponding nodes in correspondance, when the _strong/weak_ relation is equivalently defined in both constituents, or else (when the _strong/weak_ relation is defined differently on the daughters of corresponding nodes) setting the trochaic constituent in correspondance to the strong element of the iambic constituent.

We sought justification for this conception of congruence as the basic step in combining tune and text, both in the patterns of association that arise in simple cases, and in the fact that tonal constituents do not cross-cut textual constituents, nor do textual constituents cross-cut tonal constituents, although one constituent of either type may span several of the other type.

The necessity of pattern/grid alignment also plays a crucial role in tune-text association, according to our theory, explaining both the specific segmental location of a tone associated (by congruence) with a complex textual constituent, and the patterns exhibited by tonal deletion.

The necessary imposition of a metrical pattern on temporally ordered behavior results in the phenomenon of linguistic stress patterns. The structural (rather than segmental) nature of metrical patterns was argued to explain certain typological features of the rules by which individual languages assign "stress," including those features which have been taken (in conventional theories) to argue for cyclic rule application. We proposed a reformulation of the English phrasal stress rules, and a partial reformulation of the English word stress rules, in light of which the need for the phonological cycle as a theory of rule-ordering was called into question; we proposed that metrical rules could be allowed to apply simultaneously to their maximal domain. An attempt to collapse the metrical CSR with the word-level node-labelling principle, led to the hypothesis that the level at which a metrical rule applies should be seen as defining which elements it must consider as primitive (=lacking internal structure). This hypothesis clearly has a close relationship to the concept of "strict cyclicity," an aspect of the theory of the cycle which is logically separable from the issue of ordering the application of rules according to the size of their domain. This observation leads toward a view of stress rules which affirms a version of strict cyclicity, but rejects the cycle itself.

Certain aspects of English phrasal stress (the "thirteen men" rule) were argued to result from a class of difficulties that arise due to the constraints on <u>pattern/grid</u> alignment. This suggests that the alignment process is a rather abstract one; its role in tune-text association suggests the same thing.

Concerning the nature of the tonal system of English, considered apart from its instantiation in association with any particular text, we have proposed a set of preliminary hypotheses. We have suggested a tonal phonology and a tonal morphology; higher levels of organization have not been given any theoretical status, more out of ignorance than design. The phonology for tone in English specifies four distinct tones, resulting from the interplay of two binary features $\begin{bmatrix}\pm \text{High}\end{bmatrix} \begin{bmatrix}\pm \text{Low}\end{bmatrix}$. We suggested that if these features are general, then the tetratonal linguistic system, and the puzzling naturalness of minor thirds, may be related. The morphology defines the canonical form for a tonal word as $T_1T_2T_3$, where T_2 is stressed. We proposed that the organization of the tonal lexicon is fundamentally ideophonic, resulting in a complex of properties which it shares with non-tonal ideophonic systems: metaphorical (rather than referential) meaning, non-arbitrary sign-meaning correspondances, and non-concatenative meaningful elements. These properties were argued to result from a basically iconic mode of meaning, which is conventionalized to various degrees in particular cases.

We observed the existence of the phenomenon of boundary tones, and made a preliminary proposal for their theoretical representation. It was noted that the theoretically motivated boundary elements seemed to take up metrical space, unlike other sorts of boundaries -- the existence of a phonetic zero was proposed, to serve as a terminal placeholder for these boundary elements. It was observed that a treatment of intonation breaks as boundaries of this sort may explain an interesting fact (about their effect on tune-text association) noted in chapter 1.

The system outlined above answers, at least in a preliminary way, most of the questions raised in chapter 1. We have provided a theory of underlying representation for stress and for tune, and a theory of tune-text association which accounts for their interrelations. We have suggested a preliminary hypothesis about the nature of phrasing, in at least some of its manifestations, which helps to define its relations to stress and tune -- this aspect of our theory has not been developed very far, since it immediately leads us into domains of syntax and semantics which cannot be discussed within the scope of this work.

We have discussed some aspects of intonational derivations, but have undoubtedly left the process hanging somewhat short of its lowest systematically significant level, so that we cannot claim to have advanced a phonetic representation for intonation.

The major remaining issue, of those we raised at the beginning of this work, is the place of intonation in grammar. Since "intonation," pretheoretically defined as "the way something is said," turns out to reflect a complicated intersection of a number of different factors, this question must be redefined in order to be answered in any detail, but a general point is in order before we proceed to give such detail as we intend to.

We have not discussed the relationship of the intonational factors we have been investigating to other linguistic phenomena, real or alleged (semantic interpretation, syntactic transformations, etc.), because we were not really forced to do so in order to understand them. If tune-text association were ordered in the syntactic cycle, for example (assuming the syntactic cycle to exist), it is unlikely that we could get very far without taking cognizance of the fact.

I think that there is excellent reason to believe that "intonation," considered (falsely) as a whole, is pretty much independent, in theoretical terms, of the rest of language. Obviously the various systems interact extensively on the surface -- for some odd reason this is often taken to indicate theoretical interdependence, as if systems independently defined on the same domain could somehow fail to give rise to patterns of correlation. This style of argument (which is astoundingly common in linguistics) seems comparable to the claim that the emergence of moire patterns when two pieces of silk are placed in contact, demonstrates that there could never have been more than one piece of fabric all along.

If the separation of intonational phenomena from the rest of grammar can be maintained, an enormous reduction in the number of possible grammatical systems results, since rules in one domain are not allowed to make reference to features (either primitive or constructional) in the others. This fact establishes an initial presumption in favor of separateness, to be maintained unless and until strong evidence to the contrary can be found.

The tonal system of English can be argued to be distinct from the rest of the grammar on a number of grounds. Among the arguments:
1) phonological rules never need to refer to tonal features;
2) tune-text association does not depend in any way on the derivational history, phonologically speaking, of the text;
3) tune-text association is not influenced by the transformational history (real or alleged) of the text, other than to the extent that this history defines the surface structure which serves as the textual basis of the association;
4) the "meanings" of tunes do not interact with semantic interpretation of the text, but only serve to define conditions on its appropriate use in a discourse -- certain apparent exceptions to this principle are spurious, as was argued in Liberman and Sag 1974.

In other words, tune and text touch only at their derivational periphery. The only apparent exception to this, that I know of, is that the pragmatic value of a given "word" in a complex tune is defined on the textual constituent with which it is associated; but this is not much of a concession, since it is hard to see how else tonal meanings could be defined.

The assignment of metrical patterns to textual material is a rather different matter, since it is an operation defined purely in the textual domain. Still, the most restrictive hypothesis says that metrical bracketing and labelling is not interspersed with the rest of the system, but takes place at one well-defined point in the derivation. Bresnan has argued that stress-assignment is ordered in the syntactic cycle, on the basis of an interpretation of the distribution of main stress in cases where originally clause-final material has (in some ies of syntax, to which she thus hopes to give credence) been removed. I do not find her arguments especially compelling, although her proposal seems preferable to the alternatives that have been advanced against it by Lakoff, Berman and Szamosi, and Bolinger.

The main issue here seems to me to be that whatever is true, must be true universally. The examples which Bresnan cites in favor of her proposal are in too narrow a part of the language, are too rare in actual usage, and are too subject, in individual cases, to alternative explanations in terms of information focus or some different structural principle, for a child learning English ever to hit on the idea of applying the NSR in the syntactic cycle, unless Universal Grammar offered no alternative.

If stress (of the sort in question) is a feature of structure, as a metrical theory claims, then one might argue that it is in fact a humanly necessary feature of structure, at whatever level of abstractness. If this is true, and if Bresnan's syntactic theories are correct, then the required universality might be ascribed to her ordering hypothesis. Given the attractiveness of this idea, I wish I was a little more convinced by the facts of the case.

The boundaries to which we have ascribed the role of delimiting intonational phrases seem (in some cases) to condition phonological rules (the case of "pausal forms"). Since some of these boundaries also seem to be associated with features of surface syntax, it seems likely that they are in place in surface structure prior to the application of phonological rules. According to our hypothesis about the nature of tune-text association, they must be in place in the text at the point at which congruence of metrical patterns is defined.

The work presented in this thesis does not so much solve a set of problems, as redefine them. We have provided at least provisional answers to most of the questions asked in chapter 1, but in the process, a host of other problems, whose existence we did not originally suspect, arose to take their place in the list of tasks imposed on a theory of intonation. In science, this multiplication of problems is generally a sign of progress. A cynic might argue that it is so considered primarily because further work is created for the profession, which thus perpetuates itself through a chain of inquiries in which ultimate success would spell the extinction of the enterprise. A more sympathetic way to put it, would be that the motivation of most scientists is not so much to arrive at answers, as to discover more interesting questions.

I have heard that a certain mathematician, when asked (on his arrival at an English university) what benefit he intended to derive from his studies, responded that there were two questions for which he hoped to find an answer -- "what is number, that man should know it? and what is man, that he should know number?" We began our investigation with a question comparable to the first of these -- "what is intonation, that man should know it?" It is fitting to end the story by asking a similarly modified version of the second one.

6.2 The Metrical Organization Hypothesis.

In one of his letters (Epistola CI) St. Augustine wrote that in omnibus rerum motibus quid numeri valeant, facilius consideratur in vocibus; "the role that meter* plays in all movements, is seen more readily in speech." We have made explicit reference to music in develop-

*Augustine does not use the term rhythmus which he defines as an infinite sequence of quantitative patterns; nor the term modus, which represents a particular finite "measuring" of such a sequence; but rather the term numerus (lit. "number"), which seems to have its origins in the practice of "counting time," and by which he presumably means some non-quantitative kind of temporal structure, comparable perhaps to our metrical grid, although he demonstrates no awareness of the pattern/grid distinction, and therefore might be taken to refer to all non-quantitative aspects of temporal patterning.

ing our ideas about the metrical system; but the concept of such a system, as a complementary and mutually necessary structuring of event-sequences and of the time-line, is not inherently restricted to language and music. It would be surprising if this way of abstractly representing the temporal structure of complex behavior were to have been invented out of nothingness, to fulfill some new need defined by speech or song. Humanity (and its predecessors) was engaged in complex temporally-structured behavior long before any voice attempted to talk or sing, and it is hardly to be believed that the techniques evolved over the eons to cope with the problem of action, should have been precipitously abandoned in favor of an entirely new invention during the development of human vocal play and vocal communication. The apparent existence of similar structures in dance is a telling point -- one of the closest approximations (that I have seen) to the metrical system proposed in this work, is contained in Aldrich's discussion of 17th century Italian dance forms, which he reproduces in a notation derived from treatises of the period.

It is fashionable (and for good reason, since it is often true) to stress how different language is from other aspects of human behavior, and to insist on the particularity and (often) even on the arbitrariness of linguistic mechanisms. In the present case, however, the opposite conclusion seems forced on us. The metrical system, which stands as the organizing principle through which intonational phenomena are given form, shows unmistakable similarities (independent of the truth of our particular formulation of it) to the organizing principles of several other kinds of human behavior. There is some reason (evolutionary parsimony, if nothing else) to suppose that the cases of language, music, and dance demonstrate, in a more intuitively accessible way, a system which in fact is the organizing principle of all temporally ordered behavior. It is proper to give this supposition form, and a name:

The Metrical Organization Hypothesis: All temporally ordered behavior is metrically organized.

In its most general interpretation, this hypothesis would simply suggest the existence of a way of imposing abstract structure on the behavior in question; a way of imposing abstract structure on the time-line; and a way of relating the two types of structure. More particular versions of the hypothesis would claim general validity for our particular conceptions of metrical patterns, metrical grids, congruence, and alignment.

There is a corollary to the hypothesis, which suggests that the perception of temporally complex entities is also metrically organized; again, the nature of the different aspects of the system is independent of the hypothesis in its general form, but the suggestions made in this work constitute a starting-point.

The Metrical Organization Hypothesis, as we have stated it, suggests that the source and origin of "rhythm" is quite abstract, especially with respect to the fundamental division between event-structure (pattern) and time-structure (grid). It is worth noting that this division confers enormous practical advantage. A temporally complex action or event can be given a structure which is independent of whether it is performed rapidly or slowly, evenly or haltingly, with parts temporally expanded or contracted to suit the needs of the occasion. The viewing of such patterning as accomplished through something like tree-structure, is a necessary condition on the ability to invent new patterns of action, rather than relying on a limited repertoire of fixed gestures. An independent structuring of time helps organize the coordination of different abstract patterns of action, either by an individual or among friends.

It is worth pointing out that this tremendously advantageous system does seem to impose one requirement -- that the behavior in question be abstractly segmented into discrete subunits. An interesting consequence of this is that there could be no music without notes (structured, say, along the lines of Bolinger's gestalt contour-feature theory of intonation), no language without words and phonemes (in which the meaningful units were irretrievably smeared throughout a temporally unsegmented noise), and no dance without "steps."

Although I am not familiar with the literature involved (in whatever field such ideas are considered both relevant and permissible), I suspect that ideas along the general lines of our previous argument are not uncommonly advanced. St. Augustine certainly had the same basic intuition -- others have undoubtedly felt it too. The principal contribution of this present document, in my opinion, is to have taken a general intuition about the organization of temporally ordered behavior, given it a sparse and preliminary formalization, and made this formalization do a considerable amount of work in the description of stress, tune and phrasing in English, a description which has previously been accomplished on other principles, principles which were both more complicated and less related to any general consideration of the nature of the behavior described.

These first and faltering steps in the direction of metrical explanation have succeeded rather better than could have been expected, if the general approach were not correct. So we are in a position to make the knife cut the other way; since meter has taught us something about language, perhaps it is now time to allow language to teach us a thing or two about meter.

If St. Augustine was right, and the metrical structure which is seen in speech functions "in all motions of things," in some more-or-less modified form, then the study of speech and language, where such a large array of perceptually distinct and nameable elements serves as the theatre of metrical operations, may be able to shed an entirely unexpected kind of light on the structure of behaviors whose abstract metrical primitives are inaccessible to conscious inspection.

It would be nice to discover that there is at least that much connection between talk and action.

FOOTNOTES

1\. 1 e.g. Schmerling 1974.

2 Williams 1975.

3 The F_0 contours used in examples in this work were obtained by means of a hardware pitch-extracter (developed at Lincoln Labs) used in conjunction with a PDP-9 computer (Moses Reading Machine). The display and print-out programs used were developed by Douglas O'Shaugnessy.

4 The Sound Pattern of English, Chomsky and Halle. The notation presents 1 as the highest value of stress; successively subordinate levels of stress have successively higher numbers.

5 Lasnik 1972.

6 Liberman 1975.

7 Leben 1973, Goldsmith 1975, Williams 1971.

8 Crystal 1969, p. 151; Bailey 1971, pp. 45-48.

9 Crystal 1969, pp. 217-21.

10 Goldsmith 1974, 1975; Haraguchi 1975.

2\. 1 Halle and Keyser 1971.

2 e.g. Maling 1972.

3 The exposition here comes dangerously close to confounding "metrical pattern" with "metrical grid;" caveat lector.

4 This example is wrongly parsed, by the word-stress rules in 4.1, but the argument is still valid -- a correct example is cited as 2.4.4/6a.

5 This view was once the norm in American linguistics, and now seems to be undergoing a revival.

3\. 1 O'Connor and Arnold 1961, p. 71.

2 ibid. p. 15.

3 Crystal 1969, p.207.

4 *ibid*. p. 233.

5 *O'Conncr and Arnold* p. 4.

6 *Goldsmith* 1975.

7 *Pike* 1945, p. 26.

BIBLIOGRAPHY

Aldrich, Putnam (1966) Rhythm in Seventeenth-Century Italian Monody. W.W. Norton and Company.

Aronoff, Mark (1974) Word-Structure. Unpublished MIT dissertation.

Bailey, Charles-James N. (1970) A New Intonation Theory to Account for Pan-English and Idiom-Particular Patterns. in University of Hawaii Working papers in Linguistics, vol. 2 no. 3

(1971) Intonation. in University of Hawaii Working Papers vol.3 no.5

Berman, Arlene and Szamosi, Michael (1972) Observations on Sentential Stress. Language 48.304-25.

Bolinger, Dwight L. (1951) Intonation: Levels vs. Configurations. Word 7.199-210.

(1958) A Theory of Pitch Accent in English. Word 14.109-49.

(1965) Pitch Accent and Sentence Rhythm. in Forms of English, Harvard University Press.

(1972) Accent is Predictable (if you're a mind-reader). Language 48.633-45.

Bresnan, Joan W. (1972a) On Sentence Stress and Syntactic Transformations. in Brame ed., Contributions to Generative Phonology. University of Texas Press.

(1972b) Stress and Syntax: a Reply. Language 48.326-42.

Chomsky, Noam and Halle, Morris (1968) The Sound Pattern of English. Harper and Row.

Cooper, G. and Meyer, L. (1971) The Rhythmic Structures of Music. University of Chicago Press.

Crystal, David (1969) Prosodic Systems and Intonation in English. Cambridge University Press.

Diffloth, Gerard (1971) Notes on Expressive Meaning. CLS 8 (Chicago Linguistic Society).

Goldsmith, John (1974) English as a Tone Language. Unpublished Mimeo.

(1975) An Overview of Autosegmental Phonology. Unpublished Mimeo.

Halle, Morris and Keyser, Samuel Jay (1971) English Stress: its Form, its Growth, and its Role in Verse. Harper and Row.

Haraguchi, Shosuke (1975) The Tone Pattern of Japanese: An Autosegmental Theory of Tonology. Unpublished MIT dissertation.

Koster, W.J.W. (1962) Traite de Metrique Grecque. A.W. Sythoff.

Lakoff, George (1972) The Global Nature of the Nuclear Stress Rule. Language 48.285-303.

Lasnik, Howard. (1972) Analyses of Negation in English. Unpublished MIT dissertation.

Leben, W.R. (1973) Suprasegmental Phonology. Unpublished MIT dissertation.

(1975) Tone Patterns in English. Unpublished Mimeo.

Liberman, Mark (1975) On Conditioning the Rule of Subj.-Aux. Inversion. in NELS V (New England Linguistic Society).

and Ivan Sag (1974) Prosodic Form and Discourse Function. CLS 10.

Maling, Joan (1972) The Theory of Classical Arabic Metrics. Unpublished MIT dissertation.

O'Connor, J.D. and Arnold, G.F. (1961; new edition 1963) Intonation of Colloquial English: A Practical Handbook. Longmans.

Ross, J.R. (1972) A Reanalysis of English Word Stress. in Brame, ed. Contributions to Generative Phonology. University of Texas Press.

Schmerling, Susan F. (1974) A Re-examination of "Normal Stress." Language 50.66-74.

Singer, Alice (1974) The Metrical Structure of Macedonian Dance. Ethnomusicology Sept. 1974.

Sag, Ivan and Liberman, Mark (1975) The Intonational Disambiguation of Indirect Speech Acts. CLS 11.

Steele, Joshua (1775) An Essay Towards Establishing the Melody and Measure of Speech. Facsimile edition, Scolar Press 1969.

Williams, Edwin S. (1971) Tone in Margi and Igbo. Unpublished paper.

(1974) Rule Ordering in Syntax. Unpublished MIT dissertation.

Waite, William G. (1954) The Rhythm of Twelfth Century Polyphony. Yale University Press.